Bayesian Analysis (2009) **4**, Number 1, pp. 45–54

Comment on Article by Craigmile et al.

Alexandra M. Schmidt*

I would like to start by congratulating the authors for their interesting contribution, and thanking Brad Carlin for the opportunity of commenting on it. In reading this article (hereafter CCLPC) and the related references, one realizes how powerful the available tools for Bayesian analysis are, and how we can now tackle important problems in more realistic ways than only some years ago. Contributions such as this one illustrate the great development Bayesian methods have been experiencing since (Gelfand and Smith 1990) was published. However, despite these recent advances, there are many practical issues which still need to be addressed. And I understand, as outlined by the authors, that this work aims to open a dialog on practical strategies for hierarchical modelling.

My comments below follow the organization of the article and are based partly on my experiences in Brazil.

1 Model building

In an interesting section on exploratory data analysis and model building (Section 3), CCLPC makes clear that when tackling complex problems we should conduct the model fitting in compartmentalized fashion, separately validating and assessing the model fit of each component. I found this point extremely important. Although obvious, it bears repeating that life is easier when we start by solving simpler problems. From these simpler problems we gather a better understanding of the process(es) being studied and learn how better to communicate with the experts, which in turn may allow us to propose a more realistic model.

This section indicates that "model building should be a combination of EDA and scientific knowledge", but does not make very clear to the reader how the subject-specific knowledge on arsenic pathways was acquired by the modellers. Expert opinion should be part of the exploratory analysis; therefore, readers would likely benefit from a description of any behind-the-scenes communication between themselves and the subject experts, as well as how these interactions were structured.

Prior specification

Although we, as Bayesians, claim the benefits of Bayesian inference, most of us, including CCLPC, still make use of noninformative priors. In this sense, I wonder if we are fully using the advantages that the Bayesian paradigm provides. Given that most of the parameters in CCLPC enter as coefficients of linear models, they may be interpretable

*Instituto de Matemática, Universidade Federal do Rio de Janeiro (UFRJ), Brazil, http://www.dme.ufrj.br/~alex

 DOI:10.1214/09-BA401C

as effects of individual covariates; elicitation may thus be helpful to establish the priors. Similarly, I wonder if informative priors would not help in improving the estimation of the missing data.

Some Bayesian analysts have made important contributions on how to elicit prior distributions for such complex models. In particular, Tony O'Hagan and Jeremy Oakley have proposed an elicitation tool named `SHELF`*. A thorough introduction to elicitation is provided by O'Hagan et al. (2006). Although elicitation can be a challenging process, I believe we should be discussing more broadly how to incorporate it more frequently and more effectively in applications requiring expert knowledge.

Two doubts regarding the prior distributions arose while I was reading the paper. First, the prior variances for the baseline parameters are set to 1000 with the exception of that of the global-water parameter, α^W. Why was α^W assumed to be much more concentrated around 0, and how sensitive are the results to the choice of this prior variance for α^W? Second, it was not clear to me, even after reading (Santner et al. 2008), why the measurement-error precisions (ω_j) for the various media were assumed known. Would there not be information in the data to use instead the values of the relative standard deviation associated with each medium as prior information and assign some uncertainty to the ω_js?

Persistence of the effect of arsenic

Both Cressie et al. (2008) and CCLPC focus on the strength of the linear relationships between different stages of exposure to infer the importance of the various pathways. Individuals were monitored for seven consecutive days and urine samples were collected on the third and seventh days of the study. The response variable was the mean of the log-transformed arsenic concentrations in the urine of each individual.

It was unclear to me whether the level of arsenic assimilated by an individual over a short time period is reflected only in measures for that period or instead propagates to future observations as well. In a time series context, Alves et al. (2009) proposed a dynamic transfer model to capture the effect of carbon monoxide on counts of infant deaths. This kind of model allows for estimation of the temporal evolution of regressor impacts on a response variable. It would be interesting to combine the pathway and transfer function models with a view to examining variation in regressor impacts. However, a time series would be required to implement this combined model.

Handling different spatial scales and the proper CAR prior

Analysts are frequently confronted with observations made at different spatial scales. For example, Ravines et al. (2008) proposed a joint spatio-temporal model for the runoff of a basin as a function of rainfall. However, rainfall was measured at fixed monitoring locations spread across the basin, whereas the runoff measurements were taken at a

*Visit `http://www.tonyohagan.co.uk/shelf/` for further details.

single downstream location; this situation exemplifies the change of support problem (Cressie 1993). This problem was handled by overlaying a regular grid of locations over the basin and obtaining basinwide rainfall as the sum of spatially-interpolated values based on the rainfall measurements made at monitoring locations.

