# What we talk about when we talk about fields

## Ewan Cameron

Department of Zoology, University of Oxford, Tinbergen Building, South Parks Road, Oxford, OX1 3PS, United Kingdom
email: `dr.ewan.cameron@gmail.com`
website: `astrostatistics.wordpress.com`

**Abstract.** In astronomical and cosmological studies one often wishes to infer some properties of an infinite-dimensional field indexed within a finite-dimensional metric space given only a finite collection of noisy observational data. Bayesian inference offers an increasingly-popular strategy to overcome the inherent ill-posedness of this signal reconstruction challenge. However, there remains a great deal of confusion within the astronomical community regarding the appropriate mathematical devices for framing such analyses and the diversity of available computational procedures for recovering posterior functionals. In this brief research note I will attempt to clarify both these issues from an "applied statistics" perpective, with insights garnered from my post-astronomy experiences as a computational Bayesian / epidemiological geostatistician.

**Keywords.** Methods: data analysis–methods: statistical

## 1. Introduction

The potential afforded by Bayesian techniques for inferring the properties of infinite-dimensional mathematical structures, such as random fields (to be understood here as random functions defined at each point of some finite-dimensional metric space), has long been recognised by both probability theorists, e.g. O'Hagan (1978), and practitioners: with the first wave of practical applications in geoscience (e.g. Omre (1987), Handcock & Stein (1993)) and machine learning (e.g. Rasmussen & Williams (1996), Neal (1997)) contemporaneous with the advent of sufficiently powerful desktop computers. Cosmologists were at this time notable as *'early adopters'* and pioneers of the new techniques for field inference. Indeed, the Monte Carlo methods for constrained simulation from Gaussian random fields developed by Bertschinger (1987) and Hoffman & Ribak (1991) remain key tools for efficient conditional simulation, cf. Doucet (2010).

However, over the past decade the sophistication of statistical analysis techniques brought to bear on the study of cosmological fields has not kept pace with progress outside of astronomy. With modern tools such as the Integrated Nested Laplace Approximation (INLA; Rue *et al.* (2009)), 'variational inference' (Hensman *et al.* (2013)), particle filtering (Del Moral *et al.* (2007)), and Approximate Bayesian Computation (ABC; Marjoram *et al.* (2003)) almost entirely ignored to-date by the cosmological community we have, in my opinion, become the *'laggards'* of the technology adoption lifecycle.

There are multiple factors seemingly to blame for this divergence: (i) the emergence of an isolationist attitude to the practice of cosmological statistics; (ii) an over-emphasis on the path integration-based conceptulisation of random fields, rather than the measure-theory-based mathematics of mainstream statistics; and (iii) an under-appreciation of the potential for stochastic process priors (including, *but not limited to*, the Gaussian process) as flexible modelling components within the hierarchical Bayesian framework. With the first already being well fought back against by inter-disciplinary programming

in conferences such as the IAUS306 and the SCMA series I will therefore focus in this proceedings (as in my contributed talk) on the latter two. In particular, I aim to clarify a number of mathematical concepts crucial to a high-level understanding of Bayesian inference over random fields and measures (Section 2), and then to highlight just a few of the exciting techniques to have recently emerged in this area (Section 3).

## 2. The Mathematics of Bayesian Field Inference

Cosmologists and astronomers already well-versed in the practice and theory of Bayesian statistics in the finite-dimensional setting will typically have one of two contrasting experiences upon first attempting to extend these ideas to infinite-dimensional inference problems. The pragmatist will happily observe that the mechanics of computation are little changed (e.g. the Gaussian random field at finite sample points is distributed just as the familiar multivariate Gaussian), while the cautious theorist will more likely be overwhelmed by a first acquaintance with measure-theoretic probability (i.e., probability triples and the algebra of sets). But ideally one will have both experiences, since each offers an equally important perspective, as I discuss in this Section.

### 2.1. *Distributions over Infinite-Dimensional Space*

In formal statistics the core of probabilistic computation is framed within the language of measure theory: the key object being the *'probability triple'* of (i) a sample space, $\Omega$, i.e., some non-empty set; (ii) a $\sigma$-algebra, $\Sigma$, i.e., a collection of subsets of $\Omega$ with $\emptyset, \Omega \in \Sigma$, closed under the formation of complements, countable unions and intersections; and (iii) a probability measure, $P$, i.e., a countably additive set function from $\Sigma$ to [0,1] for which $P(\emptyset) = 0$ and $P(\Omega) = 1$. In this context Carathéodory's Extension Theorem provides the theorist with the machinary to build complex probability triples and forge a rigorous notion of random variables as measures on the pre-images of Borel $\sigma$-algebra sets of the real numbers; and from this to the familiar mechanics of probability densities defined with respect to the Lebesgue measure (e.g. the standard Normal with $f(x) = \frac{1}{\sqrt{2\pi}} \exp -\frac{x^2}{2} dx$). Nevertheless, with the Lebesgue measure behaving intuitively as a product measure in $\mathcal{R}^n$, and with Lebesgue and Riemann integration interchangable in practice for all but a few rare cases, the pragmatist can safely ignore these theoretical foundations in the study of 'real-world' problems in finite-dimensional settings.

