Bayesian Analysis (2016) **11**, Number 1, pp. 303–305

Comment on Article by Page and Quintana*

Brian J. Reich† and Montserrat Fuentes‡

Congratulations to the authors on this nice contribution! The proposed model fuses ideas from spatial methods and Bayesian nonparametrics to produce a new approach for heterogeneous areal and geostatistical data. The paper is quite thorough, with extensive study of the model's properties, simulation studies to evaluate performance in different settings, and multiple worked examples. It is welcomed addition to the spatial literature.

A key argument is that connected partitions are too restrictive. As the authors acknowledge, disconnected partitions such as those considered in the paper prohibit continuous (even locally) spatial processes. We tend to agree with the authors that most environmental data are either measured with spatially-uncorrelated errors or include spatial variation at a scale smaller than can be captured in a finite sample, and so discontinuous models are often a reasonable choice. However, there are counterexamples where a continuous process maybe preferred, such as numerical climate model output. For these cases, it would be interesting to compare the proposed method the spatial tessellation model of Kim et al. (2005) (or perhaps the similar construction in Gramacy (2007) and Gramacy and Lee (2008) for computer model output). On a related note, Table 4 shows that the proposed method dramatically outperforms the usual spatial regression model (SR) when data are generated with a single cluster. This is surprising because SR is the true model in this case with cluster mean μ_1^* as the overall mean.

Selecting the number of clusters is a crucial step in classical clustering analysis. Similar to the non-spatial Dirichlet process, the authors avoid fixing the number of clusters by assuming the number of clusters in the population is infinite and studying the prior and posterior effective number of clusters that partition the n data points (k_n). The Bayesian model then averages over the effective number of clusters. However, this flexibility comes at the price of having to specify the prior over the number of clusters, which in this formulation boils down to the hyperparameters M and α. The authors examine sensitivity to the choice of M both in the prior and posterior, and find that for prediction the results are slightly sensitive to M; likely inference on the number and location of clusters is highly sensitive to M.

An alternative to fixing M is to treat it as an unknown parameter to be estimated in the Bayesian model. It is not clear how to implement this though because the prior for the partition in (2) contains an unspecified normalizing constant that is a complicated function of M. One approach to estimating M is to use a truncated stick-breaking representation of the DP. If we assume there are at most H clusters in

*Main article DOI: 10.1214/15-BA971.

†Department of Statistics, North Carlina State University, Campus Box 8203, Raleigh, NC 276945, bjreich@ncsu.edu

‡Department of Statistics, North Carlina State University, Campus Box 8203, Raleigh, NC 276945, fuentes@ncsu.edu

 DOI: 10.1214/15-BA971B

the population, the proposed model (3) permits the hierarchical representation (Park and Dunson, 2010) $\mathbf{s}_i|c_i = h \sim q(\mathbf{s}_h^*, \epsilon)$ and $\text{Prob}(c_i = h) = V_h \prod_{l<h}(1 - V_l)$, where $V_1, ..., V_{H-1} \overset{iid}{\sim} \text{Beta}(1, M)$ and $V_H = 1$. In this hierarchical formulation, if M has a $\text{Gamma}(a_M, b_M)$ prior then its full conditional distribution is conjugate and updating M is straight-forward. Of course, this approach increases the number of parameters in the MCMC algorithm and requires specification of H, a_M, and b_M, but it also offers increased modeling flexibility.

When the cohesion is a function of location through distance penalties the interpretation of M is not very intuitive and it is difficult to determine how the expected number of clusters would potentially grow as a function of M, complicating prior specification and posterior inference. Thus, as an alternative approach, instead of putting a prior on M, one could consider putting the prior directly on the number of clusters modelled using local spatial stationary Gaussian processes, and adopting a Voronoi partitioning scheme as done by Kim et al. (2005) that is tractable and computationally efficient. The number and center location of the clusters that define the tessellation would be unknown and estimated using a tailored RJMCMC to move between models of different dimensions. An advantage of this approach is the flexibility, interpretability and the computational advantages by having to invert several covariance matrices of small dimensions rather than the potentially very large global covariance matrix.

