Modeling spatial patterns of mixed-species Appalachian forests with Gibbs point processes

By

Kevin C. Packard

Dissertation submitted to the faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Forestry

Philip J. Radtke, Chair
Harold E. Burkhart
Thomas R. Fox
Scotland Leman
David L. Loftis
Marion R. Reynolds, Jr.

February 5, 2009
Blacksburg, Virginia

Keywords: Gibbs point processes, Appalachian hardwood forests, spatial point pattern, Ripley's *K*-function, Markov Chain Monte Carlo, triplets process

Modeling spatial patterns of mixed-species Appalachian forests with Gibbs point processes

Kevin C. Packard

(ABSTRACT)

Stochastic point processes and associated methodology provide a means for the statistical analysis and modeling of the spatial point pattern formed from forest tree stem locations. Stochastic Gibbs point processes were explored as models that could simulate short-range clustering arising from reproduction of trees by stump sprouting and intermediate-range inhibition of trees that may result from competition for light and growing space. This study developed and compared three pairwise interaction processes with parametric models for 2nd-order potentials and three triplets processes with models for 2nd- and 3rd-order potentials applied to a mixed-species hardwood forest in the Southern Appalachian Mountains of western North Carolina. Although the 2nd-order potentials of both the pairwise interaction and triplets processes were allowed to be purely or partially attractive, the proposed Gibbs point process models were demonstrated to be locally stable. The proposed Gibbs point processes were simulated using Markov Chain Monte Carlo (MCMC) methods; in particular, a reversible-jump Metropolis-Hastings algorithm with birth, death, and shift proposals was utilized. Parameters for the models were estimated by a Bayesian inferential procedure that utilizes MCMC methods to draw samples from the Gibbs posterior density. Two Metropolis-Hastings algorithms that do this sampling were compared; one that estimated ratios of intractable normalizing constants of the Gibbs likelihood by importance sampling and another that introduced an

auxiliary variable to cancel the normalizing constants with those in the auxiliary variable's proposal distribution.

Results from this research indicated that attractive pairwise interaction models easily degenerate into excessively clustered patterns, whereas triplets processes with attractive 2^{nd}-order and repulsive 3^{rd}-order interactions are more robust against excessive clustering. Bayesian inference for the proposed triplets models was found to be very computationally expensive. Slow mixing of both algorithms used for the inference combined with the long iteration times limited the practicality of the Bayesian approach. However the results obtained here indicate that triplets processes can be used to draw inference for and simulate patterns of mixed-species Appalachian hardwood forests.

Acknowledgements

I would like to express my sincere gratitude to my adviser, Dr. Phil Radtke, who provided me the freedom to pursue my interest in bridging the abstract and mathematical world of stochastic point processes with forest biometrics for mixed hardwood forests. His patience, encouragement, guidance, and skill in programming have been invaluable for the production of this research. I wish also to express my gratitude to my committee of Drs. Harold Burkhart, Tom Fox, Scotland Leman, David Loftis, and Marion Reynolds for their insights and expertise in forestry and statistics.

I am also appreciative for several other faculty members here at Virginia Tech with whom I had fruitful discussions. In particular I wish to thank Dr. Carolyn Copenheaver of the Department of Forestry for reminding me of the ecological context of mixed Appalachian hardwoods; Dr. Dan Spitzner of the Department of Statistics (now at the University of Virginia) for teaching me the fundamentals of measure and integration theory, and Dr. Royce Zia, Department of Physics, for providing a meaningful context for what Gibbs distributions were designed to do.

Last but not least I would also like to thank current and past members of the "Forest Biometrics Club" at Virginia Tech: Jungkee Choi, Jason Henning, Charles Sabatia, Guillermo Trincado, Curtis VanderSchaaf, Nate Herring, and Matt Russell. Their willingness to discuss this research despite its mathematical intensity provided a great source of feedback and encouragement for a truly challenging thesis topic.

Dedication

To my dear wife Star

艳新

and daughter Anya

安雅

the love and joy in my life.

Table of Contents

List of Figures

List of Tables

1. Introduction

Mixed Appalachian hardwood forests constitute a large and ecologically diverse forest resource encompassing oak-hickory, oak-chestnut, and northern hardwood forest types (Fralish 2003, p. 103). Within the portion of the Southern Appalachian Assessment area in Virginia, West Virginia, North Carolina, Tennessee, South Carolina, Georgia, and Alabama there are approximately 5.4 million hectares of oak, 1.3 million acres of mixed mesophytic hardwood, and 249 thousand hectares of northern hardwood timberland (SAMAB 1996). In addition to being a large and economically important resource, it is very ecologically diverse, supporting over 2,250 species of vascular plants, 65 species of mammals, 175 species of birds, 80 species of reptiles and amphibians, and 25,000 species of invertebrates, including 51 terrestrial species listed as federally threatened or endangered (SAMAB 1996).

The diversity of mixed Appalachian hardwood forests coupled with the wide geographic distribution of these forests across mountainous topography necessitates flexible forest growth models that can account for multiple species coexisting in spatially complex ways. A variety of natural disturbances, and varying levels of intensity of forest management, will also have impact on species composition, tree growth and mortality (Copenheaver et al. 2006). By modeling growth and mortality at the resolution of individual trees, individual tree growth and yield models offer an attractive choice as a support tool for forest management decision making (Porté and Bartelink 2002). Particularly relevant are distance-dependent individual tree growth and yield models. These models provide a means to simulate forest growth and yield with a high degree of

detail by specifying forest spatial structure through each tree's coordinate location. Generally, a distance-dependent tree model is initialized through a user specified or model generated list of trees with coordinate locations. Tree growth is predicted by regression equations that relate the growth of individual tree diameter and height with forest or stand attributes, a measure of site quality, and the competitive effects from neighboring trees. Mortality and ingrowth may be specified deterministically or stochastically from stand and tree information. After a desired growth interval, a tree list is obtained that represents the projected forest or stand conditions. To estimate yield, tree volume or taper equations are applied to the final tree list, which can be expanded to a per unit land area basis (Avery and Burkhart 2002).

Currently there are no distance-dependant individual tree growth and yield models for mixed Appalachian hardwoods, although this approach has been applied in the United States to mixed northern hardwoods (Ek and Monserud 1974) and mixed bottomland hardwoods of the lower reaches of the Mississippi River (Zhao 2003; Zhao et al. 2006). Factors that have been cited as major drawbacks to the use of distance-dependent tree models are the high cost of development, the high cost of implementation, requiring detailed tree and stand data, and longer model execution times and "sophisticated computing equipment" needed to run the model (Avery and Burkhart 2002). However as landowner and public interests in Appalachian forests change, demand for information about these forests will increase to levels where it can only be satisfied by more sophisticated models that account for the spatial complexity and species diversity inherent to these forests.

Stochastic point process models provide an opportunity to reduce the implementation cost of distance-dependent individual tree models, by replacing costly measured tree stem locations (Reed et al. 1989) with stochastically generated maps of tree stems. Stochastic point processes have been used to describe and simulate patterns of trees since the pioneering works of Matérn (1960) and Warren (1972), but their implementation in conjunction with individual tree models remains limited (Pretzsch 1997; Stoyan and Penttinen 2000). One reason that stochastic processes have not seen wide use in forest growth models is that forestry applications of stochastic point processes are too simplistic (Comas and Mateu 2007); typically the process will describe patterns of one or two species of trees that show either strongly clustered or regular patterns. In mixed Appalachian hardwoods, stump sprouting is a common means of reproduction for several tree species where new shoots emerge from the cut stump of a parent tree (Wendel 1975; Beck 1977; Lamson 1983; Beck and Hooper 1986). Wendel (1975), for example, reported that after clearcutting 80% of all cut stumps had sprouted with an average of 22 – 42 sprouts per clump immediately following harvest. After 10 years of observation, an average of 3 – 7 sprouts per clump remained, varying by tree species. Sprouts that survive exist in clusters where stem locations are separated at very small distances corresponding to the ground line diameter of the parent trees. On the other hand, some studies suggest that over time density dependent mortality can result in regular spatial patterns of trees (Christensen 1977; Ward et al. 1996; Kenkel et al. 1997). In a study of an old-growth oak-hickory forest in the North Carolina piedmont, Christensen (1977) found a shift from random to a regular pattern over a 22 year study period, where trees in large diameter classes showed significant regularity in pattern.

Hence a stochastic process used to realistically model patterns of mixed Appalachian hardwoods should be flexible enough to account for small scale clustering from stump sprouting and intermediate scale spatial inhibition of trees from competition.

Aside from their use in forest growth and yield modeling, stochastic point process methodology provide a powerful means of quantifying the spatial structure of an observed pattern of trees. To this end, a number of tools are available including 1) summary statistics that look at distances to nearest neighboring events of a process from a "typical" event of the process (Clark and Evans 1954), 2) statistics for distances from random point locations to nearest neighboring events (Pielou 1959), 3) probability density functions of nearest-neighbor distances (Stoyan 2006), and 4) functions that assess second-order characteristics of a point process including the *K* and *L* functions (Ripley 1976, 1977) and the pair correlation function (Stoyan et al. 1993). The *K* and *L* functions, which look at the expected number of events found at increasing distances from a typical event of the process, are widely used in forestry and ecological research (Sterner et al. 1986; Getis and Franklin 1987; Moeur 1993; Haase 1995; Gavin and Peart 1997; Batista and Maguire 1998; Hanus et al. 1998; Pélissier 1998; McDonald et al. 2003; North et al. 2004; Doležal et al. 2006; Mason et al. 2007; Dagley 2008). The pair correlation function, which is proportional to the derivative of the *K*-function, also has seen use in forest and plant ecology studies (Penttinen et al. 1992; Gavrikov and Stoyan 1995; Pélissier 1998; Mason et al. 2007).

