Bayesian Analysis (2016) **11**, Number 1, pp. 315–323

Rejoinder*

Garrit L. Page†§ and Fernando A. Quintana‡¶

1 Introduction

We would very much like to thank Marina Vannucci, the Editor of Bayesian Analysis, for the opportunity of receiving additional feedback regarding our work on spatial product partition models. The comments made by the discussants were insightful and thought provoking. For this we would like to thank Robert Gramacy, Herbie Lee, Brian Reich, Montserrat Fuentes, Carlo Gaetan, Simone Padoan and Igor Prünster for the time spent on reading and critiquing the paper. Our rejoinder is organized according to specific comments made by each discussant.

As a general preliminary comment we would like to emphasize that our main goal was to develop a probability model for partitions that takes into account spatial dependence when forming clusters. This, however, does not preclude the use of various types of sampling models that may be used in tandem with the sPPM prior, including areal and even count data, which may require the use of models beyond Gaussian processes. Therefore, the sPPM prior generically introduces spatial dependence in a statistical model and as a result complements additional spatial structure that may be considered at other stages of a hierarchical model.

2 Gramacy and Lee

Gramacy and Lee (GL)'s comments focused on comparing methods that incorporate regional partitions to produce flexible nonstationary spatial models to the sPPM. The nonstationarity is induced by fitting independent "local" models (e.g., Gaussian process) at each regional partition. Among the methods that were mentioned are local approximate Gaussian processes and treed Gaussian processes (tgp). We would like to thank GL for bringing these methods to our attention as they were not discussed in the main article.

The tgp approach builds a partition by recursively splitting the space so that subset boundaries are parallel to coordinate axes. Within partitions, tgp fits a Gaussian process to the set of responses. A variation of this idea was considered in Kim et al. (2005) with partitions defined in terms of Voronoi tessellations. The sPPM prior provides support that in addition to these, includes subsets of potentially different shapes, according to the spatial features encouraged by the definition of cohesion functions adopted. We

*Main article DOI: 10.1214/15-BA971.

†Department of Statistics, Brigham Young University, Provo, Utah, page@stat.byu.edu

‡Departamento de Estadística, Pontificia Universidad Católica de Chile, Santiago, Chile, quintana@mat.uc.cl

§Partial support provided by FONDECYT Grant 11121131 and ANILLO Project SOC 1107.

¶Partial support provided by FONDECYT Grant 1141057.

 DOI: 10.1214/16-BA971REJ

could also envision combining the sPPM prior with flexible Gaussian processes for each subset in the partition.

GL noted that fitting a tgp model to the first 600 observations of the SIMCE data produced a lower mean squared prediction error (MSPE) on the last 615 SIMCE observations than the conditional model described in Section 4.2 of the main article except for the Conditional Model with Prior Spatial Structure (CPS) 4. Our main concern with the comparisons made in GL's comments deals with the information used to partition the feature space. According to our understanding, the entire feature space is partitioned for tgp which in the example considered includes mother's education in addition to spatial coordinates. The conditional models of Section 4.2 based on the sPPM only incorporate spatial coordinates in the partition model (i.e., mother's education only appears in the likelihood). To make comparisons more equitable the partition model of sPPM would need to depend on mother's education.

As mentioned in the last section of the main article making this extension is fairly straightforward if the covariate dependent product partition model ideas of Müller et al. (2011) are considered. All that is required is to multiply the cohesion function of the sPPM by a nonnegative function $g(\cdot)$ that measures the similarity of the mother's education values (x's) that are in $\boldsymbol{x}_j^\star = \{x_i : i \in S_j\}$. This would change equation (3.1) of the main article to

$$Pr(\rho) \propto \prod_{j=1}^{k_n} C(S_j, \boldsymbol{s}_j^\star) g(\boldsymbol{x}_j^*).$$

$g(\cdot)$ is referred to as a similarity function and in principle can be any nonnegative function that produces larger values when the $\boldsymbol{x}$'s in $\boldsymbol{x}_j^*$ are more similar. That said, here we opt to employ the so called double dipper similarity (the same object used to construct C4, see Quintana et al. 2015) which has the following form

$$g(\boldsymbol{x}_j^*) = \int \prod_{i \in S_j} q(x_i|\xi_j) q(\xi_j|\boldsymbol{x}_j^*) d\xi_j.$$