In the context of probabilistic inference over *fields*, however, one must proceed with care as there exists no equivalent to the Lebesgue measure to serve as a natural reference for defining densities in an infinite-dimensional Banach space (e.g. the $L^p$ function spaces). Hence, for Bayesian analysis in infinite-dimensional space we must be deliberate in our choice of reference measure, which we encode into the prior. Typically we will do this indirectly by assigning as prior the implicit measure (or 'law') belonging to a given *stochastic process* (e.g. a Gaussian process, or Poisson process) having sample paths within the field space under study. Although quite technical the distinction between this formal statistical approach and the path integration-based language of cosmological papers, e.g. Enßlin *et al.* (2009), Kitching & Taylor (2011), is important if we are to connect with, *and thereby benefit from*, the rich body of applied statistics literature on infinite-dimensional inference. Worth noting also is that the measure-theoretic equivalent of the probability density is the 'Radon-Nikodym (R-N) derivative', with a trivial but illustrative example being that of the R-N derivative of posterior against prior given by the likelihood function divided by the marginal likelihood, c.f. Cotter *et al.* (2009).

Finally, the measure theoretic definition of a stochastic process is a collection of random variables *indexed* by a set; here all points of the physical metric space over which

the field problem is to be studied. A key theoretical tool for the construction of stochastic processes, which is greatly illustrative of their behaviour, is Kolmogorov's Extension Theorem. The theorem gives conditions under which a rule for assigning the distributions of the finite-dimensional projections of an infinite-dimensional indexing set can be considered sufficient to define a proper stochastic process. The key condition here is one of *mutual consistency* of 'coarser-binned' finite-dimensional projected distributions with respect to 'finer-binned' ones; this idea was well-understood by the MaxEnt pioneers, cf. Skilling (1998): "the prior must depend ... on the pixel size $h$ in such a way that subsidiary pixelisation is immaterial".

## 2.2. *Hierarchical Bayesian Models with Stochastic Process Priors*

As mentioned earlier for the pragmatist the mechanics of Bayesian field analysis need differ little from those of finite-dimensional inference; especially when we are able to write our prior–likelihood pairing via a stochastic process embedded within a *hierarchical Bayesian model*. A typical hierarchical model for Bayesian field inference is the following from Gething *et al.* (2010) in the context of epidemiological geostatistics,

$$
\begin{aligned}
N_i^+ &\sim \text{Bin}(N_i, p(x)) \\
p(x) &= g^{-1}(f(x)) \\
f|\phi &\sim \text{GP}(M_\phi, C_\phi).
\end{aligned}
$$

We can read this model from the top down to understand the generative process for the data: binomial sampling from a population of $N_i$ at each site with underlying prevalence (probability), $p(x)$, depending on the location of the site, $x$. The prevalence field, $p(x)$, is in turn the realisation of a Gaussian random field, GP, transformed to the range, (0,1), by the *link function*, $g^{-1}(\cdot)$. A prior on the parameters of the mean and/or correlation function of the GP would be a natural extension of this particular model; while natural extensions to other problems using the GP could include, e.g., exponentation of $f$ via $g^{-1}(\cdot)$, use of a Poisson process likelihood function, or some penalisation of the GP sample paths. An even greater diversity of hierarchical forms can then be built with the addition of *non*-Gaussian processes: cf. systematic error analysis (Burr & Doss (2005)) and *non*-proportional hazards modelling (De Iorio *et al.* (2009)) with Dirichlet processes, online Bayesian classification with Mondrian processes (Lakshminarayanan *et al.* (2013)), and so on (Poisson processes, Gamma processes, Negative binomial processes, etc.).

For complicated real-world problems there is rarely an analytical solution to the resulting posterior so, as in ordinary Bayesian analysis, one will almost always turn to computational methods. Some of these are well-known to cosmologists already, such as Gibbs sampling (e.g. Wandelt *et al.* (2004)) and Hamiltonian Monte Carlo (e.g. Hajian (2007), Jasche & Kitaura (2010)); however, a much greater number of specialised techniques for Bayesian field inference remain largely undiscovered. I highlight just a few of these briefly in the following Section.