In addition to quantifying the spatial structure in observed point patterns, point process methodology provides a rich arena for stochastic models of trees, allowing for the incorporation of other characteristics such as tree diameter, height, or species to be

incorporated as “mark” variables (Mateu et al. 1998; Degenhardt 1999; Kokkila et al. 2002; Picard et al. 2009), the development of dynamic temporal changes in the spatial point structure (Renshaw and Särkkä 2001), or the incorporation of both of these features in marked spatio-temporal point processes (Särkkä and Renshaw 2006). Comas and Mateu (2007) provide a recent review of applications of point process models in forestry, with noteworthy examples including the use of inhomogeneous Poisson processes in tropical forests (Batista and Maguire 1998), and Cox processes in natural stands of longleaf pine (Rathbun and Cressie 1994).

Gibbs point processes are a particularly popular class of point process model, which has been used to model regular patterns resulting from spatial inhibition of trees through competition for light and growing space. Forestry examples of Gibbs point processes cover a variety of forest types, with applications to birch, spruce, and pine forests in southern Finland (Tomppo 1986; Särkkä and Tomppo 1998), a stand of Norway spruce (*Pice abies* Karst.) in Germany (Penttinen et al. 1992), tropical evergreen forests in southern India (Pélissier 1998), a research plot of *Quercus ilex* L. and *Pinus halepensis* L. located in the Mediterranean (Mateu et al. 1998), a subset of hickory trees (Stoyan and Stoyan 1998) from a research plot established in a oak-hickory stand located in the Duke Forest in Durham, North Carolina (Bormann 1953), a 150 year old beech (*Fagus sylvatica* L.) forest and a mixed ash (*Fraxinus excelsior* L.) and linden (*Tilia cordata* Mill.) forest, both located in Germany (Degenhardt 1999), a thinned plantation of Scots pine (*Pinus sylvestris* L.) in Germany (Degenhardt and Pofahl 2000), a mixed Scots pine-hardwood forest in Finland (Kokkila et al. 2002), and tropical lowland forests in Brazil (Neeff et al. 2005).

Despite the flexibility of stochastic point processes as models of forest tree pattern, Comas and Mateu (2007) note that applications in modeling forest dynamics remain limited. Some reasons that may have limited the use of stochastic point processes in forest growth models are the computational demands of these methods, the lack of biological mechanisms incorporated into the mathematical models (Comas and Mateu 2007), and not least, the mathematical rigor needed to define stochastic point processes, which can result in the primary literature being "not intelligible to foresters" (Stoyan and Penttinen 2000). Despite these obstacles, the use of spatial statistics continues to grow in forestry and in many other scientific disciplines (Schabenberger and Gotway 2005). As the power of spatial statistical tools improves, there will be greater opportunities for application of these methods to address forest management challenges.

The overall objective of this research was to apply stochastic Gibbs processes to simulate patterns of mixed Appalachian hardwood forests, with three specific objectives pertinent to this. Objective 1 was to explore the spatial structure of Appalachian forests by means of Ripley's *L*- function and the pair correlation function, using a dataset from a long-term ecological study of Appalachian hardwoods in western North Carolina. Specific to this objective is the assessment of the effect of tree diameter on spatial pattern by evaluating the spatial summary functions on subsets of the data using successively larger minimum tree diameter requirements. Objective 2 was the development of mathematically valid models for Gibbs point processes with 2^{nd}- and 3^{rd}-order interaction that can simultaneously account for small scale aggregation due to stump sprouting and large scale repulsion due to intertree competition. To demonstrate that the proposed Gibbs point processes can generate patterns comparable to observed patterns of

Appalachian hardwoods, Markov chain Monte Carlo (MCMC) methods were used to draw samples from the models across a matrix of parameter values representing spatial behavior from strong clustering to strong point repulsion. Objective 3 was to apply the MCMC methods for Bayesian statistical inference of Gibbs point processes proposed by Bognar (2004, 2005) and Møller et al. (2006) in order to explore the behavior of the proposed Gibbs models on simulated point patterns. Specific to this objective, the effect on the mixing of Bognar's and Møller et al.'s MCMC algorithms and the quality of fit of the proposed Gibbs point processes to observed and simulated data was assessed as a consequence of the following three factors: 1) point density of simulated test point patterns, 2) the spatial structure of simulated test patterns, comparing strongly attractive with purely repulsive spatial interaction, and 3) the effect of the choice of the interaction distance parameter of the Gibbs process. In addition to testing Bognar's and Møller et al,'s algorithms on simulated test patterns, both of these algorithms were used to test the fit of the proposed Gibbs point process models with 2^{nd}- and 3^{rd}-order interactions to observed Appalachian hardwood data.

Pursuant to these objectives, the organization of this thesis is as follows. First, the mathematical background of stochastic point processes is presented in Chapter 2. In addition to providing a primer on the fundamentals of measure theory necessary for a study of point processes, Chapter 2 presents Gibbs point processes formally and reviews both frequentist and Bayesian approaches to statistical inference for these models. Since the Bayesian approaches studied in this thesis depend heavily on MCMC methodology, a brief review of some key points of the theory of Markov chains and Metropolis-Hastings samplers are presented. Chapter 3 details the specific methodology used in this research,

describing the proposed Gibbs point process models and their stability properties. Chapter 3 additionally presents the Metropolis-Hastings algorithm developed by Geyer and Møller (1994), which is used to draw sample point patterns from Gibbs processes, along with the specific details of Bognar's and Møller et al.'s algorithms used to perform Bayesian inference. In Chapter 4, the results of all the statistical analyses are presented, with discussion of these results following in Chapter 5.

2. Background

2.1 Stochastic Point Processes

2.11 Measure theory fundamentals

In a stochastic spatial model of trees, the locations of tree stems can be considered as a pattern of points occurring in some subset of two- or three-dimensional Euclidean space which can represent the forest itself. By considering each tree stem location as an event determined from some random (stochastic) mechanism, the collection of all trees in a forest or stand can be mathematically viewed as a random set of point locations, or a stochastic point process. The development and use of point processes requires careful mathematical treatment however, so that the concept of randomly located points in space has meaning and is consistent with the laws of probability. To provide an illustration of this, consider a hypothetical disk with an arrow at the center that spins freely on an axis. The circle is marked like a clock with numbers from 1 to 12, and the arrow in the center has an infinitely small tip so that it can land precisely on any real number (or point) in that interval. If the arrow is spun, what is the probability of landing on any particular number? There are uncountably infinite real numbers in any finite length interval, so if we say that this probability is something greater than zero, then the sum probability of the infinite numbers must be infinite, which violates the notation that the sum probability of all random events must equal one. Yet on the other hand, if we say that any real number has probability zero, then all outcomes of the experiment have zero probability and their total is zero, not one. While this is a purely academic example, the notion of assigning

probabilities to points in a subset of the real line is not as simple as it may seem. By extension, the notion of probability for a collection of points in a subset of two-dimensional Euclidean space also requires care. To avoid pathological cases such as the spinner example, the fundamental theorems of probability and measure must be visited so that a stochastic point process can be meaningfully defined.

Generally speaking, a stochastic process is a sequence of random variables where each random variable can be referenced in another set. Stochastic point process treat points in time, space, or space-time as random variables where each is referenced against its position in time (temporal point processes), space (spatial point processes), or space-time (spatio-temporal point processes) respectively. As a model of tree spatial pattern, an observed collection of trees can be viewed as a realization of a spatial point process, where trees are treated as points that are referenced by their stem locations in two-dimensional Euclidean space. Treating tree locations as random variables in this way allows for statistical inference on the pattern as a whole, for instance by answering questions as to whether the observed pattern of trees is spatially random or not. Furthermore, the fitting of spatial point process models can allow for the simulation of the pattern of trees that are probabilistically consistent with the spatial structure found in observed tree stem maps.

To make a definition of stochastic point processes meaningful, the concept of a random variable becomes critical. A random variable may be defined as "a function from a sample space *S* into the real numbers" (Casella and Berger 2002, p.27). In familiar statistical settings reference is made to random variables without explicitly specifying the underlying experiment and its sample space that gave rise to the random variable. For

example, tree diameter may be thought of as a random variable (a function) that gives in its range only positive real numbers. We could have a sample of *n* tree diameters, and calculate the sample mean as the average of *n* independently and identically distributed random variables. The sample mean (a statistic) would in turn become another random variable with an associated probability distribution that we could use to make inference about the parameters of this probability distribution. For example we could perform a t-test on whether the mean diameter of the sampled trees is greater than some hypothesized value of interest. Notice that in this setup no specific reference has yet been made to the sample space *S* or any mathematical definition of the function that gives rise to tree diameter being a random variable. To understand stochastic point processes, it is necessary to make explicit these underpinnings.

