Discussion of paper "nonparametric Bayesian inference in applications" by Peter Müller, Fernando A. Quintana and Garritt L. Page

Athanasios Kottas

Statistical Methods & Applications
Journal of the Italian Statistical Society

ISSN 1618-2510
Volume 27
Number 2

Stat Methods Appl (2018) 27:219-225
DOI 10.1007/s10260-017-0398-7

Springer

Stat Methods Appl (2018) 27:219–225
https://doi.org/10.1007/s10260-017-0398-7

ORIGINAL PAPER

Discussion of paper "nonparametric Bayesian inference in applications" by Peter Müller, Fernando A. Quintana and Garritt L. Page

Athanasios Kottas[1]

Accepted: 27 August 2017 / Published online: 4 September 2017

Abstract This is an invited discussion of review paper "Nonparametric Bayesian Inference in Applications" by Peter Müller, Fernando A. Quintana and Garritt L. Page.

Keywords Bayesian nonparametrics · Dirichlet process mixtures · Integro-difference equation models · Spatial Poisson processes · Spatio-temporal modeling

We commend the authors for an interesting review of applications of nonparametric Bayesian (NPB) modeling. The authors have made fundamental contributions to the methodology and applications of Bayesian nonparametrics, including the application areas discussed in this paper, and it is always stimulating to read on their perspective on NPB modeling and inference. In this discussion of the paper, we provide additional details on Bayesian nonparametric modeling methods for spatial data analysis, the area on which the authors focus for more detailed discussion and literature review. One means to organize discussion of NPB spatial models revolves around the different data structures: point-referenced data, areal unit data, and spatial point patterns (see, e.g., Banerjee et al. 2015). In all three cases, the data structure is typically enriched in applications with spatial data recorded over time, requiring appropriate elaborations of the spatial models. We have recently reviewed NPB methods for disease mapping,

The author wishes to thank Igor Prünster and the editors Tommaso Proietti and Ruggero Bellio for the invitation to write this discussion. This work was supported in part by the National Science Foundation under award SES-1631963.

✉ Athanasios Kottas
thanos@soe.ucsc.edu

[1] Department of Applied Mathematics and Statistics, University of California, Santa Cruz, CA 95064, USA

one of the main areas of application involving areal unit data (Kottas 2016). Here, we provide some additional discussion and references that complement the paper in the direction of spatio-temporal modeling for point-referenced data. Moreover, the discussion of NPB modeling for spatial point processes contributes an additional component to the review.

Spatio-temporal modeling for point-referenced data

The authors have provided a detailed review of the modeling aspects of various NPB methods for spatial point-referenced data. In many applications, arising, for instance, in ecology, epidemiology and environmental sciences, data is collected over both space and time. For discrete-time settings, it is possible to incorporate spatial NPB priors into a dynamic spatial process model. For instance, Kottas et al. (2008) use the spatial DP prior (Gelfand et al. 2005) for the spatial process that generates the innovations, $\eta_t(\boldsymbol{s})$, of the transition equations for spatial random effects, $\theta_t(\boldsymbol{s}) = \rho\,\theta_{t-1}(\boldsymbol{s}) + \eta_t(\boldsymbol{s})$. However, general modeling methods for spatio-temporal data seek to capture the interactions between the spatial and temporal components. Cressie and Wikle (2011) provide a systematic treatment of spatio-temporal models. Here, we focus on integro-difference equation (IDE) models, a particular class of dynamical spatio-temporal models, and discuss recent work that incorporates NPB components to enhance the inferential scope of the IDE modeling framework.

Denote by $X_t(\boldsymbol{s})$ the process at spatial location $\boldsymbol{s}$ and time t, and by $Y_t(\boldsymbol{s})$ the associated response. In the simplest form, the data observations can be related to the process through $Y_t(\boldsymbol{s}) = X_t(\boldsymbol{s}) + \varepsilon_t(\boldsymbol{s})$, where $\varepsilon_t(\boldsymbol{s})$ is (independent) observation noise. (For simpler notation, we ignore covariate information which may be incorporated at the observation level of the model.) Under an IDE model formulation, the evolution of the process is specified as follows:

$$X_t(\boldsymbol{s}) = \int k(\boldsymbol{u} - \boldsymbol{s} \mid \theta) X_{t-1}(\boldsymbol{u})\, \mathrm{d}\boldsymbol{u} \; + \; \omega_t(\boldsymbol{s})$$

where $\omega_t(\boldsymbol{s})$ is a zero mean, possibly spatially correlated, error process, and k is a redistribution kernel with parameters θ, typically assumed to belong to a location family of distributions (as the notation above suggests).