## 3. Some Computational Techniques for Bayesian Field Inference

From a geostatistical perspective it is difficult to overstate the revolutionary impact lately effected by the emergence of INLA (Rue *et al.* (2009)) and the stochastic partial differential equation (SPDE) approach to Gaussian processes (Lindgren *et al.* (2011)). Using the machinery of the finite-element method already familiar to astronomers, the SPDE approach aims to identify discretely-indexed Gaussian *Markov* random fields providing weak approximations to their continuously-indexed counterparts. The result is a

system for approximate Bayesian inference over random fields amenable to *fast* computation via sparse matrix operations; whereas Hamiltonian MCMC over field posteriors is typically the reserve of cluster-computing, the INLA method enables approximate field inference on desktop computers with run-times small enough to allow for the important follow-up inference steps of model testing and refinement. A first glimpse of the potential for INLA in cosmological applications was provided in a 2010 study of CMB reconstruction by Wilson & Yoon (2010); but as yet this pioneering work remains unappreciated (or, at least, uncited).

Another seminal technique for 'Big Data' Bayesian inference over random fields is that of 'variational inference' (Hensman *et al.* (2013)), in which approximation to the full field posterior is made using sets of data-driven 'inducing points'. Ongoing research efforts into the application of Approximate Bayesian Computation (Soubeyrand *et al.* (2013)) and Kalman filtering (Särkkä *et al.* (2014)) in the spatial domain are also well worth keeping an eye on.

## References

Bertschinger, E. 1987, *ApJ*, 323, L103–L106

Burr, D. & Doss, H. 2005, *J. Am. Statist. Assoc.*, 100, 242–251

Cotter, S. L., Dashti, M., Robinson, J. C., & Stuart, A. M. 2009, *Inverse Probl.*, 25, 115008

De Iorio, M., Johnson, W. O., Müller, P., & Rosner, G. L. 2009, *Biometrics*, 65, 762–771

Del Moral, P., Doucet, A., & Jasra, A. 2007, 'Sequential Monte Carlo for Bayesian Computation' in *Bayesian Statistics 8*, J.M. Bernardo, M.J. Bayarri, J.O. Berger, A.P. Dawid, A.F.M. Smith & M. West, eds., OUP, 1–34

Doucet, A. 2010, 'A Note on Efficient Conditional Simulation of Gaussian Distributions', Departments of Computer Science and Statistics, University of British Columbia

Enßlin, T. A., Frommert, M., & Kitaura, F. S. 2009, *Phys. Rev. D*, 80, 105005

Handcock, M. S. & Stein, M. L. 1993, *Technometrics*, 35, 403–410

Gething, P. W., Patil, A. P., & Hay, S. I. 2010, *PLOS Computat. Biol.*, 6, e1000724

Hajian, A. 2007, *Phys. Rev. D.*, 75, 083525

Hensman, J., Fusi, N., & Lawrence, N. D. 2013, 'Gaussian Processes for Big Data' in *Association for Uncertainty in Artificial Intelligence*, UAI2013, 244

Hoffman, Y. & Ribak, E. 1991, *ApJ*, 380, L5–L8

Jasche, J. & Kitaura, F. S. 2010, *MNRAS*, 407, 29–42

Kitching, T. D. & Taylor, A. N. 2011, *MNRAS*, 410, 1677–1686

Lakshminarayanan, B., Roy, D. M., & Teh, Y. W. 2013, 'Top-down Particle Filtering for Bayesian Decision Trees' in *International Conference on Machine Learning*, ICML2013

Lingren, F., Rue, H., & Lindström, J. 2011, *J. R. Statist. Soc. B*, 73, 423–498

Marjoram, P., Molitor, J., Plagnol, V., & Tavaré, S. 2003, *PNAS*, 100, 15324–15328

Neal, R. M. 1997, 'Monte Carlo Implementation of Gaussian Process Models for Bayesian Regression and Classification', *Tech. Rep. 9702*, Department of Statistics, University of Toronto

O'Hagan, A. 1978, *J. R. Statist. Soc. B*, 40, 1–42

Omre, H. 1987, *Math. Geol.*, 19, 25–39

Rasmussen, C. E. & Williams, C. K. I. 1996, 'Gaussian Processes for Regression' in *Advances in Neural Information Processing Systems 8*, eds. D.S. Touretzky, M.C. Mozer, M.E. Hasselmo, MIT Press, 514–520

Rue, Y., Martino, S., & Chopin, N. 2009, *J. R. Statist. Soc. B*, 71, 319–392

Särkkä, S., Solin, A., & Hartikainen, J. 2014, 'Spatio-Temporal Learning via Infinite-Dimensional Bayesian Filtering and Smoothing', to appear in The IEEE Signal Processing Magazine

Skilling, J. 1998, *J. Microscop.*, 190, 28–36

Soubeyrand, S., Carpentier, F., Guiton, F., & Klein, E. K. 2013, *Stat. Appl. Genet. Mol. Biol.*, 12, 17–37

Wandelt, B. D., Larson, D. L., & Lakshminarayanan, A. 2009, *Phys. Rev. D*, 70, 083511

Wilson, S. P. & Yoon, J. 2010, preprint(arXiv:1011.4018)