Fast Simulation of Markov Cluster Processes

Rafaela M. Guidi and Ranjan Maitra *

Abstract

Spatial point processes have been used by scientists in various disciplines such as astronomy, ecology, geology and urban planning. Cluster processes are a special case where the point patterns to be represented have an affinity for being close to each other. However, analytical inference in such models are intractable, because in many cases, the likelihood function is known only up to a normalizing constant. Stochastic methods of inference are then the only hope. This paper studies single-point simulation schemes and illustrates their inadequacies in the presence of strong clustering preferences. Augmenting the point process distribution with partially decoupled auxiliary variables while also maintaining the marginal distribution of the cluster process, is seen to perform very well.

Keywords: cluster process, Strauss process, birth-and-death process, auxiliary variables, partial decoupling

1 Introduction

The characterization of spatial point patterns has long been an object of interest in several fields, starting from ecology [11] where the objective was to describe the spatial distribution of plant material, to cosmology where Neyman and Scott [18] hypothesized the universe as a realization of a stochastic process. Some more applications are provided by Holder *et al.* [14] with a view to study spatial patterns in archaeological data or by Dacey [7] to study the impact of population density on the location of primary shopping areas.

Statistical interest in the theory of spatial point patterns owes its genesis to the seminal paper of Ripley [19] which reviewed stochastic models such as cluster processes, hard-core models with fixed-range and pairwise interactions as well as mixtures of such models. In a complementary development, Ripley and Kelly [21] introduced the Markov point process opening a rich source of models with the possibility of characterizing interactions and inhibitions between points. They showed that these models can be simulated by stochastic methods as the equilibrium distribution of a birth-and death process. [15]. This is particularly useful because most of these models have likelihoods that are known only up to a normalizing constant, which in turn is impractical to obtain either analytically or computationally [1, 9, 17].

Baddeley and Møller [1] widened the class of Markov point processes by specifying models with the *nearest-neighbor* Markov property. When one simulates from such models, inclusion of a new point has no effect on the interaction between two other points unless they are both neighbors of the new point. Further a nearest-neighbor Markov Point Process can, under certain conditions, be simulated by a birth-and-death process. The class of Markov point processes with the nearest-neighbor property has a number of important members such as area-interaction processes, cluster processes, Dirichlet tessellations and line-segment process.

In this paper, we study some simulation algorithms for cluster processes and show how their efficiency, especially in the presence of strong interactions can be improved by extending the theory of auxiliary variables [13, 22] and partial decoupling [13] to this framework. The goal behind inclusion of a partially decoupled auxiliary variable is to partially break the strong interaction component in the likelihood function, permitting greater mobility of the birth-and-death process around the state-space. The augmentation of the process with partially decoupled auxiliary variables is done so that the limiting marginal distribution of the point process is maintained. Test experiments reported in this paper show excellent performance. For convenience, simulations have usually been done conditional on the number of points [1, 19, 20] and we follow this practice here. We also extend our development to the general case of an unconditional point process. Throughout this paper, we use the term *conditional point process* to refer to a point process conditional on the number of points. Point processes not conditional on the number of points are referred to as *unconditional point processes.*

This paper is organized in three more sections. Section 2 introduces some preliminary notations, definitions, existing simulation algorithms, and also illustrates the poor performance of single-point simulation schemes in the presence of high interactions. Section 3 extends the partially decoupled auxiliary variables approach [13, 22] and illustrates the performance for both conditional and unconditional cluster processes. Fi-

*Rafaela M. Guidi is Graduate Teaching Assistant and Ranjan Maitra is Assistant Professor in the Department of Mathematics and Statistics, University of Maryland, Baltimore County, Baltimore, MD, 21250. The first author is also a Junior Advisor in the Control Department, Bank of Brazil, Brasilia, Brazil.

nally, Section 4 discusses the results in this paper and poses issues for further research.