To proceed, some basic understanding of measure theory (Billingsley 1995) is needed. When one speaks of a sample space of an experiment, one is referring to a universal set that contains all the possible outcomes of that experiment. Typically this set is denoted as Ω. From this set, one could define a class of subsets that would be of particular interest and would have useful mathematical properties. For instance, one such class, called a field, can be defined as being the class of subsets of Ω that contains Ω and is closed under the elementary set operations of finite unions and complementation. This can be expressed mathematically as the class of sets, $\mathcal{F}$:

i) $\Omega \in \mathcal{F}$

ii) $A \in \mathcal{F}$ implies $A^C \in \mathcal{F}$

iii) $A_1, A_2, \ldots, A_n \in \mathcal{F}$ implies $\bigcup_{i=1}^{n} A_i \in \mathcal{F}$

By extending property iii) from finite unions to countable unions, a new class of sets, called a σ-field or a σ-algebra, can be defined as the class of sets, $\mathcal{F}$:

i) $\Omega \in \mathcal{F}$
ii) $A \in \mathcal{F}$ implies $A^C \in \mathcal{F}$
iii) $A_1, A_2, \ldots \in \mathcal{F}$ implies $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

Now consider the notion of some class of sets ($\mathcal{A}$) of Ω that is of interest. From this class of sets a σ-field could be defined using elementary set operations. While this σ-field may be useful, it is not unique in the sense that there may be other σ-fields that may also contain $\mathcal{A}$ along with other sets. To have the "smallest" possible σ-field that still contains $\mathcal{A}$, define a new σ-field that is formed from an arbitrary (possibly uncountable) intersection of *all* σ-fields that contain $\mathcal{A}$. This new σ-field is known as the σ-field generated by $\mathcal{A}$ and is usually denoted as σ ($\mathcal{A}$) (Billingsley 1995, p.21). A very important example of a generated σ-field is the Borel subsets. Let Ω be defined as *d*-dimensional Euclidean space ($\mathbb{R}^d$), and define $\mathcal{R}^d$ to be the class of all mutually disjoint, half-open rectangles in $\mathbb{R}^d$. The Borel subsets may then be defined as the σ-field generated from $\mathcal{R}^d$, that is σ ($\mathcal{R}^d$). As described by Billingsley (1995, p.158-160), the Borel subsets are not restricted to half-open rectangles. In fact, the Borel subsets can also be generated from the classes of open or closed sets of $\mathbb{R}^d$. The implication is that the Borel subsets is a very large class of subsets of $\mathbb{R}^d$ and contains practically all subsets of interest.

The concept of a measure is fundamental to measure and probability theory. A measure is a set function μ defined on a σ-field $\mathcal{F}$ of subsets of Ω that satisfies the following properties (Billingsley 1995, p. 160):

i) $\mu(A) \in [0, \infty]$ for $A \in \mathcal{F}$
ii) $\mu(\varnothing) = 0$
iii) If $A_1, A_2, \ldots$ is a disjoint sequence of $\mathcal{F}$-sets and $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

then $\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i)$

A measure then is a set function on $\mathcal{F}$ that is non-negative and countably additive. In addition to these properties, some measures have other properties that will be useful. If the sample space can be written as a finite or countable sequence of sets in $\mathcal{F}$, that is $\Omega = \bigcup_{i=1}^{n} A_i$ or $\Omega = \bigcup_{i=1}^{\infty} A_i$, where all $A_i \in \mathcal{F}$, then the measure μ is said to be σ-finite if $\mu(A_i) < \infty$ for all A_i. When a measure has the property that $\mu(\Omega) = 1$, μ is said to be a probability measure. In such a case, the properties of the probability measure are known as the "Axioms of Probability" and μ is called a probability function (Casella and Berger 2002, p. 7). There are numerous types of measures that can be defined depending on the topological properties of Ω and $\mathcal{F}$. Two important examples of measures encountered frequently in the study of stochastic point processes are counting measure and Lebesgue measure. Counting measures are a class of measures that are a special case of atomic measures. An atomic measure, defined on the Borel subsets of $\mathbb{R}^d$, is a measure that puts all of its "mass" on a countably infinite collection of points $\{x_1, x_2, \ldots\} \in \mathcal{F}$. A simple

counting measure is an atomic measure that gives a value of 0 or 1 to each of those points (Stoyan et al. 1995). In contrast, Lebesgue measure puts its mass on regions of Ω rather than on points. Formally, Lebesgue measure, defined on the Borel subsets of $\mathbb{R}^d$, gives $\mu_L(A) = \prod_{i=1}^{d}(b_i - a_i)$, where A is a bounded rectangle given by $A = \prod_{i=1}^{d}[a_i, b_i]$ (Billingsley 1995) . To provide some intuition to Lebesgue measure, suppose $d = 1$. Then the set $A = [a_1, b_1]$ is simply a closed interval on the real line and $\mu_L(A) = b_1 - a_1$ is the length of the interval. If $d = 2$, then $A = [a_1, b_1] \times [a_2, b_2]$ is a closed rectangle given by the vertices (a_1, a_2), (b_1, a_2), (a_1, b_2), and (b_1, b_2). Lebesgue measure of this rectangle is its area, that is $\mu_L(A) = (b_1 - a_1) \times (b_2 - a_2)$. In a similar fashion, when $d = 3$, Lebesgue measure gives volume of cubes, and when $d \geq 4$, Lebesgue measure gives hypervolumes of hypercubes. To apply the concepts of simple counting measure and Lebesgue measure in a forestry context, consider a stem-mapped forest of 4 hectares that contains 2,000 trees. Consider each tree as a point within the forest, given by its stem location and define the set A to be the collection of these trees. Simple counting measure applied to A will give the value 2,000. Lebesgue measure of this same set is zero, however, since a point contains no area, regardless of how many points in A there are. Now define the set B to be the 4 hectares of land of the forest itself. Here simple counting measure of B is zero since simple counting measure is only non-zero for points in two-dimensional space. Lebesgue measure of B will be 4 hectares.

A sample space Ω combined with a σ-field of subsets of Ω is referred to as a measurable space, and is denoted $(\Omega, \mathcal{F})$. When combined with a measure μ, the triplet $(\Omega, \mathcal{F}, \mu)$ is called a measure space. A probability space is the particular measure space

$(\Omega, \mathcal{F}, \mathbb{P})$ where $\mathbb{P}$ is a probability function. With this structure in mind, a random variable can be defined as a measurable function that maps from a probability space to another measurable space (Θ, $\mathcal{G}$) where Θ is often taken to be the real numbers $\mathbb{R}$. The key requirement in the definition is that the random variable is a measurable function, a concept that is subtle in nature. Formally, a function $f : \Omega \to \Theta$ is measurable if for any set $A \in \mathcal{G}$, the inverse-image of A through f is an element of $\mathcal{F}$, that is $f^{-1}(A) \in \mathcal{F}$. Hence a measurable function takes elements from the σ-field of subsets of Ω and maps them to the σ-field of subsets of Θ. As shown in Billingsley (1995, p. 182-186), all continuous functions that map from $\mathbb{R}^d$ to $\mathbb{R}$ are measurable, as are functions constructed from sums and products of measurable functions. Consequently, most functions that one would typically use could be used to define a random variable mapping from a probability space to another measurable space.

Now consider the situation where a random variable f maps from a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to a measurable space (Θ, $\mathcal{G}$). Define the set function $P : \mathcal{G} \to \mathbb{R}$ by $P(A) = P\left(f^{-1}(A)\right)$ for any $A \in \mathcal{G}$. This function, referred to as an *induced* probability function (Casella and Berger 2002, p.28), is a probability measure (Billingsley 1995, p. 185-186) and it defines the probability distribution of the random variable f. In standard statistics texts, e.g. Casella and Berger (2002), the notation X is used instead of f, and with $\mathcal{G}$ taken to be the Borel subsets of the real line, σ ($\mathcal{R}^1$), the probability distribution is then $P(A) = \mathbb{P}(X \in A)$ for any $A \in \sigma(\mathcal{R}^1)$. If the set A is chosen to be $(-\infty, x]$, then $P(A) = P(-\infty, x] = \mathbb{P}(X \le x)$ defines the cumulative distribution function (cdf) of the random variable X (Billingsley 1995, p. 188).

Another concept important to the study of random variables is the notion of a probability density function (pdf). In typical applications of statistics, a probability density function for a continuous random variable X is the function f_X that satisfies

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt \qquad [1]$$

where $F_X(x)$ is the cumulative distribution function (cdf) of the random variable X (Casella and Berger 2002, p. 35). In the context of probability and measure theory, the concept of a density is more general. Specifically, the density f_X exists and satisfies Eq. 1 if the induced probability measure $P(A)$ of the random variable X is σ-finite and absolutely continuous with respect to Lebesgue measure, that is for any set $A \in \sigma\left(\boldsymbol{\mathcal{R}}^1\right)$, $\mu_L(A) = 0$ implies that $P(A) = 0$. The existence of f_X under these conditions is a result of the Radon-Nikodym Theorem, and the density f_X is referred to as a Radon-Nikodym derivative. More generally, the Radon-Nikodym Theorem states that given any σ-finite measures μ and v on a measurable space $(\Omega, \boldsymbol{\mathcal{F}})$, where v is absolutely continuous with respect to μ, then there exists a non-negative density $f_X : \Omega \to \mathbb{R}^*$ such that

$$v(A) = \int_A f_X(x)\,\mu(dx) \qquad [2]$$

for all $A \in \boldsymbol{\mathcal{F}}$ (Billingsley 1995, p. 422-423), where $\mathbb{R}^*$ is the extended real line.