Notice that this similarity has the functional form of a posterior predictive distribution, but we are not assuming the $\boldsymbol{x}$'s to be random. We are simply using a "likelihood" and "posterior" to measure closeness of covariate values. For mother's education we set $q(x_i|\xi_j = (m_j, v_j^2)) = N(x_i|m_j, v_j^2)$. According to suggestions in Müller et al. (2011) we set $v_j = 10s$, where s is the empirical standard deviation of mother's education scores. Using a Gaussian "prior" for m_j produces $q(\xi_j = m_j|\boldsymbol{x}_j^*) = N(m_j|m_j^*, v_j^{2*})$ where m_j^* and v_j^{2*} are the typical weighted average and variance found in the posterior distribution using a Gaussian likelihood and Gaussian prior on the mean.

Finally, retaining the assumption of a global likelihood variance and slope, once cluster labels are introduced the hierarchical specification of the CPS model becomes

$$\begin{aligned} y_i|x_i, \mu_j^*, \sigma^2, c_i = j &\sim N(\mu_j^* + \beta x_i, \sigma^2) \text{ with } \sigma \sim UN(0, 10) \\ \mu_j^* &\sim N(\mu_0, \sigma_0^2) \end{aligned}$$

$$\mu_0 \sim N(0, 100^2)$$
$$Pr(\rho) \propto \prod_{j=1}^{k_n} C(S_j, \boldsymbol{s}_j^\star) g(\boldsymbol{x}_j^*).$$

We will refer to this hierarchical model as sPPMx.

The MSPE values produced by fitting sPPMx to the SIMCE data treating the first 600 observations as training and the remaining 615 as testing are provided in Table 1. We also list the tgp MSPE as reported by GL. It appears that Cohesions 1 and 4 of sPPMx provide lower MSPE values, but tgp performs better relative to Cohesions 2 and 3. Expecting to outperform tgp for all cohesions would not be realistic as the tgp incorporates very flexible cluster specific Gaussian process fits. As noted by GL the partition used to split the SIMCE dataset into testing and training observations influences the MSPE values. Therefore, we also considered 100 randomly generated partitions of the of the data into testing and training observations and fit the models to each in addition to fitting a tgp model using the default settings of the `btgp` function found in the `tgp` package (Gramacy 2007 and Gramacy and Taddy 2010) for `R` (R Core Team 2015). The average MSPE values are provided in the second column of Table 1. Here tgp is more competitive, but Cohesions 3 and 4 still outperform tgp in terms of MSPE.

Model	MSPE	$\overline{\text{MSPE}}$
CSP1	361.1	346.1
CSP2	398.4	409.5
CSP3	371.3	345.7
CSP4	360.9	345.6
TGP	365.0	345.8

Table 1: MSPE values when fitting sPPM with 4 cohesions and tgp to the SIMCE data. The data was divided into 600 training and 615 testing observations. The first column corresponds to the first 600 observations being training. The second column is the average MSPE value over 100 different partitions of the data into testing and training observations.

The comment regarding the utility of nonstationary methods to model the SIMCE dataset is warranted. We believe that the data set is useful to make comparisons between different spatial partitioning methods, but upon further exploration, we noticed that mother's education tends to explain much of the of the spatial structure in SIMCE math scores. Therefore, conditional on mother's education level simple spatial models for SIMCE math scores are very competitive. This motivated the inclusion of the scallops data in the main article.

We finish our response to GL's comments by noting that even though the sPPM does produce competitive MSPE values relative to tgp and spatial stick breaking (SSB) of Reich and Fuentes (2007), our main focus was to develop probabilistic partition models that incorporate spatial information in a principled way. That is, a model that places higher probability on spatially "pleasing" clusters (only through the process of developing the sPPM did we become aware of some of the nice spatial properties that resulted in fitting a model on top of the spatially referenced partition). Therefore, the comment regarding interpreting the spatially referenced clusters is highly relevant. In the supplementary file of the main article we provide figures that contain estimated partitions of the schools (employing Dahl 2006's least-squares method). The estimated partitions are quite different for each of the cohesions, but all grouping generally followed the socioeconomic "boundaries" that are generally believed to influence school performance in Santiago.

