Atomic Spatial Processes

Sean Jewell SEAN.JEWELL@STAT.UBC.CA
Neil Spencer NEIL.SPENCER@STAT.UBC.CA
Alexandre Bouchard-Côté BOUCHARD@STAT.UBC.CA
Department of Statistics, University of British Columbia, Vancouver, BC V6T 1Z4, Canada

A. Justification of Model Comparisons

To compare predictive posterior likelihoods in Section 6, we use log predictive likelihoods for the held out points. For mixed distributions, the combination of atoms and densities can make these comparisons difficult. The following work shows how we compare models with atoms in the same locations.

Consider an atomic urban spatial data set X ($|X| = N$). Let $D = \{D_1, \ldots, D_n\}$ denote another dataset, held-out from X, consisting of n_1 observations ($D_{1:n_1}$) with locations observed in X, and $n_2 = n - n_1$ observations ($D_{(n_1+1):n}$) with locations not observed in X. Consider fitting two different ASP models M_1 and M_2 to X. The models M_1 and M_2 differ in which parameters were resampled, or through the prior on G_0 (parametric prior or nonparametric). Consider an $\epsilon > 0$ such that the Euclidean distance between any two unique locations in X exceeds ϵ. Let y_ϵ denote a square of side lengths ϵ centred at y. Equation 1 demonstrates a posterior predictive probability assigned to the area y_ϵ given X in ASP models (the hyperparameters differ between M_1 and M_2).

$$\mathbb{P}(Y \in y_\epsilon | X = x) = \frac{1}{N + \alpha_0} \sum_{i=1}^{N} \mathbb{1}(x_i \in y_\epsilon) + \frac{\alpha_0}{N + \alpha_0} H'(Y \in y_\epsilon | X).$$

Here, K denotes the number of mixture components in the realization of G_0 which correspond to at least one element of X, the a_i denote the atoms in the data, $|B_i|$ is the number of atoms in mixture component i, $T = \sum_{i=1}^{K} |B_i|$ is the number of unique locations, and

$$H'(Y \in y_\epsilon | X) \approx \frac{\epsilon^2}{T + \alpha_0^\pi} \left(\sum_{i=1}^{K} |B_i| m(y | a_{B_i}) + \alpha_0^\pi m(y) \right).$$

Note that for the models with parametric priors on G_0, K is fixed at 1 and $\alpha_0^\pi = 0$. The approximation in Equation 2 comes from the fact that the density of the DPM realization G_0 is assumed to be constant within y_ϵ. Since any G_0 considered in an ASP is continuous, this approximation can be made arbitrarily accurate by choosing ϵ small enough. In addition, as ϵ shrinks, the second summand goes to 0.

B. Notes on Statistical Computation

To determine the posterior predictive densities for graffiti illustrated as Figures 2b and 2d, Markov Chain Monte Carlo runs consisting of 20000 sweeps were used. A burn-in of 1000 sweeps was used, and the samples were thinned to 1 in every 500 to create the plots. The initializations for the hyperparameters were determined through short runs to discern which values led to fast mixing, the values were not necessarily the same for the Vancouver and Manhattan runs. The trace plots shown as Figure S.1 and Figure S.2 illustrate the values of the hyperparameters for the Vancouver and Manhattan runs, respectively. For the experiments comparing models, if a hyperparameter was not resampled, they were held fixed at their initialization. See Table S.1 for all initial hyperparameter values and the associated normal random walk kernel specification. Different thinning levels were used to create the trace plots and to create the posterior predictive. The posterior predictive is computationally intensive to compute whereas the trace plots were thinned mainly for readability.

	Initial value		MCMC proposal SD	
	Vancouver	New York	Vancouver	New York
α_0	1000	10000	50	60
α_0^π	1.000	1.000	1.000	1.000
c	0.500	3.000	0.500	0.500
κ	0.500	0.001	0.100	0.100
ν	2.000	2.000	0.100	0.100

Table S.1. Initial hyperparameter values and their corresponding normal random walk MCMC kernel standard deviations.