The integral notation in Eq. 2 refers to an indefinite integral that is constructed in a different manner than the Riemann integrals familiar in calculus. To provide some understanding, let $(\Omega, \mathcal{F}, \mu)$ be a measure space and consider the case when f is restricted to being a non-negative measurable *simple* function:

$$f(x) = \sum_{i=1}^{n} a_i \; I_{E_i}(x) \qquad [3]$$

where $a_1, a_2, \ldots, a_n$ are non-negative real numbers, $E_1, E_2, \ldots, E_n$ form a finite decomposition of Ω into $\mathcal{F}$-measurable sets, and $I_{E_i}(x)$ is an indicator function so that $I_{E_i}(x) = 1$, if $x \in E_i$ and 0 otherwise. Then the definite integral of f with respect to μ is defined by Stroock (1994) as the following:

$$\int f(x)\,\mu(dx) = \sum_{i=1}^{n} a_i \; \mu(E_i) \qquad [4]$$

Hence, the definite integral is a finite sum of the product of non-negative real numbers and the value of the measure μ applied to the sets E_i.

To provide meaning for a probability density function as defined in Eq. 2, the definite integral must be extended from non-negative simple functions given by Eq. 3 to non-negative measurable functions generally. This is done by approximating any given non-negative measurable function $f(x)$ by a sequence of non-negative simple measurable

functions $g_n(x)$ for $n = 1, 2, \ldots$ where $0 \le g_n(x) \le f(x)$ and $\lim_{n\to\infty} g_n(x) = f(x)$ for all $x \in \Omega$. The definite integral for non-negative measurable functions can then be defined as (Stroock 1994):

$$\int f(x)\,\mu(dx) = \lim_{n\to\infty}\left[\int g_n(x)\,\mu(dx)\right] \qquad [5]$$

This definition of the definite integral is not the only one that can be constructed, however. Billingsley (1995, p. 199-200) defines the definite integral for non-negative measurable functions by setting the real number a_i as the infimum (greatest lower bound) of the function $f(x)$ evaluated over all x present in each set E_i, so that:

$$\int f(x)\,\mu(dx) = \sup\left\{\sum_{i=1}^{n}\left[\inf_{x\in E_i} f(x)\right]\mu(E_i)\right\} \qquad [6]$$

where the supremum (sup) means the least upper bound of the sums within the braces. Regardless of whether the definite integral is defined through Eqs. 5 or 6, Billingsley (1995, p. 203) shows that these definitions are equivalent, and both allow for a definition of the indefinite integral as

$$\int_A f(x)\,\mu(dx) = \int I_A\, f(x)\,\mu(dx) \qquad [7]$$

where I_A is the indicator function for $A \in \mathcal{F}$.

In comparison to the Riemann integral, an intuitive difference is that the measure theoretic integral partitions the range of *f* and approximates *f* with simple functions to those partitions, whereas Riemann integrals partition the domain of *f* (restricted usually to a closed interval [*a*, *b*] of the real numbers) and approximates *f* with Riemann sums. When a non-negative measurable function with domain $\Omega = [a,b] \subset \mathbb{R}$ is integrated with respect to Lebesgue measure, then the indefinite integral given in Eq. 7 is referred to as a Lebesgue integral and is equivalent to the Riemann integral (Billingsley 1995, p. 221-222):

$$\int_{[a,b]} f(x)\,\mu_L(dx) = \int_a^b f(x)\,dx \qquad [8]$$

While the foregoing excursion into integration theory may seem to distract from an understanding of stochastic point processes, an important feature of measure theoretic integrals is that their definition no longer relies on the function *f* having its domain be on the real line. This allows for integrals to be applied to functions defined on very general measure spaces including those that give rise to stochastic point processes.

2.12 Point process definition and characterizations

With all of this technical machinery in place, now consider the definition of a stochastic point process given by (Stoyan et al. 1995, p.99) as a random variable that can take on values in a measurable space consisting of 1) the family of all simple, locally finite sequences of points in *d*-dimensional Euclidean space, and 2) the smallest σ-algebra on that family that makes measurable mappings onto the bounded Borel subsets. Formally, a stochastic point process Φ is a random variable that maps from the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to a measurable space $(\mathbb{N}, \mathcal{N})$, where $\mathbb{N}$ is the set of all simple and locally finite sequences of points $\mathbf{x} = \{x_1, x_2, \ldots, x_n\}$ in $\mathbb{R}^d$ and $\mathcal{N}$ is the smallest σ-field on $\mathbb{N}$ that makes the mappings measurable.

Applying this definition to patterns of trees, the "universal" set Ω may be taken as two-dimensional Euclidean space, $\mathbb{R}^2$, where the trees are observed in some forest $F \subseteq \mathbb{R}^2$. The random sequences defined above are subject to two regularity conditions: i) sequences **x** are simple – each point in a sequence has its own unique location in $\mathbb{R}^d$ and ii) sequences are locally finite – any bounded subset of $\mathbb{R}^d$ contains at most a finite number of points. In plain language, these conditions say that within any fixed region of the forest *F*, there are at most a finite number of trees, and no two trees occur at the same location.

Due to historically different approaches to stochastic process theory, there are two alternative characterizations of a spatial point process. One characterizes the spatial point process as a random set, that is a random selection of one **x** from $\mathbb{N}$ (Stoyan et al. 1995).

The other characterizes the spatial point process Φ as a random measure $N(B)$, counting points x in Borel subsets B of $\mathbb{R}^d$, that is

$$N(B) = \sum_{x \in \Phi} I_B(x) \qquad [9]$$

where I_B is the indicator function for Borel set B. In this characterization, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathbb{G}$ be the set of all boundedly finite counting measures, where a counting measure, $N(B)$, is said to be boundedly finite if $N(B) < \infty$ for any Borel set B. Define $\mathcal{G}$ as the σ-field generated from the class of sets $\{N \in \mathbb{G} : N(B) = n\}$ so that $(\mathbb{G}, \mathcal{G})$ is a measurable space. A stochastic point process can then be defined as a measurable mapping from $(\Omega, \mathcal{F}, \mathbb{P})$ to $(\mathbb{G}, \mathcal{G})$ (Cressie 1993, p.619). Notice in this definition the stochastic point process is a random counting measure, and is not restricted to simple counting measures as described previously. While Daley and Vere-Jones (1988) provide background on the theory of general counting measures, for the purposes here, $N(B)$ is assumed to be a simple counting measure and the stochastic point processes of interest are also simple (Daley and Vere-Jones 1988, pp.197-199) .

The presence of these alternate characterizations of stochastic point processes unfortunately results in a considerable array of notation used to develop point process theory (Stoyan et al. 1995). Regardless of how spatial point processes are characterized, the key point is that a spatial point process is a random variable. Being a random variable, the stochastic point process induces a probability measure $P(A) = \mathbb{P}(\Phi \in A)$ for any $A \in \mathcal{G}$ that is the probability distribution of the process. The induced probability measure

allows for a meaningful definition of probability for a collection of tree locations in space. For example, if $B_1, B_2, \ldots, B_k$ represent k bounded regions of a forest, then $\mathbb{P}(N(B_1) = n_1, N(B_2) = n_2, \ldots, N(B_k) = n_k)$ is the joint probability that $B_1, B_2, \ldots, B_k$ contain $n_1, n_2, \ldots, n_k$ trees respectively. This joint probability is called the finite-dimensional distribution and it uniquely determines the probability distribution of a given point process (Stoyan et al. 1995, p. 100).

Two important properties that a spatial point process may have relate to how the probabilities of the point process change if the point pattern is subject to translations and rotations. A point process is said to be stationary if the probability of the process does not change if the point pattern is shifted (translated) by some finite amount. Mathematically, represent the point process Φ as the random selection of a finite sequence $\mathbf{x} = \{x_1, x_2, \ldots, x_n\}$ in $\mathbb{R}^d$. Define the point process Φ_h as the random selection of a finite sequence $\mathbf{x} + \mathbf{h} = \{x_1 + h_1, x_2 + h_2, \ldots, x_n + h_n\}$ in $\mathbb{R}^d$, where the point configuration $\mathbf{x}$ has been translated by the vector $\mathbf{h} \in \mathbb{R}^d$. The point process Φ is intrinsically stationary if $\mathbb{P}(\Phi \in Y) = \mathbb{P}(\Phi_h \in Y)$ for all $Y \in \mathcal{N}$. Similarly, if a point process is subject to a rotation about the origin and the probabilities of the stochastic point process do not change, the process is said to be isotropic. When a point process is both isotropic and stationary the stochastic process is then said to be motion-invariant (Stoyan et al. 1995, p. 102).

2.13 Moment measures and intensity functions

Since the count of the number of trees in a region $N(B)$ is a random variable, moments may be calculated for it, analogous to moments such as the mean and variance of continuous random variables. For example the first moment, also known as the mean measure or intensity measure of the process is defined (Stoyan et al. 1995, p. 102) as:

$$\Lambda(B) = \mathbb{E}\left[N(B)\right] = \int N(B) P(dN) \qquad [10]$$

where $\Phi(B)$ is the number of points of the process in Borel set B as defined in Eq. 9, $\mathbb{E}$ is the expectation operator, and P is the probability distribution of the process. If the intensity measure is σ-finite and is absolutely continuous with respect to d-dimensional Lebesgue measure, then the Radon-Nikodym Theorem implies that there is a density $\lambda(x)$ that satisfies

$$\Lambda(B) = \int_B \lambda(x)\, \mu_L(dx) = \int_B \lambda(x)\, dx \qquad [11]$$

This density is referred to as the intensity function or the first-order intensity of the point process (Schabenberger and Gotway 2005, p. 83; Baddeley 2007, p. 27). In the case of a stationary point process, the first-order intensity function is constant and is written $\lambda(x) = \lambda$.