3 Reich and Fuentes

Reich and Fuentes (RF) point out many interesting connections and extensions of the sPPM methodology. We start by noting that connected partitions, though desirable in many applications, are hard to impose through the prior. Thus, our philosophical approach considers priors that encourage connectedness but that still allow departures from this with hopefully small probability. This introduces some local correlation. Smoothness can be further controlled at a local and/or global level by conveniently chosen sampling models, as exemplified in Section 3.3 of the main article. As further discussed in Section 2 of this rejoinder our model fares well when compared to the treed Gaussian processes of Gramacy and Lee (2008).

We agree with RF that the results in Table 4 of the simulation study were unexpected and upon further investigation found a slight error in how the mean squared prediction error (MSPE) was calculated. This motivated an improved simulation study that better illustrates the utility of the methodology being developed. Since the sPPM model and the spatial regression (SR) model accommodate spatial information very differently (the former in the prior and the latter in the likelihood), using the SR model as a baseline model is not appropriate. A more appropriate baseline model would be the product partition model with no spatial information (PPM). However to illustrate sPPM's ability to accommodate non-stationary spatial processes, we retain the SR model in the simulation study. A table of new results is provided here in Tables 2 and 3.

First notice that in both tables the sPPM outperforms the PPM in terms of model fit and prediction in all data generating scenarios and for all four cohesion functions. Thus, including spatial information in the partition model is useful. Also, the sPPM incorporates spatial information more efficiently than the SSB which is evidenced by the better fits and out of sample predictions across all data generating scenarios. Next, notice that in Table 3 where the data are generated from the SR model (i.e., there is only one cluster) the SR model has the best performance in terms of MSPE which is to be expected. The fact that the sPPM is somewhat competitive is surprising since the SR model is not only the correct model, but also incorporates spatial information more directly (i.e., through the likelihood) than the sPPM. The LPML values associated with the SR model are quite sensitive to a few MCMC iterates that produce very small likelihood values and hence produce low LPML values. We also consider a scenario

Error	Cluster	Method	$M = 0.01$			$M = 0.1$			$M = 1.0$		
			RAND	LPML	MSPE	RAND	LPML	MSPE	RAND	LPML	MSPE
Gaussian	Square	CPS $C1_{\alpha=1}$	0.71	-139.79	3.43	0.63	-139.15	4.65	0.54	-139.97	6.67
		CPS $C1_{\alpha=2}$	0.45	-138.62	2.73	0.38	-138.97	3.44	0.30	-141.38	5.60
		CPS C_2	0.80	-147.81	2.00	0.78	-146.51	2.04	0.61	-146.52	2.14
		CPS C_3	0.97	-143.66	1.61	0.97	-144.01	1.70	0.97	-144.08	1.66
		CPS C_4	0.98	-142.78	1.46	0.95	-142.56	1.58	0.88	-141.60	1.62
		PPM	0.70	-160.08	14.34	0.70	-157.50	14.29	0.66	-156.09	14.07
		SSB	0.69	-160.05	14.90	0.68	-158.07	14.54	0.70	-153.53	14.39
		SR	-	-130.43	1.81	-	-98.20	1.78	-	-126.15	1.68
	Irregular	CPS $C1_{\alpha=1}$	0.73	-139.90	5.15	0.63	-138.28	5.55	0.53	-139.07	7.04
		CPS $C1_{\alpha=2}$	0.56	-140.66	4.31	0.48	-139.82	4.56	0.40	-141.84	6.13
		CPS C_2	0.70	-155.44	5.38	0.71	-154.34	5.03	0.60	-151.24	5.73
		CPS C_3	0.92	-145.50	3.61	0.96	-143.02	3.25	0.95	-143.20	3.33
		CPS C_4	0.95	-143.57	3.27	0.94	-142.12	2.98	0.87	-140.88	3.02
		PPM	0.69	-158.23	13.88	0.72	-156.13	14.41	0.68	-153.02	14.33
		SSB	0.69	-171.93	9.41	0.73	-163.47	9.84	0.74	-150.15	10.10
		SR	-	-637.28	4.14	-	-611.88	3.78	-	-540.67	3.64
Mixture	Square	CPS $C1_{\alpha=1}$	0.68	-150.21	3.87	0.61	-148.13	5.02	0.55	-150.11	6.90
		CPS $C1_{\alpha=2}$	0.44	-150.52	3.08	0.36	-149.70	3.74	0.31	-153.48	5.95
		CPS C_2	0.77	-158.03	2.29	0.76	-156.56	2.17	0.60	-156.87	2.45
		CPS C_3	0.95	-156.12	1.99	0.96	-154.58	1.93	0.96	-156.96	1.96
		CPS C_4	0.96	-155.30	1.83	0.94	-152.70	1.77	0.85	-153.44	1.94
		PPM	0.65	-169.13	14.64	0.67	-164.99	14.50	0.62	-162.98	14.13
		SSB	0.67	-167.11	15.14	0.68	-168.31	15.01	0.68	-161.82	14.39
		SR	-	-154.11	2.09	-	-168.53	2.07	-	-144.63	2.01
	Irregular	CPS $C1_{\alpha=1}$	0.70	-151.62	5.57	0.60	-148.10	6.29	0.53	-150.62	7.14
		CPS $C1_{\alpha=2}$	0.54	-152.50	4.45	0.47	-150.54	5.25	0.39	-154.60	6.39
		CPS C_2	0.69	-165.92	5.73	0.68	-162.02	6.08	0.58	-161.35	5.54
		CPS C_3	0.93	-156.49	3.65	0.92	-154.08	3.99	0.93	-154.90	3.59
		CPS C_4	0.95	-155.80	3.28	0.92	-152.59	3.66	0.85	-152.63	3.41
		PPM	0.66	-166.89	14.57	0.68	-162.92	14.70	0.62	-161.30	14.39
		SSB	0.66	-179.80	10.26	0.64	-179.93	10.29	0.68	-167.08	10.38
		SR	-	-543.82	4.11	-	-563.06	4.36	-	-593.35	4.07