I find interesting how CCLPC deals with the spatial misalignment sampling schemes in the global-soil model. Following Calder et al. (2008), CCLPC defines the topsoil process (point-referenced data) as linearly related to the stream-sediment process, which is defined over watersheds (areal data). The latent stream-sediment process is assumed to follow a proper CAR prior. I understand that the choice of a proper CAR is made because interpolations of the stream-sediment process across AZ can be obtained. Use of a proper CAR prior can be coupled with methods that exploit the sparseness of the neighbourhood matrix, which can speed up the MCMC computations dramatically (see (Rue and Follestad 2003) for details).

CAR priors are commonly used for modelling epidemiological data for which observations are available for all areal units. The aim in such studies is usually smoothing rather than interpolating. Banerjee et al. (2004) [p. 82-83] note that, under a proper CAR prior, interpolated predictions for ungauged locations are not assumed to come from the same spatial process. Gelfand (private communication, ISBA 2008) suggested as a remedy that all locations, gauged and ungauged, be introduced in the MCMC procedure at once. Schmidt et al. (2009) faced a similar situation, but because our spatial domain was not very large, we assigned a multivariate normal prior distribution for the spatial effects, with a free-form covariance matrix. We avoided using a proper CAR because we were uncomfortable with the idea of having different spatial processes for gauged and ungauged locations.

2 Practical issues

The authors successfully describe many different practical aspects of Bayesian model fitting. Advances in hardware, together with the development of powerful computational statistical methods, have made Bayesian inference a natural paradigm for modelling highly complex processes. In particular, MCMC methods are now standard tools for fitting Bayesian hierarchical models. The spread of software such as `WinBUGS` has made Bayesian inference accessible to a wide range of researchers. As cautioned by many authors (e.g., Spiegelhalter et al. (2002)), it is fundamental to stress that these tools should not be used as a black box. Below I will describe my own experiences with various issues discussed in the paper.

Data management and software

There is no doubt that `SAS` is excellent software for dealing with huge datasets. However, I work in a developing country where funds are limited; therefore, it is essential that we have access to *free* software. In such situations, it is extremely helpful when researchers make quality software freely available.

It is helpful to code a particular MCMC algorithm using different programming tools. Although I find the `WinBUGS` software extremely useful, I find it essential to understand the underlying computations to determine whether the sampling procedure is being performed efficiently. Otherwise, longer chains may be required to reach convergence. An alternative algorithm may be required altogether. In our Graduate Program it has become common practice to use the `Ox Console`[†] language to code MCMC algorithms. The console version of `Ox` is free for non-commercial use. It is very similar to `R`, is easy to use, and deals with matrices and vectors very efficiently, often running the same algorithm faster than `R`. Usually, I write my codes both in `R` and `Ox` as a means of double-checking my computations.

Program coding and sampling schemes

One of the main challenges that I find in implementing a Gibbs sampler is to ensure that the code yields reliable results. As mentioned in CCLPC, organization and clarity are fundamental when writing an MCMC algorithm to sample from the target distribution. It is very useful to explicitly write down the full conditional distributions and verifying them analytically before writing the code itself. This helps avoid different sources of errors. Also, at this stage, one can investigate ways of simplifying the algorithm. For example, Schmidt et al. (2008) avoided sampling (and monitoring the convergence of) over 25,000 parameters by using a marginalized version of the likelihood. Marginalizing also helps to make computation more stable.

I agree with the authors that when dealing with unknown full conditional posteriors it is convenient to use random walk proposal distributions in Metropolis-Hastings steps. However, it can be challenging to tune the variance of the proposal distribution. To this end, the method proposed by Roberts and Rosenthal (2001, 2006) is very useful and easy to implement.

As described in CCLPC, fitting the model to synthetic data is a useful means of initially exploring the model. I usually perform this exercise considering multiple realizations obtained from both a single set of parameters and different sets of parameters. This approach helps to check the code, explore the range of possible responses, and yield hints on possible unidentifiability problems.

Checking convergence

There are situations in which convergence can be speeded up by choosing convenient starting values. For example, in (Schmidt et al. 2008) we propose a stochastic frontier model with a spatial component. Convergence of the regressor coefficients was reached faster when the starting values were set to the OLS estimates of a multiple regression model without the inefficiency component.

The model in CCLPC has an unspecified, but large, number of parameters. This

[†]Visit `www.doornik.com` for more details.

raises a general issue: how can one check for convergence of *all* parameters for models or submodels having hundreds or thousands of parameters? CCLPC mentions that convergence was checked mainly running multiple chains starting from multiple starting locations. It would be helpful if the authors discussed further how they chose the starting values and specified how many chains were run. I too, check convergence using multiple chains (usually two or three), but, additionally, use the convergence diagnostics available in `BOA` (Smith 2005) for those parameters that are key for the inference procedure.