The concept of an IDE model is that the process at time $t-1$ is propagated in time by re-weighting it using the kernel k, which acts in space. This model structure provides a simple, yet effective and interpretable way to express the relationship between the space and time components, using linear dynamics. As the kernel shifts or expands, the nature of the dependence between the process at a specific location and the process at previous time points changes accordingly. In fact, the first two moments of the IDE kernel distribution determine, respectively, the advection and diffusion of the process (Brown et al. 2000; Storvik et al. 2002). The IDE model can be made more flexible through spatially dependent kernel parameters, $\theta_{\boldsymbol{s}}$ (e.g., Xu et al. 2005). In particular, the covariance function of an IDE process $X_t(\boldsymbol{s})$, with a stationary initial process $X_0(\boldsymbol{s})$, is non-stationary in space for all $t > 0$, when the IDE kernel parameters depend on spatial location (Richardson et al. 2016).

In essentially all applications of IDE models, Gaussian kernels are used, owing to the convenience in their specification and computation. In Richardson et al. (2017a), it is shown that IDEs with non-Gaussian kernels allow for the description of higher order properties in the spatio-temporal process, and, in particular, the skewness and tail behavior of the IDE kernel can be associated with such properties. Thus, although complex spatial fields will often exhibit non-linear dynamics, non-Gaussian kernels can help explain some of the complexity. This feature can be further enhanced by extending the IDE kernel parameters to vary across the spatial region. From a NPB modeling perspective, such considerations naturally motivate exploring nonparametric prior models for the IDE kernel.

Richardson et al. (2016) studied nonparametric mixture prior models for the IDE kernel, in the special case of one-dimensional space (therefore, the location s is scalar). More specifically, the kernel density was modeled with a Gaussian mixture, $k(u - s \mid \sigma^2, G_s) = \int \mathrm{N}(u - s \mid \mu, \sigma^2)\, \mathrm{d}G_s(\mu)$, with a location-dependent mixing distribution G_s assigned a spatial DP prior (with a baseline Gaussian process defined on one-dimensional space). The reason for the restriction to one-dimensional space involves the computational challenges in implementation of the model for full posterior inference. IDE models are typically applied to data using a common orthonormal basis representation for the process $X_t(\boldsymbol{s})$ and the IDE kernel k (e.g., Wikle 2002). This representation results in a dynamic linear model for the data vectors collecting all responses across space at each time point, albeit with the IDE kernel parameters entering the distribution for the corresponding state vectors in a highly non-linear fashion. To facilitate implementation, Richardson et al. (2016) used Hermite polynomials (which are orthogonal with respect to the weight function $\exp(-x^2)$) to define a basis representation that is more compatible with the Gaussian densities driving the mixture model for the IDE kernel, as well as Hamiltonian MCMC steps in the posterior simulation method. Nevertheless, the extension to two-dimensional space remains challenging for the spatial DP mixture IDE kernel model. The univariate case is not without merit, as there are environmental applications where variables are recorded across time by altitude, length or depth. In particular, the data example of Richardson et al. (2016) involves ozone pressure data obtained by releasing a balloon in the air that collects ozone pressure measurements at certain intervals throughout its flight. For this data example, the location-dependent DP mixture model was shown to outperform in prediction stationary IDE models with different parametric kernels, as well as the non-stationary Gaussian IDE kernel model with location-dependent mean and variance parameters.

To achieve balance between model flexibility and computational feasibility for a two-dimensional space, Richardson et al. (2017b) propose a semiparametric approach. The IDE kernel is modeled with the bivariate stable distribution, which is defined through a location parameter vector $\boldsymbol{\mu} = (\mu_1, \mu_2)$, a scalar parameter $\alpha \in (0, 2]$ that controls tail behavior, and a finite measure Γ on $[0, 2\pi]$ that controls skewness, orientation and spread. Note that the basis representation for the IDE kernel requires only the characteristic function of the kernel distribution, and thus the lack of a closed-form expression for the bivariate stable density does not pose any difficulties in this context.