Table 2: Simulation study results when data are generated with four clusters.

			$M = 0.01$		$M = 0.1$		$M = 1.0$	
Error	Cluster	Method	LPML	MSPE	LPML	MSPE	LPML	MSPE
Gaussian	Square	CPS $C1_{\alpha=1}$	-130.04	1.00	-128.05	1.03	-120.70	1.01
		CPS $C1_{\alpha=2}$	-129.52	0.97	-123.90	1.00	-118.40	0.98
		CPS C_2	-140.56	1.01	-138.54	1.02	-134.69	1.00
		CPS C_3	-142.98	1.03	-142.58	1.05	-140.97	1.01
		CPS C_4	-139.64	1.00	-138.10	1.01	-133.87	0.98
		PPM	-144.10	1.05	-142.66	1.06	-137.02	1.04
		SSB	-140.49	1.04	-140.28	1.06	-133.81	1.04
		SR	-281.24	0.84	-293.87	0.85	-292.41	0.84
	Irregular	CPS $C1_{\alpha=1}$	-129.04	0.98	-123.08	0.95	-116.23	1.01
		CPS $C1_{\alpha=2}$	-126.81	0.94	-120.82	0.92	-113.70	0.97
		CPS C_2	-138.08	0.99	-136.48	0.95	-130.78	1.00
		CPS C_3	-140.68	1.00	-140.80	0.97	-139.06	1.01
		CPS C_4	-138.06	0.98	-137.10	0.95	-133.39	0.97
		PPM	-143.73	1.07	-141.41	1.01	-137.10	1.06
		SSB	-139.23	1.03	-138.50	1.00	-132.00	1.05
		SR	-203.34	0.76	-200.72	0.73	-199.90	0.77
Mixture	Square	CPS $C1_{\alpha=1}$	-142.22	1.29	-136.85	1.28	-133.89	1.26
		CPS $C1_{\alpha=2}$	-142.55	1.25	-137.22	1.25	-133.32	1.23
		CPS C_2	-153.68	1.30	-149.58	1.28	-144.73	1.25
		CPS C_3	-157.41	1.31	-153.87	1.29	-152.11	1.26
		CPS C_4	-153.23	1.29	-148.70	1.27	-143.97	1.22
		PPM	-155.25	1.32	-151.11	1.32	-145.80	1.29
		SSB	-152.85	1.32	-152.05	1.31	-145.92	1.29
		SR	-476.61	1.13	-331.10	1.14	-387.94	1.10
	Irregular	CPS $C1_{\alpha=1}$	-141.53	1.21	-136.98	1.23	-132.90	1.26
		CPS $C1_{\alpha=2}$	-140.48	1.16	-136.10	1.19	-132.36	1.22
		CPS C_2	-150.64	1.21	-146.96	1.23	-144.37	1.25
		CPS C_3	-153.69	1.23	-152.58	1.25	-153.29	1.25
		CPS C_4	-151.16	1.20	-148.71	1.22	-147.97	1.22
		PPM	-153.81	1.29	-149.66	1.28	-147.42	1.32
		SSB	-151.71	1.27	-149.96	1.27	-146.85	1.31
		SR	-468.13	1.00	-383.73	1.03	-436.47	1.06