3 Analyzing the results

Model checking

An essential (and challenging) task in Bayesian model fitting is checking whether the model provides a reasonable fit. Chapter 6 of (Gelman et al. 2004) illustrates several useful approaches to model checking. The approach selected for model checking will generally depend on the structure of the observations (e.g., presence of spatial or temporal correlation; nested observations). I also follow the approach pursued by the authors, of performing some kind of cross-validation, especially when dealing with point-referenced data. Cross-validation is very useful to investigate influential observations. For time series, I often check the model by considering one-step-ahead predictive distributions (e.g, Sansó et al. (2008)). Whenever missing data in a time series are being imputed within the inference procedure, I like to visualize the completed series by examining a graphical summary that includes predictive distribution at missing points, to detect possible discrepancies between observed and imputed data. Figure 1 shows an example of such a plot for a time series of concentrations of particulate matter at a monitoring station in Rio de Janeiro.

CCLPC makes use of standardized residuals obtained from *particular* realizations from the posterior distribution of the parameters. A potential difficulty with this approach is that it may be difficult to choose which and how many realizations should be used to adequately characterize the distribution of residuals. As an alternative, CCPLC mentions the use of posterior predictive checks which would be involved because of the great amount of missing and censored data. Gelman et al. (2005) extend posterior predictive checking to situations with missing or latent data by including unobserved data in the model checks.

Model comparison

In Section 4.5, CCLPC explores alternative models which assume proper CAR prior specifications for the intercepts of each medium in the LEB model. An interesting aspect is that model comparison is performed by examining the residuals and posterior summaries of the parameters obtained under the different specifications. It would be helpful to discuss further this important aspect of Bayesian model fitting. Despite the many recent advances in Bayesian modelling, there is no agreement yet in the literature

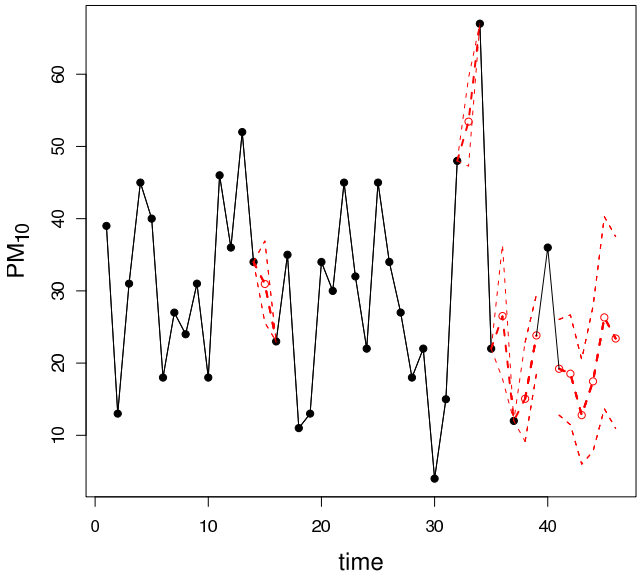

Figure 1: Example of a time series where the imputed missing data (open circles) are shown with their associated 95% posterior credible interval (dashed lines). Solid lines and filled circles represent observed data.

on how to compare different model specifications. Although there are advantages to using Bayes factors, it is well known that they are sensitive to the use of vague prior distributions. There are well known alternatives in the literature. See, for example, Chapter 7 of (O'Hagan and Foster 2003) for a nice review of the many different criteria available for model comparison.

When modelling spatio-temporal data, it can be useful to compare models through their predictive distributions. The scoring rules proposed recently by Gneiting et al. (2007) are relatively easy to implement once a sample from the posterior distribution has been obtained. See (Gschlößl and Czado 2007) for details on implementation.

Presentation of results

The complexity of the models we deal with also challenges us to think carefully about how to present the results. The project described in CCLPC and related references provides a good example of this challenge. Cressie et al. (2008) report summaries of the posterior distributions of regressor coefficients together with the acyclic directed graphs, whereas in CCLPC, summaries of the posteriors are presented in a different format (e.g., compare Figure 3 of (Cressie et al. 2008) with Figure 9 of CCLPC). The figures in (Cressie et al. 2008) emphasize the dependencies among variables, whereas those in CCLPC emphasize the visualization of posterior uncertainty while still depicting

the dependencies among variables. Ultimately, results can often be shown in a number of different ways; we should always be concerned about how to make visualization and understanding easier to the reader.

4 Concluding remarks

Once again, I thank the authors for opening this debate on practical issues related to Bayesian analysis of complex hierarchical models. We are dealing more and more often with richly structured data which require highly dimensional models to describe their underlying correlation structures. The Bayesian paradigm provides a natural way to model such complex structures. A Bayesian analysis generally involves four main steps: building the model, specifying a prior, obtaining a sample from the resultant posterior, and analysing the results, including model checking and model comparison. As CCLPC illustrates, we must keep in mind that each of these steps requires *attentiveness* and *reflection*.