Analogous to the k^{th} moment of continuous and discrete random variables (Casella and Berger 2002, p. 59), the k^{th} moment measure for a point process can be defined as:

$$\begin{aligned}\Lambda_k\left(B_1 \times B_2 \times \ldots B_k\right) &= \mathbb{E}\left[N\left(B_1\right)N\left(B_2\right) \ldots N\left(B_k\right)\right] \\ &= \int N\left(B_1\right)N\left(B_2\right) \ldots N\left(B_k\right)P\left(dN\right)\end{aligned} \qquad [12]$$

An important case is when $k = 2$ so that the variance of $N(B)$ is given by

$$\begin{aligned}Var\left(N\left(B\right)\right) &= \mathbb{E}\left[N\left(B\right)N\left(B\right)\right] - \left(\mathbb{E}\left[N\left(B\right)\right]\right)^2 \\ &= \Lambda_2\left(B \times B\right) - \left(\Lambda\left(B\right)\right)^2\end{aligned} \qquad [13]$$

Similarly the covariance of $N(B_1)$ and $N(B_2)$ is given by

$$\begin{aligned}Cov\left(N\left(B_1\right), N\left(B_2\right)\right) &= \mathbb{E}\left[N\left(B_1\right)N\left(B_2\right)\right] - \mathbb{E}\left[N\left(B_1\right)\right]\mathbb{E}\left[N\left(B_2\right)\right] \\ &= \Lambda_2\left(B_1 \times B_2\right) - \Lambda\left(B_1\right)\Lambda\left(B_2\right)\end{aligned} \qquad [14]$$

Just as moment measures are analogous to moments of continuous or discrete random variables, another class of measures, factorial moment measures, are the point process analogues to factorial moments (Casella and Berger 2002, p.83). In particular the second factorial moment measure can be defined as (Baddeley 2007, p.33)

$$\begin{aligned}\Lambda_{[2]}(B_1 \times B_2) &= \mathbb{E}\left[N(B_1)N(B_2)\right] - \mathbb{E}\left[N(B_1 \cap B_2)\right] \\ &= \Lambda_2(B_1 \times B_2) - \Lambda(B_1 \cap B_2)\end{aligned} \quad [15]$$

Notice that when evaluated on the product set $B \times B$ the second factorial moment measure is

$$\begin{aligned}\Lambda_{[2]}(B \times B) &= \mathbb{E}\left[N(B)N(B)\right] - \mathbb{E}\left[N(B \cap B)\right] \\ &= \mathbb{E}\left[N(B)^2\right] - \mathbb{E}\left[N(B)\right] \\ &= \mathbb{E}\left[N(B)\left(N(B)-1\right)\right]\end{aligned} \quad [16]$$

hence the name "factorial moment measure". While factorial moment measures can be generalized to higher-orders and for non-negative measurable functions, they will not be considered here. If the second factorial moment measure is σ-finite and is absolutely continuous with respect to 2*d*-dimensional Lebesgue measure, then the Radon-Nikodym Theorem implies that there is a density $\rho_2(x_1, x_2)$ that satisfies

$$\Lambda_{[2]}(B_1 \times B_2) = \int_{B_1}\int_{B_2} \rho_2(x_1, x_2)\, dx_1\, dx_2 \quad [17]$$

This density is known as the second-order product density (Stoyan et al. 1995, p. 112) or the second-order intensity function (Schabenberger and Gotway 2005, p. 99) and may be interpreted as giving the joint probability that there will be two points of the process at the specified locations x_1 and x_2 (Baddeley 2007, p. 33). When a point process is

stationary and isotropic, ρ_2 will only depend upon the distance between points x_1 and x_2, that is $\rho_2(x_1, x_2) = \rho_2(\|x_1 - x_2\|)$. In this case, Stoyan and Stoyan (1994, p. 249) propose a normalization of the second-order product density by dividing it by the square of the first-order intensity

$$g(r) = \frac{\rho_2(x_1, x_2)}{\lambda^2} = \frac{\rho_2(\|x_1 - x_2\|)}{\lambda^2} = \frac{\rho_2(r)}{\lambda^2} \qquad [18]$$

The function $g(r)$ is known as the pair correlation function and is proposed as a means to describe the second-order characteristics of a point process (Stoyan and Penttinen 2000).

2.13 Conditioning and Palm distributions

It is frequently desired to condition on an event of the point process occurring at a particular location. For example we may wish to know the probability of finding another event of the process at a distance r from a given event x or the probability of counting a specific number of events with a distance r from a given event. One difficulty in formulating such conditional probabilities is that the conditioning event that $x \in \Phi$ typically has probability zero. The need for meaningful conditional probabilities for stochastic point processes leads to the notion of Palm distributions, which roughly speaking are probability distributions of the "typical point" of the point process. One approach to defining Palm distributions is through the concept of Campbell measure.

To introduce Campbell measure, consider the calculation of the expected number of events in Borel set *B* by the intensity measure defined in Eq. 10. In this case the expected value was calculated for the random variable $N(B)$. Now suppose that it is desired to calculate the expected values for non-negative measurable functions $f(x)$ defined on the points of the point process, $x \in \Phi$. In other words replace the indicator function in Eq. 9 with a non-negative measurable function *f* so that

$$\mathbb{E}\left(\sum_{x\in\Phi} f(x)\right) = \int_{\mathbb{N}} \sum_{x\in\mathbf{x}} f(x) P(d\mathbf{x}) \qquad [19]$$

This expected value can be simplified through the Campbell Theorem (Stoyan et al. 1995, p. 103; Baddeley 2007, p. 28) as

$$\mathbb{E}\left(\sum_{x\in\Phi} f(x)\right) = \int f(x)\,\Lambda(dx) \qquad [20]$$

where Λ is the intensity measure of the process. The Campbell Theorem is an important theoretical tool in its own right, allowing for ease of calculation of expected values such as Eq. 19. For the purpose of understanding Palm distributions however, consider a more complex form of *f* as a non-negative function defined on the product space $\mathbb{R}^d \times \mathbb{N}$. The Campbell measure $\mathcal{C}$ can then be defined as a measure on the measurable space $\left(\mathbb{R}^d \times \mathbb{N}, \sigma\left(\mathbb{R}^d\right) \times \mathcal{N}\right)$ as (Stoyan et al. 1995, p. 112)

$$E\left(\sum_{x\in\Phi} f(x,\Phi)\right) = \int \sum_{x\in\mathbf{x}} f(x,\mathbf{x}) P(d\mathbf{x}) = \int f(x,\mathbf{x}) C\left(d(x,\mathbf{x})\right) \quad [21]$$

Alternatively the Campbell measure $C(B\times Y)$ can be defined as a measure on the measurable space $\left(\mathbb{R}^d \times \mathbb{G}, \sigma\left(\mathbb{R}^d\right)\times \mathcal{G}\right)$ (Cressie 1993, p.630)

$$C(B\times Y) = \mathbb{E}\left[N(B) I_Y\right] = \int N(B) I_Y P(dN) = \int_Y N(B) P(dN) \quad [22]$$

for all Borel sets B of $\mathbb{R}^d$ and sets $Y \in \mathcal{G}$ so that the Campbell measure is the expected value of the product of the number of events of the process in B and the indicator function denoting whether set Y occurs or not.

Now if the intensity measure Λ of the point process exists and is σ-finite, then the Campbell measure will be σ-finite. It is also absolutely continuous with respect to Λ, so the Radon-Nikodym Theorem implies that there exists a density P_x that satisfies

$$C(B\times Y) = \int_B P_x(Y)\Lambda(dx) \quad [23]$$

This density is the Palm distribution of the point process and provided "regular conditional probabilities" are satisfied, Eq. 23 defines the Palm probability measure $\mathbb{P}_x$ on (Ω, $\mathcal{F}$) so that $\mathbb{P}_x\left(\Phi \in Y\right) = P_x\left(Y\right)$ that can be defined as giving the conditional probability that event Y occurs given that the point process Φ has an event at x (Stoyan et al. 1995, pp. 117-121; Baddeley 2007, p. 44-46).

An additional distribution that is useful is the reduced Palm distribution, where the event x is removed from the distribution itself. Formally, let $\mathbf{x} - \delta_x$ denote the point configuration formed from $\mathbf{x}$ with the point x removed from it. Then the reduced Campbell measure $C^!$ may be defined (Stoyan et al. 1995, p. 112) through

$$\begin{aligned}\mathbb{E}\left(\sum_{x\in\Phi} f\left(x,\Phi-\delta_x\right)\right) &= \int\sum_{x\in\mathbf{x}} f\left(x,\mathbf{x}-\delta_x\right)P\left(d\mathbf{x}\right) \\ &= \int f\left(x,\mathbf{x}\right)C^!\left(d\left(x,\mathbf{x}\right)\right)\end{aligned} \qquad [24]$$

Alternatively the reduced Campbell measure can be defined (Baddeley 2007, p. 53) as

$$\begin{aligned}C^!\left(B\times Y\right) &= \mathbb{E}\left[\sum_{x\in\Phi} I_B\left(x\right)I_Y\left(\Phi-\delta_x\right)\right] \\ &= \int\sum_{x\in\mathbf{x}} I_B\left(x\right)I_Y\left(\mathbf{x}-\delta_x\right)P\left(d\mathbf{x}\right)\end{aligned} \qquad [25]$$

Again assuming that the intensity measure of the process exists and is σ-finite then by the Radon-Nikodym Theorem there exists a density that satisfies

$$C^!(B\times Y)=\int_B P_x^!(Y)\Lambda(dx) \qquad [26]$$

This density, $P_x^!(Y)=\mathbb{P}_x(\Phi-\delta_x\in Y)$, is known as the reduced Palm distribution of the point process, defined for all $Y\in\mathcal{F}$. With the notions of Palm and reduced Palm distributions in place, conditional probability for events Y can be defined with or without the point x included in the respective conditional probability distributions.