2 Background

2.1 Preliminaries

A two-dimensional cluster process, also referred to as a connected component point process [5], is a finite simple point process $\mathbf{X}$ in a compact region $S \subset \mathbb{R}^2$. A simple spatial point process in $\mathbb{R}^2$ is a process where almost surely at most one event occurs [6]. A realization of $\mathbf{X}$ is a finite set $\mathbf{x} = \{x_1, x_2, ..., x_n\}$ where each $x_i \in S$ and $n \geq 0$. Given a realization $\mathbf{x}$, construct the finite graph $\mathcal{G}$ with vertices x_i, and edges joining those pairs (x_i, x_j) of points that are less than $\rho > 0$ apart in Euclidean distance. Suppose that the graph $\mathcal{G}$ consists of one or more connected components $c_1, c_2, ..., c_{m(\mathbf{x})}$. Suppose also that its density can be written as:

$$f(\mathbf{x}) = \exp\{\log \alpha + n(\mathbf{x})\ log\beta + m(\mathbf{x})\ log\gamma\} \qquad (1)$$

where α is the normalizing constant, $n(\mathbf{x})$ is the number of points in the configuration $\mathbf{x}$, $\beta > 0$ is the intensity parameter, $m(\mathbf{x})$ is the number of connected component of $\mathbf{x}$, and $\gamma > 0$ is the interaction parameter.

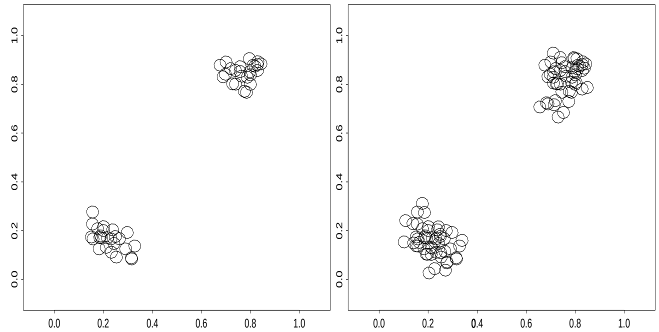
Note that γ controls the interaction between points with small values indicating a preference for realizations with fewer connected components. Further, for $\gamma = 1$ we have an homogeneous Poisson Point Process with intensity β. Also, Baddeley *et al.* [2] showed that these processes have the *nearest-neighbor* Markov property in the sense of Baddeley and Møller [1] with respect to the connected component relation. Finally, denoting the number of points in the ith cluster by n_i, the indicator function as $1_{[.]}$ and rewriting $m(\mathbf{x})$ in (1) as

$$m(\mathbf{x}) = \sum_{i=1}^{m(\mathbf{x})} \sum_{k \sim j} \frac{2}{n_i(n_i - 1)}\ 1_{[c(x_k)=c(x_j)=i]}, \qquad (2)$$

yields a special case of a Strauss process. Note that the second summation is always equal to 1 so that there is no discrepancy in the above notation with $m(\mathbf{x})$ in the limit of the sum as well as the left-hand-side of (2). For such a process with $0 \leq \gamma \leq 1$, one can specify an unique spatial birth-and-death process with unique equilibrium measure as that Strauss process [1]. This means that regardless of the initial starting point, one can obtain in the long-run, a dependent sample from the stationary distribution by drawing realizations from the corresponding spatial birth-and-death process.

2.2 Simulation Algorithms

A spatial conditional birth-and-death process can be simulated using a Hastings algorithm [12] with an initial



(a) (b)