A nonparametric prior model is used for $d\Gamma$, scaled by a parameter $c > 0$, building on ideas from Bernstein polynomial priors for densities with compact support (Petrone 1999). In particular, $d\Gamma$ is represented as a weighted combination of Beta densities with fixed parameters producing density shapes that span $[0, 2\pi]$. The weights are defined through increments of a distribution function on $[0, 2\pi]$, which is assigned a geometric weights prior (Mena et al. 2011). Kernel process convolutions (Higdon 1998) are used to model spatially dependent parameters $\mu_1(\boldsymbol{s})$, $\mu_2(\boldsymbol{s})$, $\log(c(\boldsymbol{s}))$ and $\Phi^{-1}(\eta(\boldsymbol{s}))$, where η is the random variable on $(0, 1)$ that defines the geometric weights. Using data on sea surface temperature in the tropical Pacific Ocean, it is shown that the semiparametric model has better predictive performance than the state-of-the-art Gaussian IDE kernel model with spatially varying parameters.

Bayesian nonparametric methods for spatial Poisson processes

For spatial point patterns, observed in a bounded region $\mathcal{D} \subset \mathbb{R}^k$ (with $k \geq 2$), the number of points and their locations are random. Moreover, many applications involve marks, a set of variables $\boldsymbol{y}_i$ associated with each random location $\boldsymbol{s}_i \in \mathcal{D}$; the mark $\boldsymbol{y}_i$ takes values in mark space $\mathcal{M}$ and it may comprise both categorical and continuous variables. For model-based inference, the point pattern is assumed to arise from a (marked) spatial point process and the interest lies in modeling and inference for functionals characterizing the point process. Here, we focus on spatial non-homogeneous Poisson processes (NHPPs), the most tractable class of point processes, as well as the one that has received more attention in the Bayesian nonparametrics literature.

A spatial NHPP can be defined through its intensity function, $\lambda(\boldsymbol{s})$, for $\boldsymbol{s} \in \mathcal{D}$, a non-negative and locally integrable function for all $B \subseteq \mathcal{D}$, such that: for any such B, the number of points in B, $\mathcal{N}(B)$, follows a Poisson distribution with mean $\int_B \lambda(\boldsymbol{u})\mathrm{d}\boldsymbol{u}$; and given $\mathcal{N}(B)$, the point locations within B are i.i.d. arising according to density $\lambda(\boldsymbol{s})/\{\int_B \lambda(\boldsymbol{u})\mathrm{d}\boldsymbol{u}\}$. Using the *Marking Theorem* (e.g., Møller and Waagepetersen 2004), one can conceptualize a marked Poisson process as a NHPP on the joint location-mark space $\mathcal{D} \times \mathcal{M}$ with joint intensity $\mu(\boldsymbol{s}, \boldsymbol{y})$, provided the marginal intensity $\int_{\mathcal{M}} \mu(\boldsymbol{s}, \boldsymbol{y})\mathrm{d}\boldsymbol{y} = \lambda(\boldsymbol{s})$ is locally integrable.

One of the earlier approaches to NPB inference for spatial NHPP intensities can be found in Heikkinen and Arjas (1998) where piecewise constant functions, driven by Voronoi tesselations and Markov random field priors, were used to model the intensity surface. The approach was extended in Heikkinen and Arjas (1999) to a model for spatial point patterns influenced by concomitant variables. Another approach to (approximate) Bayesian inference for spatial NHPPs is based on log-Gaussian Cox process models, where the NHPP intensity function is modeled on the logarithmic scale through a Gaussian process (e.g., Møller et al. 1998; Brix and Diggle 2001; Brix and Møller 2001). Related work can be found in Liang et al. (2009) where a Bayesian hierarchical model for marked NHPPs is developed, using an extension of the log-Gaussian Cox process to accommodate different types of covariate information. These approaches require relatively complex computational schemes for posterior inference, including reversible jump Markov chain Monte Carlo methods and Metropolis-adjusted Langevin algorithms.

A different direction involves mixture representations for the NHPP intensity function based on convolutions of a non-negative kernel (which is not necessarily a density) with a weighted gamma process (Ishwaran and James 2004) or with a general Lévy random field (Wolpert and Ickstadt 1998). Extensions of the gamma process convolution model to regression settings are considered in Ickstadt and Wolpert (1999) and Best et al. (2000). The approach of Wolpert and Ickstadt (1998) is extended by Kang et al. (2014) to model hierarchically a collection of related spatial point patterns.