One application of reduced Campbell measure is to define the notion of the conditional intensity function, which is a conditional analogue to the intensity function given by Eq. 11. For the conditional intensity function to be defined, the point process Φ must satisfy the condition that the reduced Campbell measure of the process is absolutely continuous with respect to the product measure $\mu_L\times P$, where μ_L is d-dimensional Lebesgue measure and P is the probability distribution of Φ. When this holds, the conditional intensity function λ^*, mapping from the product space $\mathbb{R}^d\times\mathbb{G}$ to the positive real numbers $\mathbb{R}^+$, is given by the Radon-Nikodym Theorem as (Baddeley 2007, p. 53)

$$C^!(B\times Y)=\int_B \mathbb{E}\left[\lambda^*(u,\Phi)I_Y(\Phi)\right]du \qquad [27]$$

Combining this definition with Eq. 25 implies

$$\mathbb{E}\left[\sum_{x\in\Phi} I_B(x)I_Y(\Phi-\delta_x)\right]=\int_B \mathbb{E}\left[\lambda^*(u,\Phi)I_Y(\Phi)\right]du \qquad [28]$$

So by extension to any integrable function f: $\mathbb{R}^d \times \mathbb{G} \to \mathbb{R}^+$, the conditional intensity function must satisfy (Baddeley 2001, 2007)

$$\mathbb{E}\left[\sum_{x\in\Phi} f\left(x, \Phi - \delta_x\right)\right] = \int_{\mathbb{R}^d} \mathbb{E}\left[\lambda^*\left(u, \Phi\right) f\left(x, \Phi\right)\right] du \qquad [29]$$

This result is sometimes referred to as the "Nguyen-Zessin formula", after the work of Nguyen and Zessin (1979), and the conditional intensity function itself is also referred to as the Papangelou conditional intensity function, after the pioneering work of Papangelou (1974). A more detailed description of the relationships of Campbell and Palm measures to the conditional intensity function and Papangelou kernels is provided by Kallenberg (1984) . In a heuristic sense, the conditional intensity function can be thought of (Van Lieshout 2000, p. 39; Baddeley 2007) as a function that provides the infinitesimal conditional probability of finding a point of a point process at a location *u*, given that the process has a point configuration **x** otherwise. That is

$$\lambda^*\left(u; \mathbf{x}\right) du = \mathbb{P}\left(N\left(du\right) = 1 \,|\, \Phi \cap \left(du\right)^{\mathrm{C}} = \mathbf{x} \cap \left(du\right)^{\mathrm{C}}\right) \qquad [30]$$

where the notation $(du)^{\mathrm{C}}$ means the complement of the infinitesimally small region *du*.

Another application of Palm distributions and associated Campbell measures develops the concept of the so-called "reduced second-moment measure". Suppose it is desired to calculate the expected number of events in a Borel set *B* given an existing event of the process at *x*. Furthermore since the event $x \in B$ is given, it is desired to count

the number of additional events, that is all the events in B excluding x itself. This expectation can be meaningfully defined in terms of the reduced Palm distribution, that is $\int N(B) P_x^!(dN)$. For a stationary point process there exists a measure $\mathcal{K}$, known as the reduced-second moment measure that satisfies

$$\lambda \mathcal{K}(B) = \int N(B) P_x^!(dN) \qquad [31]$$

where λ is the first-order intensity of the process (Stoyan et al. 1995, p. 126). Now suppose that the point process is also isotropic and stationary. Then taking the Borel set B to be the open ball of radius r centered at x, denoted $b(x, r)$ then the K-function can be defined as

$$\begin{aligned} K(r) &= \mathcal{K}(b(x,r)) \\ &= \lambda^{-1} \int N(b(x,r)) P_x^!(dN) \end{aligned} \qquad [32]$$

The K-function was first proposed by Ripley (1976, 1977), and has been used extensively in modeling forest tree patterns (Comas and Mateu 2007). The reduced-second moment measure $\mathcal{K}$, has an important relationship to the second-order product density ρ_2 for stationary and isotropic processes (Stoyan et al. 1995, p. 126):

$$\lambda^2 \mathcal{K}(B) = \int_B \rho_2(x)\, dx \qquad [33]$$

When $B = b(x, r)$, then transformation of Eq. 33 into polar coordinates implies

$$\begin{aligned}\lambda^2 K(r) &= \int_0^{2\pi}\int_0^{r} \rho_2(t)\, t\, dt\, d\theta \\ &= \int_0^{2\pi} d\theta \int_0^{r} \rho_2(t)\, t\, dt \\ &= 2\pi \int_0^{r} \rho_2(t)\, t\, dt\end{aligned} \qquad [34]$$

so that the K-function can be defined for a homogeneous point process as (Diggle 2003):

$$K(r) = \frac{2\pi}{\lambda^2}\int_0^{r} \rho_2(t)\, t\, dt \qquad [35]$$

This result, along with Eq. 18 implies that the pair correlation function and the K-function have the following relationship:

$$g(r) = \frac{1}{2\pi r}\frac{d\left(K(r)\right)}{dr} \qquad [36]$$

where $r = \|x_1 - x_2\|$, the distance between two events of the process x_1 and x_2 .

2.14 The homogeneous Poisson Process

To be able to quantify the spatial dependencies in a stochastic point process, reference must be made to the concept of a point process that is completely devoid of any spatial structure. This concept is the notion of complete spatial randomness (CSR), which is defined in terms of the homogeneous Poisson point process. The homogeneous Poisson process must satisfy two properties: i.) The number of events of the process in any bounded Borel set B follows the Poisson distribution with intensity measure $\Lambda(B) = \beta\,\mu_L(B)$ for some constant β, and ii.) Given B_1 and B_2 disjoint sets, $N(B_1) = n_1$ and $N(B_2) = n_2$ are independent of each other. Property i. implies that the probability distribution of the homogeneous Poisson process is

$$\mathbb{P}(N(B)=n)=\frac{\exp\left[-\beta\,\mu_L(B)\right]\left(\beta\,\mu_L(B)\right)^n}{n!}, \qquad n = 0, 1, 2, \ldots \qquad [37]$$

The homogeneous Poisson process is a stationary and isotropic process, and it follows that the first order intensity of the process is $\lambda(x) = \beta$. In forestry terminology, this implies that forest density, expressed in terms of the number of trees per unit area is constant and does not vary spatially within the forest or stand of interest.

As a point of reference to Gibbs point processes, the homogeneous Poisson process has a probability density function of the form

$$f(\mathbf{x}) = \alpha\beta^{n(\mathbf{x})} \qquad [38]$$

for point configurations $\mathbf{x} \in \mathbb{N}$, where $n(\mathbf{x})$ is the random number of points in **x**, and α is a normalizing constant that allows the density function to integrate to one. In the case of a homogeneous Poisson process defined on a bounded region (i.e. a forest) $F \subset \mathbb{R}^d$, the normalizing constant can be found explicitly as $\alpha = \exp\left[(1-\beta)\mu_L(F)\right]$ (Baddeley 2007, p. 62-63)

To provide some illustration of CSR, a hypothetical square forest area representing an area of 6,400 square meters was created (Fig. 1). Within this square, n = 269 points were chosen from a Poisson distribution with mean intensity of approximately 420 trees per hectare. The coordinate X and Y locations of these points were each obtained randomly from a uniform (0, 80) distribution, following the procedure of Schabenberger and Gotway (2005, p. 418). To the human eye, the resulting pattern shows no spatial trends in tree density in any region of the square. A few small clusters of 2-3 trees are apparent even though the data are generated by a uniform random process. Visual inspection of point patterns by itself is not sufficient to determine if any spatial dependency is present in an observed point pattern. Instead, spatial dependence can be statistically detected using summary functions of the spatial pattern, which are compared against the homogeneous Poisson process.