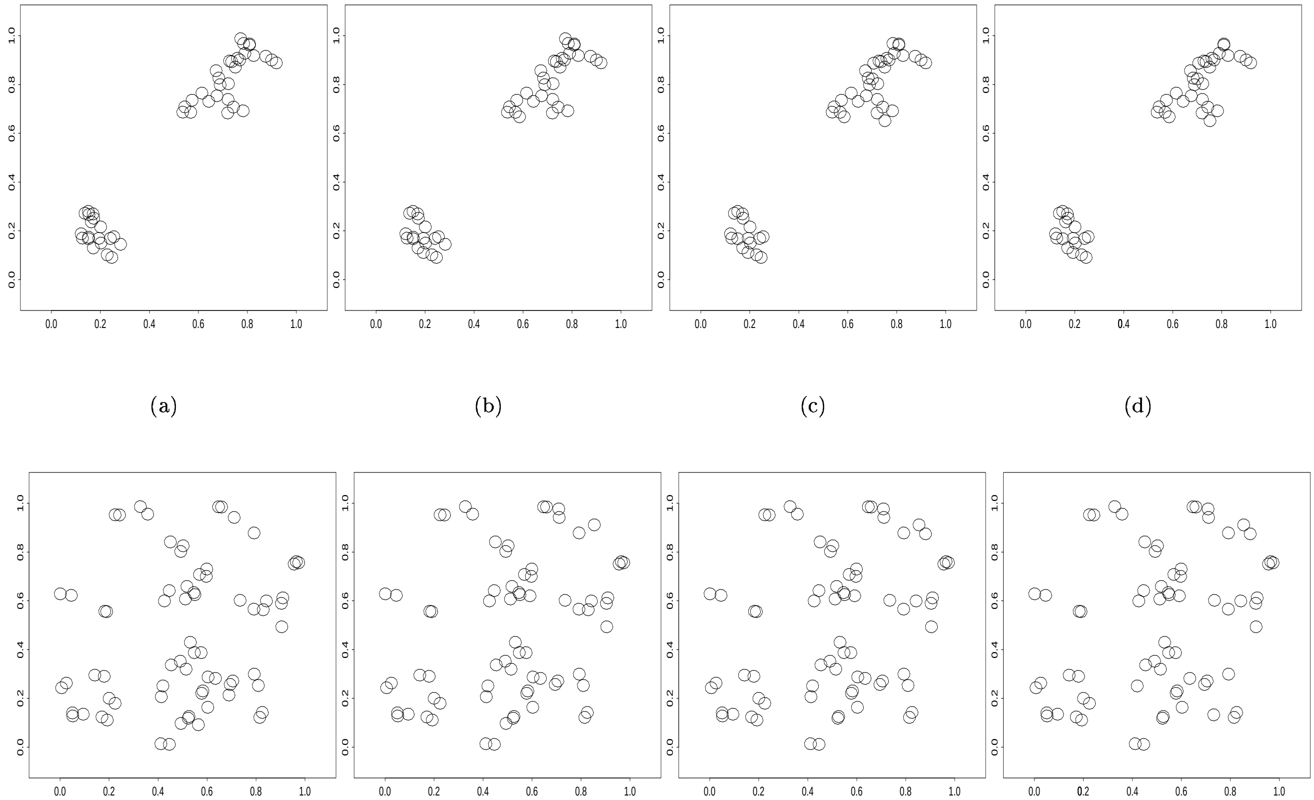
Figure 1: Initial configurations for our experiments with (a) conditional and (b) unconditional cluster process.

configuration $\mathbf{x}$ by replacing, with probability $q_i(\mathbf{x}, x^*)$, a randomly chosen member x_i of the realization with another candidate x^*, uniformly drawn on the state-space. Here $q_i(\mathbf{x}, x^*) = \min\{\frac{f(\mathbf{x} \setminus x_i \cup x^*)}{f(\mathbf{x}) + f(\mathbf{x} \setminus x_i \cup x^*)}, 1\}$. Note that $f(\mathbf{x})\mathrm{P}(\mathbf{x} \to \mathbf{x}') = f(\mathbf{x}')\mathrm{P}(\mathbf{x}' \to \mathbf{x})$, where $\mathbf{x}' = \mathbf{x} \setminus x_i \cup x^*$ with transition rates $\mathrm{P}(\mathbf{x} \to \mathbf{x}') = \frac{f(\mathbf{x}')}{f(\mathbf{x}) + f(\mathbf{x}')}$, so that the procedure satisfies detailed balance and is therefore stationary and time-reversible.

For an unconditional process, one may alternate, with equal probability between a birth, death or displacement. This is a Metropolis-Hastings-type algorithm [9] which proceeds from a configuration $\mathbf{x} = \{x_1, x_2, \ldots, x_n; n \geq 0\}$ as follows: if a birth is randomly selected to be considered, a new point is added to the configuration with probability $\min\{1, \frac{\beta\gamma^{m(\mathbf{x}^*) - m(\mathbf{x})}}{n+1}\}$, while in the potential death case, an existing randomly selected point is deleted with probability $\min\{1, \frac{n}{\beta\gamma^{m(\mathbf{x}^*) - m(\mathbf{x})}}\}$. Note that for a null configuration, $n = 0$ and $m(\phi) = 0$. The displacement step is performed for non-null configurations, conditional on the number of points in the current configuration, and using the birth-and-death process described above.