Table 3: Simulation study results when data are generated with one cluster.

where a small amount of nonstationarity is introduced in the spatial process via cluster specific intercepts. In this case C3 and C4, which more readily admit nonstationary spatial processes (see Figure 4 of the main article), produce superior MSPE values even compared to the SR model. This is quite remarkable given that sPPM is only including spatial information through the prior.

The simulation study illustrates the benefits of assigning higher probability to spatially connected partitions in a nonstationary setting (which C3 and C4 do). Since the Bayesian Kriging predictor is a linear predictor of $\boldsymbol{y}$ which assigns higher weight to "local" observations, Cohesion 1 which admits more spatially diverse partitions does not fair as well. The benefit of Cohesion 1 is being able to accommodate stacked clusters (i.e., multiple clusters in the same region) but this flexibility comes at a cost when modeling nonstationary processes. Additionally we envisioned Cohesion 2 being more useful in areal data modeling.

RF rightly point out the relevance of the number of clusters. In sPPM and related models, the maximum number of imputable clusters is n, and the choice of cohesion functions determines implicitly the prior distribution of k_n. Analytical results are unfortunately very hard to derive under the type of spatially dependent definitions of $C_j(S_h, \boldsymbol{s}_h^\star)$, $j = 1, \ldots, 4$ we tried. The prior simulation study summarized in Table 2 was partially designed to shed some light on the implied distributions for k_n, precisely because it is not clear how to interpret the M parameter. In this regard, RF suggested a hierarchical model with a stick breaking prior, as in Dunson and Park (2008). With a sufficiently large upper bound on the maximum possible number of clusters H, this indeed allows setting up a gamma prior for the M parameter, which can then be easily handled. The same idea can be readily extended to more general classes of stick breaking priors. One subtlety of this approach is that it treats spatial coordinates $\boldsymbol{s}_i$ as random, which may have undesired effects in posterior predictive inference, as discussed in Müller et al. (2011). RF also suggest the clever idea of modeling directly the number of clusters under a local spatial stationary Gaussian processes, and adopting Voronoi tessellations as in Kim et al. (2005), thus avoiding the complicated doubly intractable problem that would result when putting a prior on M under any of the C_j functions we explore. This idea is promising and certainly deserves to be further studied.

RF wonder about the use of sPPM when cluster-specific parameters are correlated to cluster centers, discussing two particular forms to achieve such goal. Such strategy can be certainly adapted to our proposed models. Although the particular details need to be sorted out, one envisions similar properties to those described in Fuentes and Reich (2013), with one big caveat: the sPPM does not involve random distributions, and the prior concentrates on the discrete space of partitions. Adding correlation to cluster-specific parameters indeed contributes to smoothing out the marginal sampling process. In addition, it is of interest in such a case to study how this correlation affects the partitions created.

4 Carlo Gaetan, Simone A. Padoan and Igor Prünster

Similar to RF, Gaetan, Padoan and Prünster (GPP) discuss interesting extensions to the sPPM and detail some additional contexts in which spatially referenced clusters might provide some benefit. This opens the door to many possible interesting models. Our default basic cohesion function $c(S) = M \times \Gamma(|S|)$ gives rise to an exchangeable partition probability function (EPPF) that is a special case of the general family of Gibbs-type random partitions. GPP explain that under an exchangeable PPM, considering a cohesion function that depends only on the subset size leads to a class that coincides with the family of Gibbs-type partitions. Of course, as GPP note, our proposed spatially dependent formulation has to go beyond the exchangeable format. But it is possible to build spatially-dependent extensions using the same general principle. For instance, taking $c(S) = (1-\sigma)_{|S|-1}$ with $\sigma < 1$ and where $(x)_m = \prod_{j=1}^{m}(x+j-1)$, one may consider a modification of C_2 in (4) as

$$\tilde{C}_2(S_h, \boldsymbol{s}_h^\star) = (1-\sigma)_{|S|-1} \times \prod_{i,j \in S_h} I\left\{\|\boldsymbol{s}_i - \boldsymbol{s}_j\| < a\right\}. \tag{1}$$