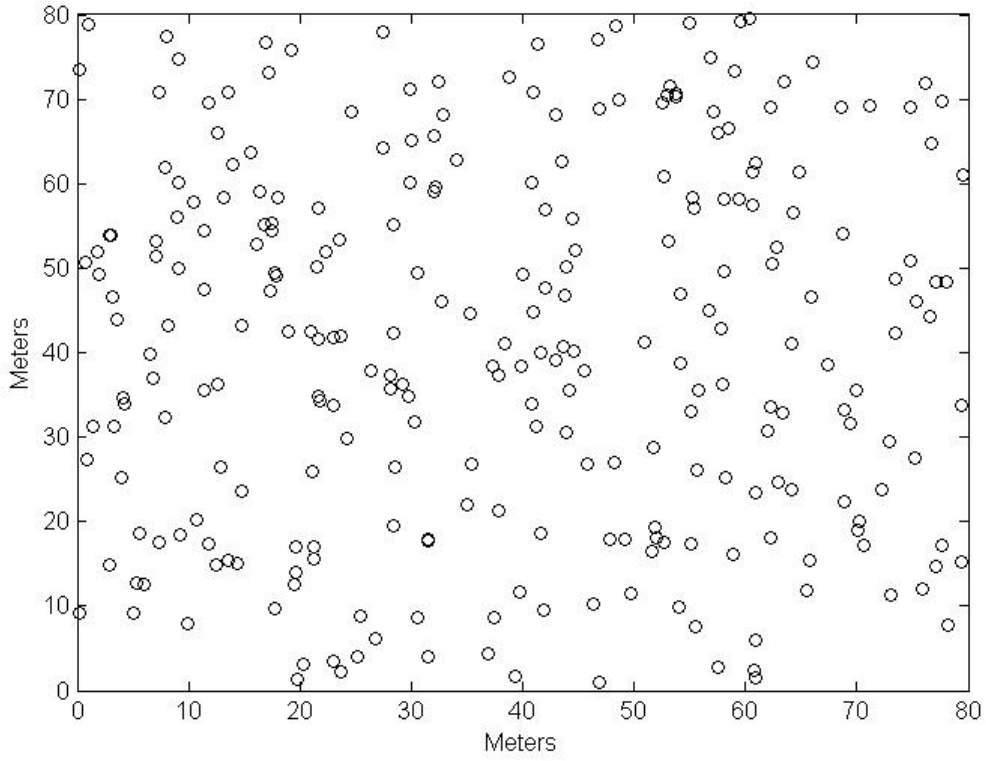

Figure 1. Simulation of a homogeneous Poisson process with λ = 420 trees ha^{-1}on a 80 x 80m square

2.15 Edge effect correction

When modeling spatial patterns of forest trees through stochastic point processes, the stochastic processes of interest are defined throughout *d*-dimensional Euclidean space, but the raw data used to model the process is observed only in some bounded window that is a subset of $\mathbb{R}^d$ such as a rectangular or circular research plot. Consequently, other trees that are just beyond the plot boundary but which are not observed may influence trees near the edge of the plot. Summary statistics such as

Ripley's *K*-function or the pair correlation function will be negatively biased because trees near the edge may be within a distance *r* of a tree that is not observed and these pairs will be omitted from the calculations of these statistics. This phenomena is known as edge bias (Radtke and Burkhart 1998). Edge effects must be considered both in summarizing the spatial structure of an observed point pattern and in fitting a stochastic point process to raw data. Frequently, edge effects and remedies for them are discussed in the context of summary statistics or indices that rely upon an observed spatial pattern such as Ripley's *K*-function (Ripley 1988; Haase 1995), the pair correlation function (Stoyan 2000), nearest-neighbor distance statistics (Gignoux et al. 1999; Pommerening and Stoyan 2006; Stoyan 2006), and competition indices (Monserud and Ek 1974; Martin et al. 1977). As pointed out by Monserud and Ek (1974), the easiest solution to edge effects is to simply ignore them. Another approach is to establish a buffer zone of some fixed width around the research plot(s) and use those trees in the buffer zone only to account for edge effects. Called the "border" method (Ripley 1988), this method has the advantage that for trees within the research plots all border trees are known, but obtaining this data can be expensive and any other information that these trees may provide is not being utilized to describe spatial pattern or to model it. Periodic and plot reflection edge effect corrections are techniques that create boundary trees by replicating the pattern of trees observed on the research plot. The periodic correction has the effect of replicating the plot about itself so that opposite edges of the plot adjoin. This method is also known as the toroidal edge correction (Ripley 1988) because for rectangular shaped plots, the joining of opposite plot edges creates a torus, a geometrical object that looks like a doughnut. Plot reflection, either at the plot boundary or at some distance from the plot

boundary replicates the plot at this boundary so that the trees outside the plot are the mirror image of those inside (Monserud and Ek 1974; Radtke and Burkhart 1998). Pommerening and Stoyan (2006) argued that periodic and plot reflection corrections were unrealistic in that these methods joined parts of the observed point pattern that do not occur. A limitation of these methods is that they work well only for rectangular plots; Williams et al. (2001) attempted to address this issue by proposing to transform circular plots into rectangular plots so that a periodic correction could be used. Ripley (1988), in the context of *K*-functions, described two edge-correction procedures based on making assumptions of isotropy or stationarity of the underlying point pattern. The isotropic edge-correction proceeds by weighing each pair of events of a point process by the reciprocal of the proportion of the circumference of the circle that falls within the plot, where the circle is centered at each event of the pair in turn and has a radius equal to the distance between the two events. The translational edge correction assumes stationarity of the pattern, and according to Ripley (1988), was originally proposed by Ohser and Stoyan (1981). Used with *K*-function analyses, this method weights pairs of events of the process based upon the proportion of the area of the plot that still remains after a translation of the plot. Some additional methods of plot correction include randomly locating trees with characteristics of those observed on the plot to positions outside the plot (Radtke and Burkhart 1998), and nearest-neighbor (NN) correction estimators (Pommerening and Stoyan 2006; Stoyan 2006; Illian et al. 2008, p. 187).

Comparative studies on the performance of edge correction methods (Radtke and Burkhart 1998; Pommerening and Stoyan 2006) have not yet identified any one method that is superior to others. Radtke and Burkhart (1998) reported that in estimating crown

canopy closure in loblolly pine plantations that periodic correction, plot reflection through a reflecting line passing through a row of planted trees, and randomly positioning trees outside the plot performed equally well, whereas ignoring edge effects resulted in biases over ten times larger. In contrast Pommerening and Stoyan (2006) compared the methods of no edge correction, periodic correction, plot reflection, the border method, and NN estimators to evaluate the bias and root mean square error (RMSE) in estimating six nearest-neighbor distance statistics using simulated point processes and two large (80m x 80m, 95m x 116m) plots in mixed oak-beech stands in Germany. Their results indicated that plot reflection resulted in the largest bias and RMSE for most indices compared across the different simulated stochastic process as well as for the "real" forest point patterns in the mixed oak stands. Surprisingly, no edge-correction often resulted in bias and RMSE quite favorably comparable to any of the edge-correction methods tested. Moreover, no method resulted in superior performance, despite the improved theoretical properties of the nearest neighbor (NN) edge-correction methods.

2.2 Gibbs point processes

2.21 Definition and formulation

Gibbs processes are a special class of point processes that arose in the field of statistical mechanics from the study of the thermodynamics of large particle systems (Ruelle 1969). Presumably, Gibbs processes should be useful in forestry applications because they allow for the modeling of the interaction between trees so that trees inhibit each other as they become closer in space. Since trees have physical dimension and compete for light, water, and other resources, this type of a regular process is somewhat suitable to simulate forest patterns.

The specific form of Gibbs point processes considered here are pairwise interaction point processes (Baddeley and Møller 1989; Diggle et al. 1994; Baddeley 2001) where the interactions are considered only to be between pairs of trees, and "triplets" processes (Geyer 1999) where the interactions are defined for triples and pairs of trees. As such, the Gibbs processes considered here are special cases of "interaction processes" (Tomppo 1986), which allow for interactions of higher order. The spatial interaction between trees are described by functions called "interaction potentials", owing to the origins of these point processes from statistical mechanics in physics, where the events of the process are molecules and the interaction potentials reflect the potential energy of the molecules that depend upon their position (Ruelle 1969; Cox and Isham 1980). When conditions are imposed upon the interaction potentials so that only 'neighboring' events of the process can interact, the point process is also a Markov point process (Baddeley and Møller 1989). In the context of forestry, these conditions imply

that trees would not interact with each other if they were separated from each other by some distance, *r*, which defines a 'neighborhood' around each tree.

To define a Gibbs process, consider point processes that are finite and simple, so that realized sequences **x** of the process represent an unordered collection of events $\{x_1, x_2, \ldots, x_n\}$ that exist in a completely separable metric space, such as *d*-dimensional Euclidean space. For the purposes here, only finite Gibbs processes defined on a bounded region (i.e. a forest) $F \subset \mathbb{R}^d$ are considered. Although Gibbs processes may be defined directly on $\mathbb{R}^d$, such infinite Gibbs processes are more challenging as they require conditioning on the behavior of the process outside of the observed data (Stoyan et al. 1995, p. 177-182). Now restrict consideration only to those point processes whose induced probability measure *P* is absolutely continuous with respect to a Poisson process with intensity measure $\Lambda(B)$ for all Borel sets $B \in \Omega$. Then by an application of the Radon-Nikodym theorem, there exists a density function, *f*, of a finite point process Φ such that

$$\mathbb{P}(\Phi \in A) = e^{-\Lambda(F)} \sum_{n=0}^{\infty} I_n(f, A) \qquad [39]$$

where $I_n(f, A)$ is defined by

$$I_0(f, A) = I_A(\varnothing) f(\varnothing)$$

when $n = 0$ and for $n = 1, 2, \ldots$

$$I_n(f, A) = \frac{1}{n!} \int_F \ldots \int_F I_A(\{x_1, x_2, \ldots, x_n\}) f(\{x_1, x_2, \ldots, x_n\}) \Lambda(dx_1) \ldots \Lambda(dx_n) \quad [40]$$

for every set $A \in \mathcal{N}$ (Baddeley 2001; Møller and Waagepetersen 2003, 2004; Baddeley 2007). The density *f* given above is referred to as a normalized density, meaning that if a specific mathematical form is desired for a density, say $g(\{x_1, x_2, \ldots, x_n\})$, *g* will only satisfy the conditions above if it is multiplied by a normalizing constant, *Z*, so that the probability density *f* will integrate to 1. Since it is desired to formulate a specific mathematical model for tree interactions, this implies that the resulting density will contain a normalizing constant *Z* given by

$$Z = \sum_{n=0}^{\infty} \frac{e^{-\Lambda(F)}}{n!} \int_F \ldots \int_F g(\{x_1, x_2, \ldots, x_n\}) \Lambda(dx_1) \ldots \Lambda(dx_n) \quad [41]$$

that depends upon an *n*-dimensional integral that is analytically intractable (Møller and Waagepetersen 2004, p. 82-83). Despite this difficulty, a Gibbs point process may be defined as a simple point process whose density *f* given above satisfies the so-called positivity condition that for any configuration $\mathbf{y} \subset \mathbf{x}$, $f(\mathbf{x}) > 0$ implies that $f(\mathbf{y}) > 0$ (Baddeley 2001).