Performance under high interactions

Hastings' procedures have been seen to perform well when their acceptance rates range from between 30 to 70% [4]. Very low or high acceptance rates are indicators of poor choice of transition matrix. Also, a good procedure should permit mobility over the state-space so that it may be possible to move from one realization to another quickly. In this section, we study a test case for simulating from a conditional as well as an unconditional cluster process. In both cases, we consider an example where the model prefers realizations with very few connected components, i.e., when γ is very small. Our initial configuration was the same for all experiments on the conditional process reported in this paper. Conditional on a total of 50 points, the initial state was



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 2: Every tenth realization, after burn-in of 10,000 iterations, of (a–d) a conditional cluster process with $n = 50,\ \beta = 1,\ \gamma = 0.001,\ \rho = 0.5$ and (e–h) an unconditional cluster process with $\beta = 200,\ \rho = 0.05,\ \gamma = 0.01$.

chosen to have two clusters, each with 25 points (Figure 1a). For all unconditional experiments, our initial configuration had 100 points, evenly split into two clusters, as shown in Figure 1b.

Figures 2a–d display every tenth successive realization, obtained after a burn-in of 10,000 iterations, from a conditional birth-and-death process with equilibrium measure as a cluster process (1) with $\gamma = 0.001$ and a connected component graph defined by $\rho = 0.05$. Here, we have assumed that the number of points $n = 50$ is fixed. Since the choice of exact $\beta > 0$ is immaterial for a conditional process, we have set $\beta = 1$. The acceptance rate for this algorithm was only 4%. Further, there is very little mobility in the state-space, and the clusters almost remain at the same place. This is also the conclusion in Møller [16] when he says that the conditional Strauss process with small $\gamma > 0$ and large ρ or n produces clusters with highly regular patterns.

The low mobility is repeated for the unconditional case, depicted in Figures 2e–h which show every tenth realization, after a burn-in period of 10,000 iterations, from (1) with $\gamma = 0.01$, $\beta = 200$, and $\rho = 0.05$. Note that relative to the conditional process case, there is more movement here, but even that is somewhat slight.

The slow movement around the state-space in the illustrations in Figure 2 is a consequence of using a single-point updating algorithm in the presence of very strong interactions. This problem can be related to critical slowing down in lattice models such as Ising and Potts' models. The problem of slow mixing in the latter models has been effectively combated by Swendsen and Wang [22] by incorporating auxiliary bond variables designed to break interactions between points while updating. This is a remarkably simple multiple-component updating algorithm that combats critical slowing down [3], but performs poorly in cases when the model has fixed boundaries, or when the external field term of the model results in a multi-modal distribution for $\mathbf{x}$. The idea of partial decoupling, introduced by Higdon [13], is designed to address such situations by introducing a partial decoupling parameter δ to control the probability of bonding between any two points when using auxiliary variables.