Different variations are also possible. One issue that remains open in this context is to study the distribution of the number of clusters for the resulting spatial-dependent partition structure. In addition, incorporating a prior distribution on σ when using the partition distribution induced by (1) is again a difficult problem, akin to the case of assuming a prior on M. The simplified version suggested by GPP with σ restricted to $\{-1, 0, 1/2\}$ is a very interesting way of combining the various types of baseline prior behavior for the number of clusters they describe.

GPP also introduced the idea of employing spatially referenced partition models in the context of spatial extremes and also of models based on max-stable random fields. We agree that this would be a promising avenue of research, including in particular the study of the effect that the interplay between likelihood- and partition-based spatial assumptions has on the observable data. GPP's suggestions highlight the fact that indeed the sky is the limit in modeling contexts once a model on ρ has been specified.

In closing this rejoinder, we observe that our proposal as well as many of the suggestions pointed out by the discussants can be extended to a spatio-temporal setting. One possible goal here would be to define a sequence of partitions that change over time, where at each time, a covariate- and/or spatially-dependent prior partition distribution is considered. A possible starting point is the dynamic generalized Pólya urn model of Caron et al. (2008), where a sequence of urn models are related in time by way of a deletion-reallocation procedure that guarantees marginal DP-style partitions at each time.

References

Caron, F., Davy, M., Doucet, A., Duflos, E., and Vanheeghe, P. (2008). "Bayesian Inference for Linear Dynamic Models with Dirichlet Process Mixtures." *IEEE Transactions on Signal Processing*, 56(1): 71–84. MR2439814. doi: http://dx.doi.org/10.1109/TSP.2007.900167. 322

Dahl, D. B. (2006). "Model-Based Clustering for Expression Data via a Dirichlet Process Mixture Model." In: Vannucci, M., Do, K. A., and Müller, P. (eds.), *Bayesian Inference for Gene Expression and Proteomics*, 201–218. Cambridge University Press. 318

Dunson, D. B. and Park, J.-H. (2008). "Kernel Stick-Breaking Processes." *Biometrika*, 95: 307–323. MR2521586. doi: http://dx.doi.org/10.1093/biomet/asn012. 321

Fuentes, M. and Reich, B. J. (2013). "Multivariate Spatial Nonparametric Modeling via Kernel Processes Mixing." *Statistica Sinica*, 100: 75–97. MR3076159. 321

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(9): 1–46. `http://www.jstatsoft.org/v19/i09/`. 317

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: 1119–1130. MR2528830. doi: `http://dx.doi.org/10.1198/016214508000000689`. 318

Gramacy, R. B. and Taddy, M. (2010). "Categorical Inputs, Sensitivity Analysis, Optimization and Importance Tempering with tgp Version 2, an R Package for Treed Gaussian Process Models." *Journal of Statistical Software*, 33(6): 1–48. `http://www.jstatsoft.org/v33/i06/`. MR2578072. doi: `http://dx.doi.org/10.1007/s11222-008-9108-5`. 317

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: 653–668. MR2160567. doi: `http://dx.doi.org/10.1198/016214504000002014`. 315, 321

Müller, P., Quintana, F., and Rosner, G. L. (2011). "A Product Partition Model With Regression on Covariates." *Journal of Computational and Graphical Statistics*, 20(1): 260–277. MR2816548. doi: `http://dx.doi.org/10.1198/jcgs.2011.09066`. 316, 321

Quintana, F. A., Müller, P., and Papoila, A. L. (2015). "Cluster-Specific Variable Selection for Product Partition Models." *Scandinavian Journal of Statistics*, 42: 1065–1077. 316

R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. `https://www.R-project.org/`. 317

Reich, B. J. and Fuentes, M. (2007). "A Multivariate Semiparametric Bayesian Spatial Modeling Framework for Hurricane Surface Wind Fields." *The Annals of Applied Statistics*, 1: 249–264. MR2393850. doi: `http://dx.doi.org/10.1214/07-AOAS108`. 317