Discussion of two Papers Read at the Society Half-Day on Inverse Problems on December 10th, 2003

Christian P. Robert†

CEREMADE, Université Dauphine, and CREST, INSEE, Paris

Both papers by Cornford *et al.* and Haario *et al.* are impressive pieces of work that are perfect representative of the issues pertaining to Inverse Problems: huge amounts of (satellite) data, intractable likelihood functions, large dimensions and speed processing requirements. They also illustrate the necessity for multiple levels of approximations that arises in such problems and the correlated difficulty in assessing the effects of such approximations. In addition, they provide sophisticated additions to the theory and practice of MCMC algorithms.

Ultimately, the approximations used by the authors in both cases are Gaussian: in Haario *et al.*, the integral defining $T^{\text{abs}}_{\lambda,l}$ is discretised into $T_l(N_l)$ and the base-level observations y_l are normal $\mathcal{N}(T_l(N_l), C_l)$. In Cornford *et al.*, a normal mixture is built for the inverse model,

$$\mathbf{v}_i | s_i^0 \sim \sum_{k=1}^{4} \alpha_k(s_i^0) \mathcal{N}_2(\mu_k(s_i^0), \sigma_k^2(s_i^0) I_2)\,, \tag{1}$$

with neural network estimation for the functions $\alpha_k, \mu_k, \sigma_k$, but the MCMC processing of the resulting model is found to be too slow and sequential Gaussian approximations are used instead for the dynamic case. (As an aside, notice that the issue of additional Gaussian noise $\mathcal{N}_2(0, \tau^2 I_2)$ in the mixture model does not modify the mixture structure since it simply transforms the variance in $(\sigma_k^2 + \tau^2)\, I_2)$.) Still, it seems that a model like

$$p(\mathbf{V}|\mathbf{S^0}) \propto \prod_i \frac{p(\mathbf{v}_i|s_i^0)}{p(\mathbf{v}_i)}\, p(\mathbf{V},\,,$$

where the $p(\mathbf{v}_i|s_i^0)$'s are the mixtures of (1) should be manageable for simulation as is, given that the parameters of the mixtures are known: the function can be computed analytically and either a missing data structure can be introduced for Gibbs sampling simulation (Diebolt and Robert, 1994) or a random walk Metropolis algorithm can be implemented. The fact that the modes of $p(\mathbf{V}|\mathbf{S^0})$ need to be known in advance is only an apparent challenge in that the bimodality is a consequence of the lack of identifiability of the direction of the wind.

Even though the papers work within the Bayesian paradigm, I find the prior input fairly vague and limited, as far as the description goes in both papers. For Haario *et al.*, although some further knowledge on the $\rho^{\text{abs}}\{z(s)\}$'s, other than its positivity, could be used (altitude and gas correlations could also appear in the prior), it seems that the main prior modelling is related to the discretisation/regularisation choice $\gamma = \pm 1, \pm 2$

$$x_i = \hat{x}_i \pm \varepsilon^{\text{reg}} \surd(\Delta z^\gamma)\,.$$

For the model of Cornford *et al.*, there also seems to be very little prior input, either in terms of spatial modelling or geophysical and historical knowledge. One reason for this limited use of the possibilities offered by the Bayesian approach is the restriction imposed by the Gaussian structure of the model, especially in the case of the scatterometer data. (The last sentence of §4 in Cornford *et al.* is also fairly intriguing in that it seems to imply that prior distributions are simply stabilising devices, rather than summaries of prior information.)

As stated above, an interesting feature of both papers is the devising of novel simulation methodologies to handle the complex posterior distributions found there. The adaptive MCMC algorithm of Haario *et al.*

†This discussion was written while the author was enjoying the hospitality of the Department of Statistics, Bristol University, via a CNRS–Royal Society grant.