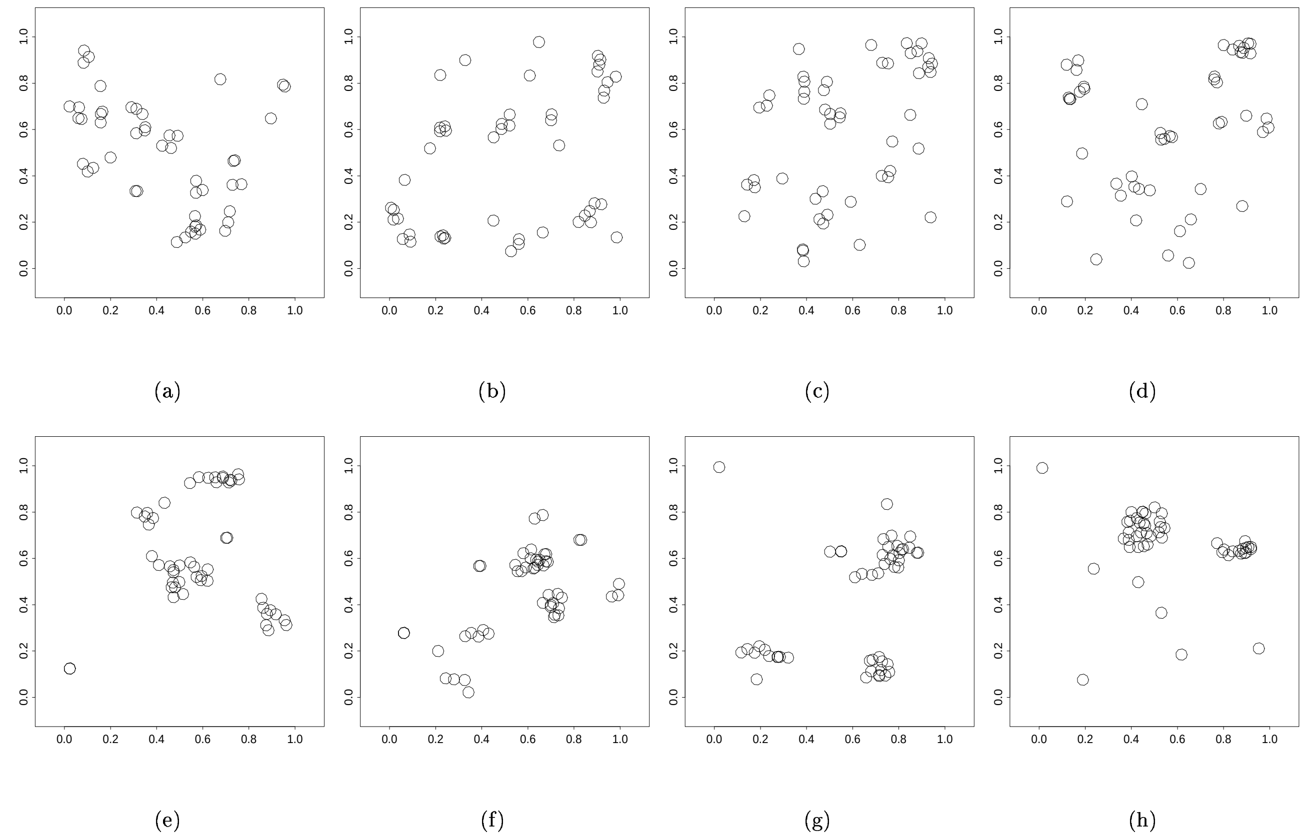


Figure 3: Every tenth realization, after burn-in of 10,000 iterations, from a conditional cluster process model, with $n = 50$, $\rho = .05$, $\gamma = .001$, using auxiliary variables with (a–d) no partial decoupling ($\delta = 1$) and (e–h) partial decoupling ($\delta = 10$).

3 Methodology

In order to extend the idea of partially decoupled auxiliary bond variables to simulations of cluster process defined by (1) and (2), we first set up auxiliary variables, conditional on the current realization $\mathbf{x}$. For any two points in $\{x_j,\ x_k\} \in \mathbf{x}$ in the ith cluster, we define independent random variables u_{ijk}s, uniform on $[0, b_{ijk}]$ where $b_{ijk} = \exp\{\frac{\delta 2 log \gamma}{n_i(n_i-1)}\ 1_{[c(x_k)=c(x_j)=i]}\}$. Then

$$f(\mathbf{x}|\mathbf{u}) \propto f_0(\mathbf{x}) \prod_{i=1}^{c_T(\mathbf{x})} \prod_{k \sim j} b_{ijk}^{1-\delta}\ 1_{[u_{ijk} \leq b_{ijk}^{1-\delta}]} \tag{3}$$

Notice that the marginal distribution of $\mathbf{x}$ is unaffected so that simulating from the joint distribution of $(\mathbf{x}, \mathbf{u})$ and restricting attention only to the $\mathbf{x}$ gives us realizations from the desired cluster process.

Equation 3 shows that points x_k and x_j in the ith cluster are bonded with probability $1 - \exp\{\frac{-2\ \delta log\ \gamma}{n_i(n_i-1)}\}$. Consider the new clusters $\zeta_{i_1}, \zeta_{i_2}, \ldots, \zeta_{i_{n^*}}$ induced by the bond variables $\mathbf{u}$'s. Note that these are different from the physical connected components earlier defined through the connected graph $\mathcal{G}$. Further, unlike the lattice models, cluster processes do not have a fixed structure, so that we can have bonds defined between points physically in different connected components of $\mathcal{G}$. We treat such points by moving all bonded points (as per u_{ijk}s) into the same physically connected component. For components ζ_{i_l} having such points, we choose a point at random from within ζ_{i_l} and move it inside the square determined by the maximum and minimum values of the points in ζ_{i_l}, recursively continuing until all u_{ijk}-bonded points are in the same physically connected component. Finally, similar to the suggestions in Higdon [13] for the Geman-McClure model [8], we shift each point in ζ_{i_l} by $r_l \sim U[-min(\zeta_{i_l}), 1 - max(\zeta_{i_l})]$, in both horizontal and vertical dimensions, so that all points are still within the state-space.