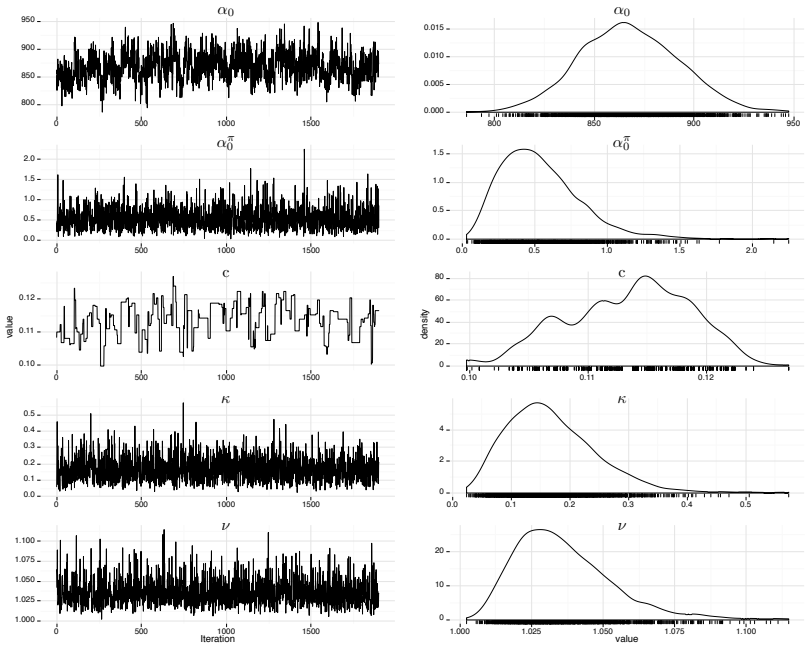

Figure S.1. Trace plot summarizing the hyperparameter values of 20 000 MCMC iterations for Vancouver, with the left side showing the chain trajectory and the right side illustrating the approximate posterior distributions. The results are thinned by using 1 in every 10 elements in the chain.

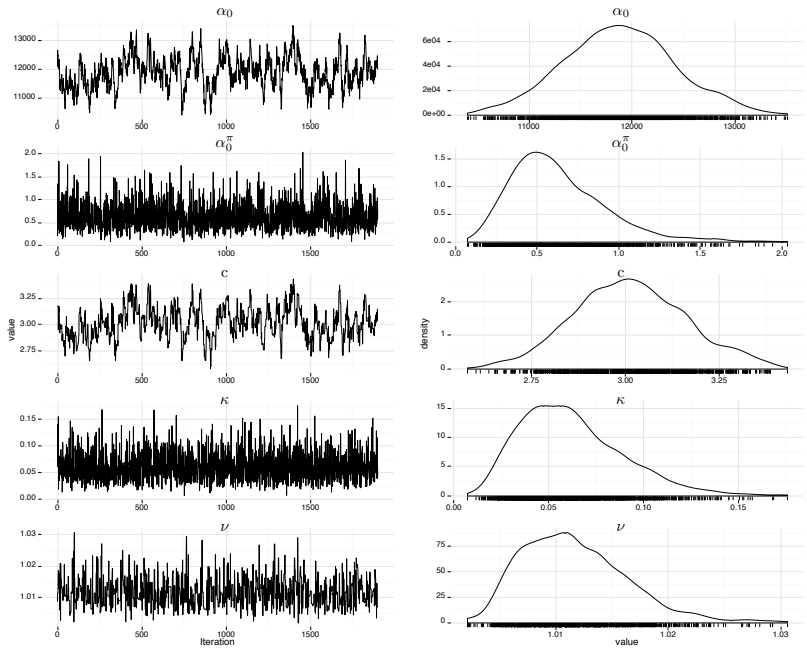

Figure S.2. Trace plot summarizing the result of 20 000 MCMC iterations for Manhattan. The results are thinned by including 1 in every 10 elements in the chain.

C. An Extended List of Competitors

The following is a summary of the results of experiments for an extended list of 11 competitor models. This consists of the four shown in the paper, and seven additional models. Performance is shown for the Vancouver and Manhattan data sets.

