The simplest case is as follows. A univariate count $y$ is related to $c$ via the density $\pi\{y;\theta_y,x(c)\}$ to a smooth latent scalar function $x(c)$ itself modelled as a stochastic process, with smoothness parameters $\theta_y$; in fact $y$ and $x(c)$ can be multivariate. One simple version leads to

$$\pi(c|y_{\text{new}},\text{data}) \approx k\,\pi(c)\int_{x(c)} \pi\{y_{\text{new}}|x(c);\theta\}\,\pi\{x(c)|\text{data};\theta\}\,\mathrm{d}x(c)$$

where $k$ is a normalizing constant. With Rue and his colleagues, we model $x(c)$ as an *a priori* Gaussian Markov random field on a finite lattice $C$; invoking the INLA, $x(c)$ is approximately Gaussian *a posteriori*, and, using the simplest of their approximations, $\pi\{x(c)|\text{data}\} \approx \pi_{\text{G}}(\hat{\theta})$, the parameter here denoting $(\hat{\mu},\hat{\tau})$ the mean and variance of $\pi_{\text{G}}(\cdot)$ evaluated at the mode. The evaluation of the integral by quadrature is fast and adequately accurate. Since $C$ is finite, we can normalize by evaluating the right-hand side for all $c \in C$.

It is possible to go further with INLA, for fast approximate cross-validation within the given data is now possible, i.e. we can evaluate $\pi(c|y_i,\text{data}_{-i})$ and compare with the known $c_i$ for every $i$. But, as we require for *each i* evaluation at *all c*, the updates in Section 6.3 no longer suffice. However, approximate fast updates $(\hat{\mu}_{-i},\hat{\tau}_{-i})$ are available, in the same spirit of the rank 1 constraint in their equation (8); see Salter-Townshend (2008) for details. A very powerful Bayesian tool is thus available without MCMC methods.

One limitation is the dimensionality of $C$. When $C$ is two dimensional on, for example, a 50-lattice, the `GMRFLib` routines are more than adequately fast. But even for a $30\times30\times30$ three-dimensional lattice we encounter problems. What might the authors recommend?

**Tom Heskes and Botond Cseke** (*Radboud University, Nijmegen*)

The authors are to be congratulated for a very interesting and stimulating paper. For the special case of sparse Gaussian processes with a small number of hyperparameters, the authors provide an automated procedure for approximate inference, producing very accurate results, which is orders of magnitude faster than Markov chain Monte Carlo methods.

We deeply appreciate the authors' efforts to relate their own approach to the deterministic approximations that have been developed in the machine learning literature. Following up on that, we shall attempt to shed some light on the link to expectation–propagation (EP) and discuss whether it could be used as an alternative to the Laplace approximation.

*Computing posterior marginals for* $\boldsymbol{\theta}$

The authors essentially apply numerical integration with the 'standard' Laplace approximation to evaluate the posterior marginal at the grid points, which they claim to be sufficiently accurate. This is somewhat contrary to the observations in Kuss and Rasmussen (2005) (see Fig. 5), where the marginal likelihood that was obtained by using the Laplace approximation seriously deviates from the truth. EP here does a much better job. See Minka (2001) for other comparisons between the Laplace approximation and EP and Seeger (2008) for an example of a Gaussian process model on which EP can still be applied but where Laplace approximation fails.

Slightly changing its standard implementation, EP can exploit the sparsity of the precision matrix in much the same way as the Laplace approximation. The computational bottleneck then also becomes the computation of the Cholesky decomposition and the marginal variances (see Section 2.1). We implemented EP for the stochastic volatility example in Section 5.3 and obtained results that were indistinguishable from a Laplace implementation with roughly the same computational complexity.

*Computing posterior marginals for the latent variables* $x_i$

For accurate approximations of the posterior marginals for the latent variables the authors have to go beyond the Gaussian approximation that they used for computing the posterior marginals for the hyperparameters. A full nested Laplace approximation is (way) too expensive, as would be a full nested EP approximation, and the authors introduce several clever tricks to obtain faster approximations thereof. Although it is not clear to us how generally applicable these approximations are, this appears to be the most important contribution of the paper. It is fair to say that (to the best of our knowledge) there are no deterministic approximations in the machine learning literature that even attempt to reach the same level of accuracy. A recent interpretation of EP as a series expansion (Opper *et al.*, 2009) may be turned into an alternative approach.

**Nils Lid Hjort** (*University of Oslo*) **and D. M. Titterington** (*University of Glasgow*)

The authors are to be congratulated for what promises to be a very influential contribution to practical Bayesian analysis. The methodology is very well thought out and the examples are convincing. Of course,