A more general structure would be to let the δ's vary for each bond. The choice of δ for our algorithm influences the performance of the simulations. For the special case when $\delta = 1$, the above algorithm is an extension of the Swendsen-Wang algorithm. For $\delta < 1$, interaction bonds are broken faster, while with increasing $\delta > 1$, the number of clusters stays low, but with good mobility around the state-space.

This algorithm can be substituted for the conditional point-process algorithm of Section 2.2. For the uncondi-

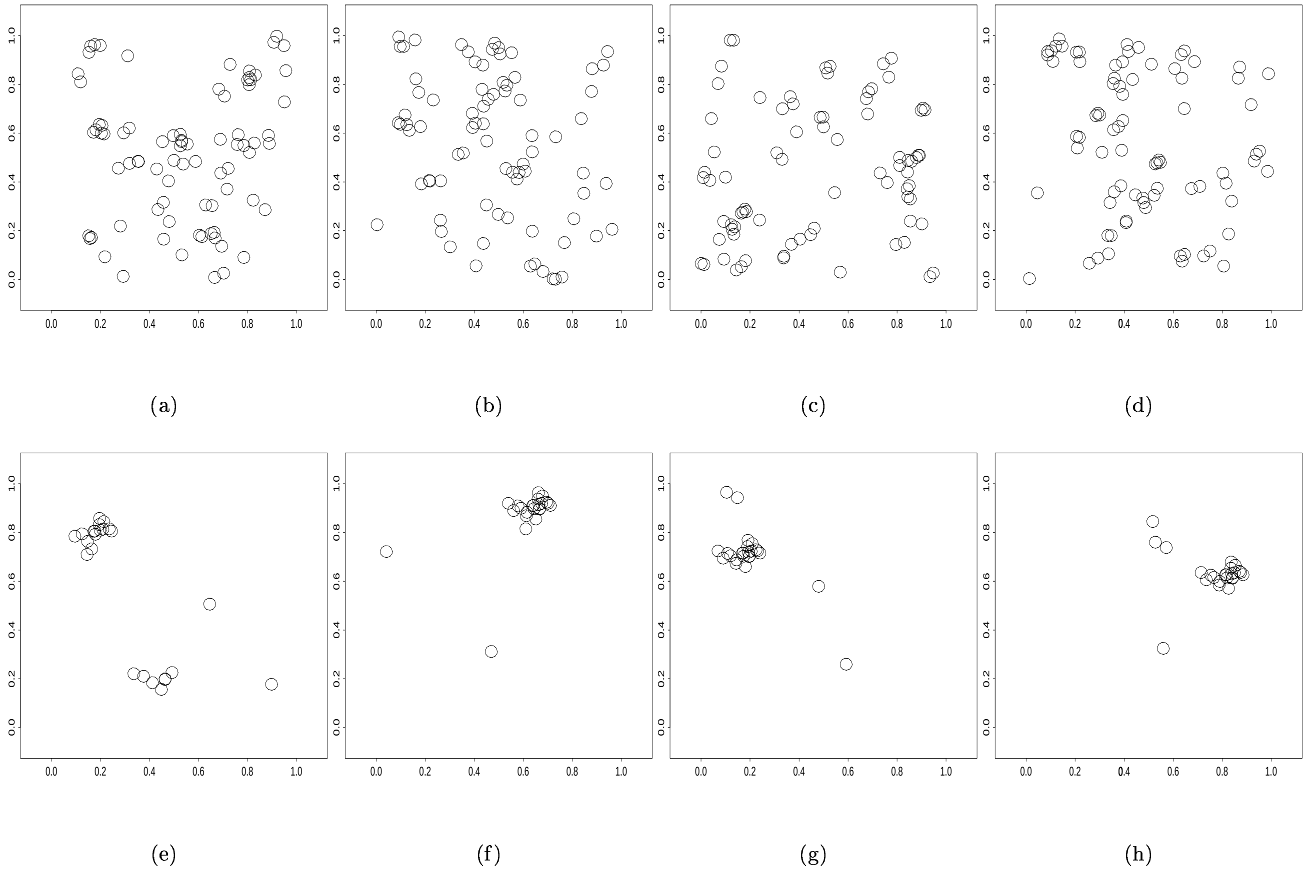


Figure 4: Every tenth realization, after burn-in of an unconditional cluster process with $\beta = 200$, $\rho = 0.05$, $\gamma = 0.01$, simulated using auxiliary variables and (a–d) no decoupling ($\delta = 1$) and (e–h) partial decoupling ($\delta = 120$).

tional simulations, the displacement step is replaced by the above. We next report results of some test simulations done to evaluate the performance of our algorithm for cluster processes.

Performance Evaluations

For the conditional cluster process, we started with the same initial configuration as in Figure 1a. Figure 3 presents the results of incorporating the suggested approach while simulating from the model with the same parameters as in Figures 2a–d. Figures 3a–d detail the case when we have no decoupling ($\delta = 1$). Note that this algorithm allows good mobility of the points in the state-space even though the number of clusters is too high for the model. Indeed, a study of the number of clusters vis-a-vis simulation number showed a very large number of clusters (ranging from 12 to 30) almost always. This is not particularly desirable for a model which prefers realizations with few connected components.

With smaller values of δ, the bonds were, as expected, even fewer and the fragmentation of clusters was more pronounced. Increasing the values of δ on the other hand produced fewer clusters, while also allowing for mobility around the state-space. Figures 3e–h show realizations from simulations done using partially decoupled auxiliary variables, with $\delta = 10$. Note that the number of clusters is few, while there is good movement of points around the state-space.

Figure 4 summarizes the results of simulations done using our suggested algorithm using auxiliary variables without (a–d) and with (e–h) partial decoupling ($\delta = 120$) for the unconditional cluster process, with the same initial configuration as in Figure 1b and parameter values as in Figures 2e–h. When $\delta = 1$, our simulations show that even though we had an initial configuration with 100 points, evenly split into two clusters, the number of clusters increases to around 30 within the first 1000 iterations and after that behaves similar to the auxiliary variables algorithm without partial decoupling for the conditional case. The number of points hovers around 70.