It is also possible to utilize more traditional nonparametric mixtures to model NHPP intensity functions by casting the inference questions in a density estimation framework. The key observation is that the intensity function can be normalized to a density: $f(\boldsymbol{s}) = \lambda(\boldsymbol{s})/\gamma$, for $\boldsymbol{s} \in \mathcal{D}$, where $\gamma = \int_{\mathcal{D}} \lambda(\boldsymbol{u}) \mathrm{d}\boldsymbol{u} < \infty$ is the total integrated intensity over $\mathcal{D}$. The NHPP definition implies that the likelihood for an observed point pattern $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N\} \subset \mathcal{D}$ can be expressed as $\gamma^N \exp(-\gamma) \prod_{i=1}^{N} f(\boldsymbol{s}_i)$, such that inference for the NHPP intensity amounts to density estimation for the NHPP density $f(\boldsymbol{s})$. Moreover, since γ only scales the intensity function, a flexible nonparametric prior model for the NHPP density (along with a prior for γ) suffices to capture general shapes for the intensity surface. This method was developed for spatial NHPPs (Kottas and Sansó 2007) and temporal NHPPs (Kottas and Behseta 2010), using DP mixtures of Beta densities to model the NHPP density, with further choices for the DP mixture kernel considered in Taddy and Kottas (2012). The method was adapted for indirectly observed spatial point patterns by Ji et al. (2009), with an application to analysis of immunological studies.

Casting the modeling for NHPPs in a density estimation setting has practical benefits, with respect to posterior simulation which draws from well established techniques for nonparametric mixture models, as well as methodological advantages, with respect to extensions to modeling marked NHPPs and to inference under hierarchical settings. Taddy and Kottas (2012) build a modeling approach for marked NHPPs from DP mixtures for the point process density over the joint location-mark space, $f(\boldsymbol{s}, \boldsymbol{y}) = \mu(\boldsymbol{s}, \boldsymbol{y})/\gamma$, with $(\boldsymbol{s}, \boldsymbol{y}) \in \mathcal{D} \times \mathcal{M}$. Here, the integrated intensity can be defined in terms of either the joint or marginal process, $\gamma = \int_{\mathcal{D}} \int_{\mathcal{M}} \mu(\boldsymbol{s}, \boldsymbol{y}) \,\mathrm{d}\boldsymbol{y}\mathrm{d}\boldsymbol{s} = \int_{\mathcal{D}} \lambda(\boldsymbol{s}) \mathrm{d}\boldsymbol{s}$. This approach can handle multivariate marks and it yields fully nonparametric inference for the marginal point process intensity, $\lambda(\boldsymbol{s}) = \gamma f(\boldsymbol{s})$, and for the conditional mark density, $h(\boldsymbol{y} \mid \boldsymbol{s}) = f(\boldsymbol{s}, \boldsymbol{y})/f(\boldsymbol{s})$. Since the prior model for the marked NHPP is built from a DP mixture, it enables hierarchical extensions through dependent DP priors (MacEachern 2000) for the mixing distributions. A practically useful extension involves dynamic modeling of marked point patterns recorded over discrete time. Two relevant applications consist of dynamic estimation of violent crime intensity surfaces (Taddy 2010) and of seasonal hurricane intensities (Xiao et al. 2015).

References

Banerjee S, Carlin BP, Gelfand AE (2015) Hierarchical modeling and analysis for spatial data, 2nd edn. Chapman & Hall/CRC, London

Best NG, Ickstadt K, Wolpert RL (2000) Spatial Poisson regression for health and exposure data measured at disparate resolutions. J Am Stat Assoc 95:1076–1088

Brix A, Diggle PJ (2001) Spatiotemporal prediction for log-Gaussian Cox processes. J R Stat Soc B 63:823–841

Brix A, Møller J (2001) Space-time multi type log Gaussian Cox processes with a view to modeling weeds. Scand J Stat 28:471–488

Brown PE, Roberts GO, Kåresen KF, Tonellato S (2000) Blur-generated non-separable space-time models. J R Stat Soc B 62:847–860

Cressie N, Wikle CK (2011) Statistics for spatio-temporal data. Wiley, New York

Gelfand AE, Kottas A, MacEachern S (2005) Bayesian nonparametric spatial modeling with Dirichlet process mixing. J Am Stat Assoc 100:1021–1035