Another area for further development is the connection between the cluster centers $\mathbf{s}_h^*$ and the cluster-specific parameters θ_h^*. For example, Fuentes and Reich (2013) extend the SSB model (Reich and Fuentes, 2007) to have cluster means following a spatial process defined over the cluster centers with $\text{Cov}(\theta_h^*, \theta_g^*) = \rho(\mathbf{s}_h^*, \mathbf{s}_g^*)$. They show that with this prior, the mixture model can approximate a continuous process arbitrarily well. It would be interesting to see if this idea can be used to approximate continuous processes within the proposed product-partition framework. Another opportunity to connect cluster centers and model parameters is via the spatial correlation parameters. For example, if ϕ_h is the spatial range parameter in cluster h, then $\log(\phi_h)$ could be modeled as a Gaussian process over $\mathbf{s}_h^*$ or even as a simple function of $\mathbf{s}_h^*$ such as $log(\phi_h) = \beta_0 + \mathbf{s}_h^{*T}\boldsymbol{\beta}_1$. In either case, plots of the posterior of ϕ_{c_i} could be used to evaluate nonstationarity.

Finally, we discuss potential computational benefits of spatial partitioning for large datasets. With n observations, evaluating the Gaussian likelihood requires operations on the dense $n \times n$ spatial covariance matrix, which is cumbersome for large n. As the authors point out, partitioning the observations into smaller blocks replaces the dense covariance matrix with the block diagonal matrix in (15) leading to a welcomed speed increase for evaluating the likelihood. In addition, allowing each cluster to have its own spatial covariance is attractive for large datasets because they are often collected over a vast and heterogeneous domain, such as North America, where stationarity is implausible. However, to exploit this potential benefit would require delicate MCMC implementation. Updating all n cluster labels in sequence would likely be too slow and a careful joint update would be required. An alternative is to fix the cluster labels based on a standard clustering algorithm. Parker et al. (2016) provide an example of this form of clustering for a non-Bayesian analysis non-stationary geostatistical data.

In summary, we reiterate our thanks and congratulations to the authors for providing such a nice paper to discuss. It is easy to envision methodological extensions such as non-Gaussian or multivariate spatial models, and applications is diverse areas such as environmental epidemiology and climate research. We look forward to following these developments.

References

Fuentes, M. and Reich, B. J. (2013). "Multivariate spatial nonparametric modeling via kernel processes mixing." *Statistica Sinica*, 23: 75–97. MR3076159. 304

Gramacy, R. B. (2007). "tgp: an R package for Bayesian nonstationary, semiparametric nonlinear regression and design by treed Gaussian process models." *Journal of Statistical Software*, 19: 6. 303

Gramacy, R. B. and Lee, H. K. H. (2008). "Bayesian treed Gaussian process models with an application to computer modeling." *Journal of the American Statistical Association*, 103: 297. MR2528830. doi: http://dx.doi.org/10.1198/016214508000000689. 303

Kim, H. M., Mallick, B. K., and Holmes, C. C. (2005). "Analyzing nonstationary spatial data using piecewise Gaussian processes." *Journal of the American Statistical Association*, 100: 635–668. MR2160567. doi: http://dx.doi.org/10.1198/016214504000002014. 303, 304

Park, J. H. and Dunson, D. B. (2010). "Bayesian generalized product partition model." *Statistica Sinica*, 20: 1203–1226. MR2730180. 304

Parker, R. J., Reich, B. J., and Eidsvik, J. (2016). "A fused lasso approach to nonstationary spatial covariance estimation." Submitted for publication. 304

Reich, B. J. and Fuentes, M. (2007). "A multivariate nonparametric Bayesian spatial framework for hurricane surface wind fields." *The Annals of Applied Statistics*, 1: 249–264. MR2393850. doi: http://dx.doi.org/10.1214/07-AOAS108. 304