Varying δ affects the behavior of the number of points as well as clusters: for $\delta < 1$, the number of points and clusters is high while with low values of $\delta > 1$, the number of points reduces for the initial iterations and then stabilizes along with the number of clusters at small values. For larger values of δ, such as $\delta = 120$, the results of which are displayed in Figure 4e–h, the number of points stabilizes at moderate values. This also produces configurations with few clusters. The fact that the number of points is low to moderate is not a reflection of the displacement step: rather, it seems to be related to the low probability of having a birth when the current realization has very few clusters and when the interactions are very strong.

4 Discussion

Simulation procedures play an important role in making inferences for a cluster process, since the likelihood model is often known only up to a normalizing constant. However, despite statistical theory postulating realizations from the target cluster process after a long run of the appropriate birth-and-death process, in reality sometimes, this is too long. Moreover, in the presence of strong interactions, there is very low mobility around the state-space, so that even after burn-in, one may have to wait interminably long to obtain a reasonably representative sample from the process. In this paper, we have presented a computationally efficient alternative to the simple birth-and-death process, by extending the theory of partially decoupled auxiliary variables [13, 22] to spatial cluster processes.

For conditional cluster processes, the algorithm works well and differently with different choices of δ. In all cases, we get low correlations and high movement among successive realizations around the state-space. However, for low values of δ, the clusters are more fragmented while for high values, the preference for clustering is pronounced. For implementation of our scheme to unconditional cluster processes, we replace the displacement step of Geyer and Møller [9] by the above algorithm. The choice of δ greatly influences the simulation results. For $\delta \leq 1$, configurations had high number of points as well as clusters, but for $\delta > 1$, the number of groups was low while the total number of points was at best moderate. However, even without any statistical scheme to analyze the performance of our algorithm, it is visually quite clear that the inclusion of partially decoupled auxiliary variables improves performance over traditional single-point updating schemes such as birth-and-death process.