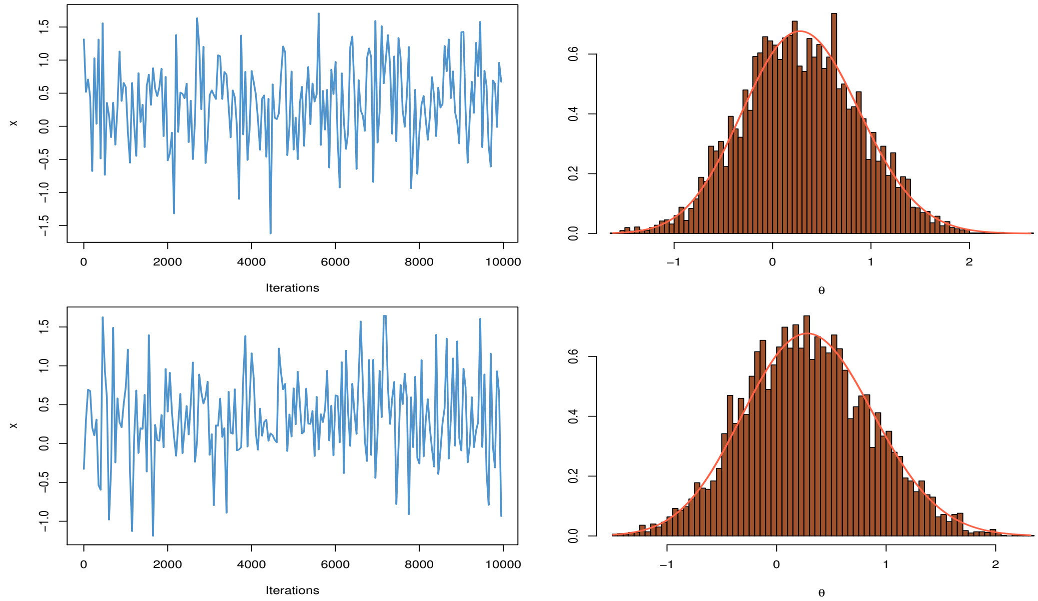

Fig. 1. Random walk Metropolis simulation of the t posterior distribution $\prod_{j=1}^{5}\left[\nu+(x_j-\theta)^2\right]^{-(\nu+1)/2}$ using a normal proposal with variance *(top)* $\hat\sigma_t^2=\frac{5}{t}\sum_{i=1}^{t}(\theta^{(i)}-\hat\mu_t)^2$, where $\hat\mu_t$ is the average of the $\theta^{(i)}$'s, and *(bottom)* a ridge-type version $\tilde\sigma_t^2=\frac{5}{t}\sum_{i=1}^{t}(\theta^{(i)}-\hat\mu_t)^2+\epsilon$, where $\epsilon=.1$. The left plots represent the sequences of simulated values and the right plots give the histograms of these values against the true posterior distribution of θ. In both cases, the fit is satisfactory, despite the lack of theoretical guarantees for the top scheme.

has been introduced in Haario, Saksman and Tamminen (1999, 2001, 2003) to overcome the difficulty of scaling the random walk Metropolis algorithm by an only adaptation

$$C_t = s_d \mathrm{cov}(X_1,\ldots,X_t) + s_d \epsilon I_d$$

and these papers are forerunners of an emerging class of more efficient MCMC algorithms (Robert and Casella, 2004). While ergodicity of the resulting process is necessary to establish the validity of the simulation method as an approximating technique, one may wonder about the practical relevance of such constraints. Figures 1 and 2 show that not all adaptive schemes are providing correct approximations to the distribution of interest. As noted in Andrieu and Moulines (2003), other schemes could be used while preserving the stability of the proper distribution. For instance, the updating of the covariance matrix of the proposal could be embedded in a grand chain $(X^{(t)},\Sigma^{(t)})$ by adding a performance component to the stationary distribution in a tempering mode, $\exp\{-\alpha H(\Sigma)\}$. It would also be of considerable interest to understand better why the single-site one-dimensional update SCAM can be so efficient in high-dimensional models since the performances of the Gibbs algorithm usually deteriorate in higher dimensions.