References

Alves, M. B., Gamerman, D., and Ferreira, M. A. R. (2009). "Transfer functions in dynamic generalized linear models." *Statistical Modelling: an International Journal (to appear)*.

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2004). *Hierarchical Modeling and Analysis of Spatial Data*. New York: Chapman and Hall.

Calder, C. A., Craigmile, P. F., and Zhang, J. (2008). "Regional spatial modeling of topsoil geochemistry." *Biometrics DOI 10.1111/j.1541-0420.2008.01041.x*.

Cressie, N., Buxton, B. E., Calder, C. A., Craigmile, P. F., Dong, C., McMillan, N. J., Morara, M., Santner, T. J., Wang, K., Young, G., and Zhang, J. (2008). "From sources to biomarkers: a hierarchical Bayesian approach for human exposure modeling." *Journal of Statistical Planning and Inference*, 137: 3361–3379.

Cressie, N. A. C. (1993). *Statistics for Spatial Data, Revised Edition*. New York: John Wiley and Sons.

Gelfand, A. E. and Smith, A. F. M. (1990). "Sampling-based approaches to calculating marginal densities." *JASA*, 85: 398–409.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004). *Bayesian data analysis*. New York: Chapman and Hall/CRC.

Gelman, A., Mechelen, I. V., Verbeke, G., Heitjan, D. F., and Meulders, M. (2005). "Multiple imputation for model checking: Completed-data plots with missing and latent data." *Biometrics*, 61: 74–85.

Gneiting, T., Balabdaoui, F., and Raftery, A. E. (2007). "Probabilistic forecasts, calibration and sharpness." *JRSSB*, 69: 243–268.

Gschlößl, S. and Czado, C. (2007). "Spatial modelling of claim frequency and claim size in non-life insurance." *Scandinavian Actuarial Journal*, 107: 202–225.

O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., Oakley, J. E., and Rakow, T. (2006). *Uncertain Judgements: Eliciting Expert Probabilities*. Chichester: John Wiley and Sons.

O'Hagan, A. and Foster, J. (2003). *Kendall's Advanced Theory of Statistics, Bayesian Inference, Volume 2B*. London: Arnold, 2nd edition.

Ravines, R., Schmidt, A. M., Migon, H. S., and Rennó, C. D. (2008). "A joint model for rainfall-runoff: The case of Rio Grande Basin." *Journal of Hydrology*, 353: 189–200.

Roberts, G. O. and Rosenthal, J. S. (2001). "Optimal scaling for various Metropolis-Hastings algorithms." *Statistical Science*, 16: 351–367.

— (2006). "Examples of Adaptive MCMC." Technical report, University of Toronto.

Rue, H. and Follestad, T. (2003). "GMRFLib: a C-library for fast and exact simulation of Gaussian Markov random fields. Version 1.07." http://www.math./ntnu.no/∼hrue/GRMFLib.

Sansó, B., Schmidt, A. M., and Nobre, A. A. (2008). "Bayesian spatio-temporal models based on discrete convolutions." *Canadian Journal of Statistics*, 36: 239–258.

Santner, T. J., Craigmile, P. F., Calder, C. A., and Paul, R. (2008). "Demographic and behavioral modifiers of arsenic exposure pathways: a Bayesian hierarchical analysis of NHEXAS data." *Environmental Science & Technology*, 42: 5607–5614.

Schmidt, A. M., Hoeting, J., Pereira, J. B. M., and Vieira, P. P. (2009). "Mapping malaria in the Amazon rain forest: a spatio-temporal mixture model." In *The handbook of applied Bayesian Analysis. Edited by A. O'Hagan and M. West*, to appear. Oxford University Press.

Schmidt, A. M., Moreira, A. B. R., Helfand, S., and Fonseca, T. C. O. (2008). "Spatial sthocastic frontier models: accounting for unobserved local determinants of efficiency." *Journal of Productivity Analysis, DOI No. 10.1007/s11123-008-0122-6*.

Smith, B. J. (2005). "Bayesian Output Analysis Program (BOA), Version 1.1.5." http://www.public-health.uiowa.edu/boa.

Spiegelhalter, D., Thomas, A., and Best, N. (2002). "`WinBugs` Version 1.4 User Manual." http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml.

Acknowledgments

These comments were prepared while Alexandra M. Schmidt was visiting Marco A. Rodríguez at the Université du Québec à Trois-Rivières, Canada. She is grateful for access to journals

and insightful discussions. The author also thanks the Brazilian research agencies CNPq and FAPERJ for supporting her research projects.