A few points are in order. The first pertains to the choice of δ. Varying δ has different effects on simulation. It may be useful to consider varying it over a range of values in some scheme for different iterations. Another issue pertains to the application of the current algorithm for unconditional cluster process simulation. It is our view, on the strength of preliminary experimental evidence that when the current configuration has very few clusters and strong interactions, a birth as in the Geyer and Møller [9] algorithm is very improbable, so that the number of points stays very low. Defining an *auxiliary process* to break the interactions while performing the birth step is a possible approach. Thus we see that while this paper does open new possibilities in the simulation of cluster point processes as in (1), there are a few issues requiring further attention.

References

[1] Baddeley, A. J. and Møller, J. (1989) Nearest-Neighbor Markov Point Processes and Random Sets. *Int. Stat. Rev.*, 57,89-121.

[2] Baddeley, A. J., van Lieshout, M. N. M. and Møller, J. (1996) Markov Properties of Cluster Processes. *Adv. Appl. Prob.*, 28,346-355

[3] Besag, J and Green, P. J. (1993) Spatial Statistics and Bayesian Computation. *J. Roy. Stat. Soc. Ser. B*, 55, 25-37

[4] Besag, J., Green, P., Higdon, D. and Mengersen, K. (1995) Bayesian Computation and Stochastic Systems (with Discussion).*Statistical Science*, 10,3-66.

[5] Chin, Y. C. and Baddeley, A. J. - On Connected Component Markov Point Processes. *Unpublished manuscript.*

[6] Cressie, N. A. C. (1993) *Statistics for Spatial Data.* Wiley.

[7] Dacey, M. F. (1966) A county seat model for the areal pattern of an urban system. *Geog. Rev.*, 56, 527-542

[8] Geman, S. and McClure, D. E. (1987), "Statistical Methods for Tomographic Image Reconstruction", *Bull. Int. Stat. Rev.*, 52:5-21.

[9] Geyer, C. J. and Møller, J. (1994) Simulation Procedures and Likelihood Inference for Spatial Point Processes. *Scand. J. Stat.*, 21:359-373

[10] Glass, L. and Tobler, W. R. (1971) Uniform Distribution of Objects in a Homogeneous field: Cities on a plain. *Nature*, 233, 67-68

[11] Gleason, H. A. (1920) Some Applications of the quadrat method. *Bulletin of the Torrey Botanical Club*, 47, 21-33

[12] Hastings, W. K. (1970) Monte Carlo Sampling Methods using Markov Chains and Their Applications. *Biometrika*, 57, 97-109

[13] Higdon, D.M. (1998) Auxiliary Variable Methods for Markov Chain Monte Carlo with Applications. *J. Amer. Stat. Assoc.*, 93,585-595

[14] Hodder, I and Orton, C. (1976) *Spatial Analysis in Archaeology.* Cambridge Univ. Press, London

[15] Kelly, F. P. and Ripley, B. D. (1976) A Note on Strauss' Model for Clustering. *Biometrika*, 63,357-360

[16] Møller, J. (1997) Markov Chain Monte Carlo and Spatial Point Processes . *Available online at http://www.math.auc.dc/~jm/MCMCpp.newandfinal.ps*

[17] Moyeed, R. A. and Baddeley, A. J. (1991) Stochastic Approximation of the MLE for a Spatial Point Pattern. *Scand. J. Stat.*, 18, 39-50

[18] Neyman, J. and Scott, E. L. (1958) Statistical approach to problems of cosmology. *J. Roy. Stat. Soc. Ser. B*, 20, 1-29

[19] Ripley, B.D. (1977) Modelling Spatial Patterns (with discussion). *J. Roy. Stat. Soc. Ser. B*, 39,172-212

[20] Ripley, B.D. (1981) *Spatial Statistics.* New York: Wiley

[21] Ripley, B. D. and Kelly, F. P. (1977) Markov Point Processes. *J. London Math. Soc.*, 15, 188-192

[22] Swendsen, R.H. and Wang,J.S. (1987) Nonuniversal Critical Dynamics in Monte Carlo Simulations. *Phys. Rev. Letters*, 58, 86-88