Heikkinen J, Arjas E (1998) Non-parametric Bayesian estimation of a spatial poisson intensity. Scand J Stat 25:435–450

Heikkinen J, Arjas E (1999) Modeling a poisson forest in variable elevations: a nonparametric Bayesian approach. Biometrics 55:738–745

Higdon D (1998) A process-convolution approach to modelling temperatures in the North Atlantic Ocean. Environ Ecol Stat 5:173–190

Ickstadt K, Wolpert RL (1999) Spatial regression for marked point processes. In: Bernardo JM, Berger JO, Dawid P, Smith AFM (eds) Bayesian statistics, vol 6. Oxford University Press, Oxford, pp 323–341

Ishwaran H, James LF (2004) Computational methods for multiplicative intensity models using weighted gamma processes: proportional hazards, marked point processes, and panel count data. J Am Stat Assoc 99:175–190

Ji C, Merl D, Kepler TB, West M (2009) Spatial mixture modelling for unobserved point processes: examples in immunofluorescence histology. Bayesian Anal 4:297–315

Kang J, Nichols TE, Wager TD, Johnson TD (2014) A Bayesian hierarchical spatial point process model for multi-type neuroimaging meta-analysis. Ann Appl Stat 8:1800–1824

Kottas A (2016) Bayesian nonparametric modeling for disease incidence data. In: Lawson AB, Banerjee S, Haining RP, Ugarte MD (eds) Handbook of spatial epidemiology. Chapman and Hall/CRC, London, pp 363–374

Kottas A, Behseta S (2010) Bayesian nonparametric modeling for comparison of single-neuron firing intensities. Biometrics 66:277–286

Kottas A, Sansó B (2007) Bayesian mixture modeling for spatial Poisson process intensities, with applications to extreme value analysis. J Stat Plan Inference 137:3151–3163

Kottas A, Duan J, Gelfand AE (2008) Modeling disease incidence data with spatial and spatio-temporal dirichlet process mixtures. Biomet J 50:29–42

Liang S, Carlin BP, Gelfand AE (2009) Analysis of Minnesota colon and rectum cancer point patterns with spatial and nonspatial covariate information. Ann Appl Stat 3:943–962

MacEachern S (2000) Dependent dirichlet processes. Technical report. Department of Statistics, Ohio State University

Mena RH, Ruggiero M, Walker SG (2011) Geometric stick-breaking processes for continuous-time Bayesian nonparametric modeling. J Stat Plan Inference 141:3217–3230

Møller J, Waagepetersen RP (2004) Statistical inference and simulation for spatial point processes. Chapman & Hall/CRC, London

Møller J, Syversveen AR, Waagepetersen RP (1998) Log Gaussian Cox processes. Scand J Stat 25:451–482

Petrone S (1999) Bayesian density estimation using Bernstein polynomials. Can J Stat 27:105–126

Richardson R, Kottas A, Sansó B (2016) Bayesian non-parametric modeling for integro-difference equations. Stat Comput. doi:10.1007/s11222-016-9719-1

Richardson R, Kottas A, Sansó B (2017a) Flexible integro-difference equation modeling for spatio-temporal data. Comput Stat Data Anal 109:182–198

Richardson R, Kottas A, Sansó B (2017b) Spatio-temporal modelling using integro-difference equations with bivariate stable kernels. Technical report, Baskin School of Engineering, University of California, Santa Cruz

Storvik G, Frigessi A, Hirst D (2002) Stationary space-time gaussian fields and their time autoregressive representation. Stat Model 2:139–161

Taddy M (2010) Autoregressive mixture models for dynamic spatial Poisson processes: application to tracking the intensity of violent crime. J Am Stat Assoc 105:1403–1417

Taddy MA, Kottas A (2012) Mixture modeling for marked Poisson processes. Bayesian Anal 7:335–362

Wikle CK (2002) A kernel-based spectral model for non-Gaussian spatio-temporal processes. Stat Model 2:299–314

Wolpert RL, Ickstadt K (1998) Poisson/gamma random field models for spatial statistics. Biometrika 85:251–267

Xiao S, Kottas A, Sansó B (2015) Modeling for seasonal marked point processes: an analysis of evolving hurricane occurrences. Ann Appl Stat 9:353–382

Xu K, Wikle CK, Fox NI (2005) A kernel-based spatio-temporal dynamical model for nowcasting weather radar reflectivities. J Am Stat Assoc 100:1133–1144