In the case of Cornford *at al.*, simulation technology is used at two levels, mode jumping MCMC (§4.1) and variational approximation (§4.2). The specific algorithm of §4.1 is not particularly appealing, given that it requires knowledge of the two modes of the posterior distribution. Local, Rao–Blackwellised and population Monte Carlo algorithms could be used as well (Cappé *et al.*, 2004, Robert and Casella, 2004). In particular, a particle filter (Doucet *et al.*, 2001) provides an efficient alternative to the two-stage approach of the authors that allows for simultaneous mode detection and simulation from the posterior. (This is even truer in a dynamic setting.) The variational Gaussian approximation in §4.2 is presented as an alternative to the costly MCMC algorithm of §4.1, but one must stress that the focus is also different, this technique providing acceptable approximations only to the first two moments of the posterior distribution and only when the latter is unimodal.

While the various approximations used by the authors are all acceptable as a result of the *Reality constraint* that makes a truly Bayesian resolution impossible (?) to implement, more elaborate assessments that

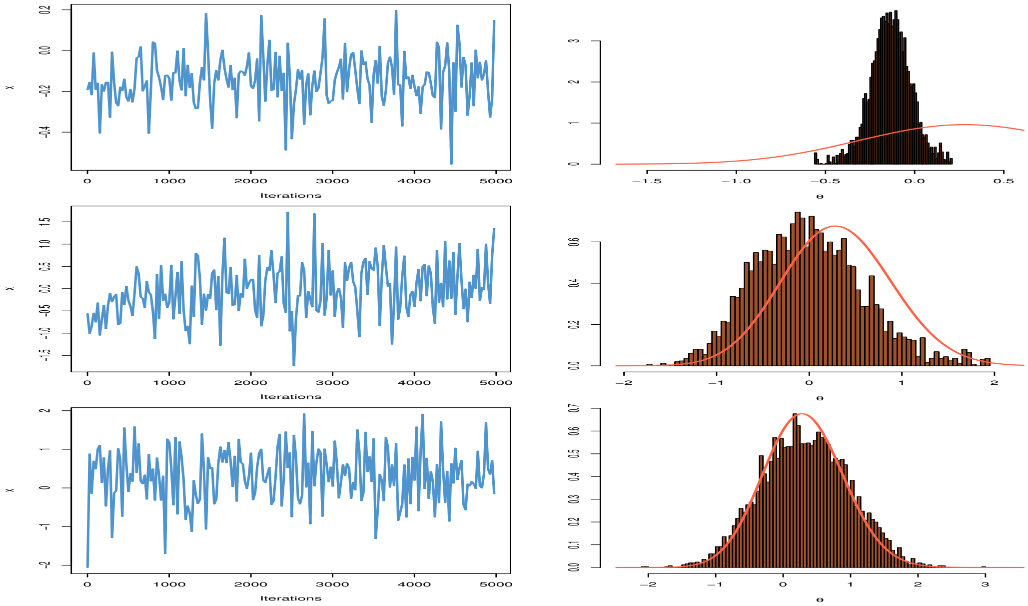

Fig. 2. Influence of the variance of the starting distribution in an adaptive MCMC algorithm with proposals $\mathcal{N}(\hat{\mu}_t, \hat{\sigma}_t^2)$, in the same setting as Figure 1. The starting variances are $.1$ *(top)*, 1 *(middle)* and 5 *(bottom)*. Even the largest variance fails to provide a convergent approximation to the stationary distribution, while the middle graph exhibits a case of poor mixing.

these approximations have limited consequences would be welcomed. Similarly, when iterative algorithms are used, the assessment (or non assessment) that convergence is not a problem should somehow appear. To conclude, I want to congratulate both groups of authors for obtaining a nonetheless satisfactory inference in such complex but comprehensive inverse problems where Statistics must play a role and I thus unreservedly propose the vote of thanks!

References

Cappé, O., Guillin, A., Marin, J.M. and Robert, C.P. (2004) Population Monte Carlo, *J. Computational Graphical Stat.* (to appear).

Diebolt, J. and Robert, C.P. (1994) Estimation of finite mixture distributions by Bayesian sampling, *J. Royal Statist. Soc.* **56**, 363–375.

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

Robert, C.P. and Casella, G. (2004) *Monte Carlo Statistical Methods* (second edition). Springer–Verlag, New York.