The new competitors are variants of those shown in the paper in which different hyperparameters are held constant instead of being resampled. The distinguishing features of these 11 models are presented in Table S.2. For ASP models, the column "Base measure" indicates whether the base measure for the GaP prior for μ is a DPM (non-parametric) or a single normal measure (parametric). The Normal-Inverse-Wishart is the base measure for DPM models, and a piecewise uniform measure is used for the empirical method. For the other columns, a ✓ indicates the hyperparameter was resampled, ✗ indicates the hyperparameter was fixed at its initial value, and – denotes that the hyperparameter is not applicable. Figure S.3 shows the predictive performance and wall-times for the additional models.

Table S.2. Model specification: ✓indicates that a hyperparameter was resampled versus fixed ✗, and – denotes not applicable.

Model	Base measure	α_0^π	α_0	c	κ	ν
ASP 1	Normal	–	✗	✗	✗	✗
ASP 2	Normal	–	✓	✓	✓	✓
ASP 3	DPM	✗	✗	✗	✗	✗
ASP 4	DPM	✓	✓	✗	✗	✗
ASP 5	DPM	✓	✓	✓	✗	✗
ASP 6	DPM	✓	✓	✓	✓	✗
ASP 7	DPM	✓	✓	✓	✓	✓
DPM 1	DPM	✗	–	–	✗	✗
DPM 2	DPM	✓	–	–	✓	✓
Empirical	Grid	–	–	–	–	–
Mixed	DPM	✓	–	–	✓	✓

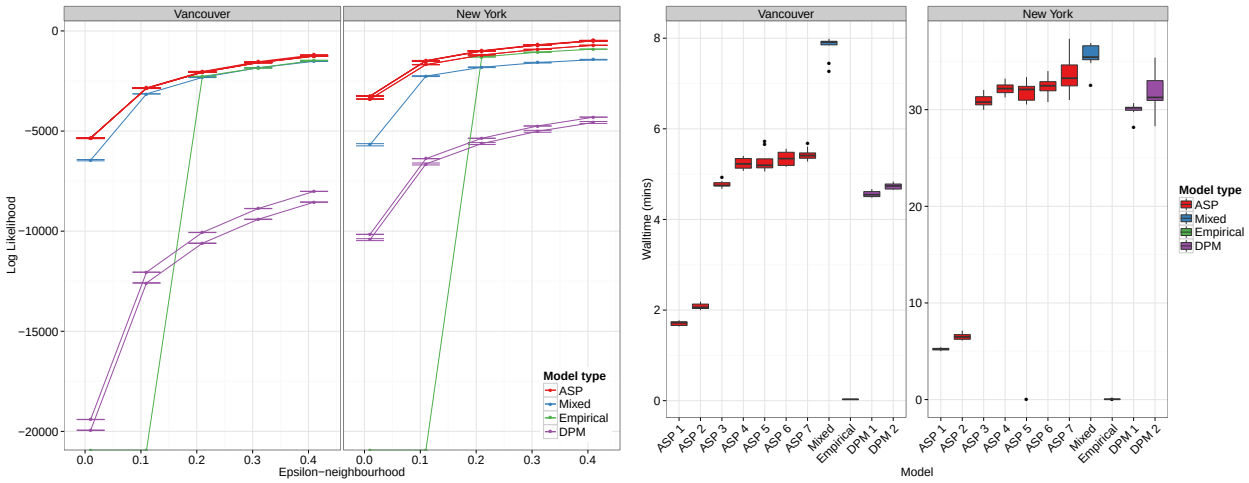

Figure S.3. (Left) Predictive log-likelihood for ϵ-neighbourhoods of held out pieces of graffiti for the extended list of competitor models. Seven different variations of the ASP model (corresponding to different resampling strategies for the hyperparameters) uniformly outperform two Dirichlet process mixture models (DPM), an empirical grid based approach, and the DPM-Empirical mixture with cross-validated mixing proportions. (Right) Walltime in minutes for each of the 11 models.