To see the connection between Gibbs processes and interaction processes, Gregori and Mateu (2002) and Baddeley (2001) note that the density for a Gibbs process as defined above may be written as the following:

$$f(\mathbf{x}) = \exp\left\{-v_0 - \sum_{i=1}^{n(\mathbf{x})} v_1(x_i) - \sum_{i<j} v_2(x_i, x_j) - \sum_{i<j<k} v_3(x_i, x_j, x_k) - \ldots\right\} \qquad [42]$$

where v_0 is a constant and the v_k, k= 1, 2, 3, … are symmetric functions mapping from $\mathbb{G}_k$ to $\mathbb{R} \cup \{-\infty\}$, where $\mathbb{G}_k = \{N \in \mathbb{G} : N(B) = k\}$ These functions are called interaction potential functions and so Gibbs processes are sometimes referred to as "interaction processes" (Tomppo 1986). Now for simplicity consider only interaction between pairs of events, that is let $v_k = 0$ for $k \geq 3$. Then Eq. 42 may be simplified to

$$\begin{aligned} f(\mathbf{x}) &= \exp\left\{-v_0 - \sum_{i=1}^{n(\mathbf{x})} v_1(x_i) - \sum_{i<j} v_2(x_i, x_j)\right\} \\ &= \exp\{-v_0\} \prod_{i=1}^{n(\mathbf{x})} \exp\{-v_1(x_i)\} \prod_{i<j} \exp\{-v_2(x_i, x_j)\} \\ &= \alpha \prod_{i=1}^{n(\mathbf{x})} b(x_i) \prod_{i<j} h(x_i, x_j) \end{aligned} \qquad [43]$$

where the constant $\alpha = Z^{-1}$ is the normalizing constant, $n(\mathbf{x})$ is the random number of events in $\mathbf{x}$, the function $b(x_i)$, sometimes referred to as the "chemical activity" (Särkkä 1993) or the "basic intensity function" (Stoyan and Stoyan 1998), controls the spatial intensity of the process, and the function $h(x_i, x_j) = \exp(-v_2(x_i, x_j))$, known as the

interaction function, controls the spatial interaction between events of the process. For pairwise interaction processes, the interaction function may be expressed in terms of a pair potential function $\phi(\cdot)$, where

$$h(x_i, x_j) = \exp\left\{-\phi\left(\left\|x_i - x_j\right\|\right)\right\} \qquad [44]$$

The pair potential function $\phi(\cdot)$ defines the spatial interaction between events of the process for all pairs of events (x_i, x_j) separated by distance $r = \left\|x_i - x_j\right\|$. The pair potential is a real-valued function with the interpretation that positive values of $\phi(\cdot)$ indicate repulsion of events, whereas negative values indicate attraction. An example of a pair potential function, proposed by Fiksel (1984), is shown in Figure 2. This pair potential function is defined to equal infinity (∞) when $r \leq r_0$. Consequently when $r \leq r_0$, $\exp(-\phi) = \exp(-\infty) = 0$, $f(\mathbf{x})$ must also equal zero, implying that a given pair of trees can never exist closer than r_0 units apart. This is known as a "hard-core" distance, and many pair potential functions have been constructed to incorporate absolute repulsion of events. Some examples of hard-core pair potentials are the hard-core Strauss potential (Strauss 1975) and the multi-scale pair potential (Penttinen 1984). Pair potentials that do not impose a hard-core distance include the "Diggle" potential (Diggle et al. 1994), the "very soft-core" potential (Ogata and Tanemura 1984), and the Lennard-Jones potential (Ruelle 1969). The pair potential function given by Fiksel (Fig. 2) is also defined to equal zero when $r > R$. The constant R is a finite range distance; it may be interpreted as distance

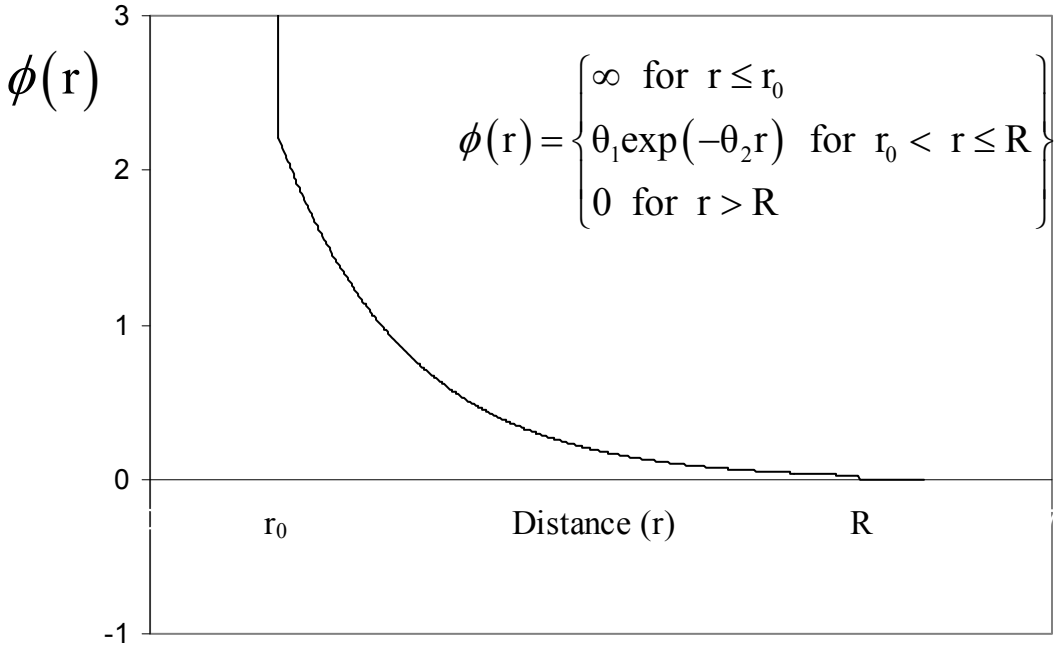

Figure 2. The pair potential function of Fiksel (1984) with $\theta_1 = 6$ and $\theta_2 = 1$.

beyond which trees no longer interact with each other. When a pairwise interaction process has a finite range of interaction, the process is said to be a Markov point process (Baddeley 2001).

While pair potentials are constructed to incorporate attraction and repulsion, they are subject to constraints that ensure that the resulting pairwise interaction process is well defined and in fact integrates to one. These constraints are referred to as stability conditions, and Møller and Waagepetersen (2003) define two types: local stability and Ruelle stability. Local stability is said to be satisfied when the conditional intensity function is uniformly bounded and the density of the process satisfies the positivity condition. Ruelle stability (Ruelle 1969) is satisfied if the density of the process is

dominated by a homogeneous Poisson process. As noted by Gregori and Mateu (2002), whenever the interaction function is purely repulsive, that is $h(x_i, x_j) \geq 0$, these stability requirements are met. When unstable potentials are used, it is possible that the point process density is not well-defined and patterns simulated from such a model may be more clustered than intended (Gates and Westcott 1986).

An important construct in the study of Gibbs processes is the conditional intensity function. As discussed in §2.13, the conditional intensity function can be interpreted as giving the infinitesimal conditional probability that a point process has a point at u given that the point process coincides with the realized pattern ($\mathbf{x}$) everywhere else. When the distribution of a Gibbs process is constructed to be absolutely continuous with respect to a Poisson process with unit intensity, that is a Poisson process with $\Lambda(B) = 1\ \mu_L(B)$, then Baddeley (2007, p. 65) showed that the conditional intensity function may be expressed in terms of the probability density of the Gibbs process, f so that

$$\lambda^*(u;\mathbf{x}) = \frac{f(\mathbf{x} \cup \{u\})}{f(\mathbf{x})} \qquad [45]$$

where $u \notin \mathbf{x}$ and $\mathbf{x}$ is any finite point configuration of the process. Applying this to pairwise interaction processes gives

$$\lambda^{*}(u;\mathbf{x}) = \frac{f(\mathbf{x}\cup\{u\})}{f(\mathbf{x})}$$

$$= \frac{\alpha\, b(u)\prod_{i=1}^{n(\mathbf{x})} b(x_i)\prod_{\substack{i=1\\ x_i\neq u}}^{n(\mathbf{x})} h(u,x_i)\prod_{i<j} h(x_i,x_j)}{\alpha\prod_{i=1}^{n(\mathbf{x})} b(x_i)\prod_{i<j} h(x_i,x_j)} \qquad [46]$$

$$= b(u)\prod_{\substack{i=1\\ x_i\neq u}}^{n(\mathbf{x})} h(u,x_i)$$

where the intractable normalizing constant α cancels out of the ratio in Eq.46 (Baddeley 2001). Consequently, parameter estimation and statistical inference for the pairwise interaction point process may be more amenable through maximizing a "pseudo-likelihood" function defined through the conditional intensity function rather than directly maximizing the true likelihood (the joint density), which requires direct evaluation of α which does not exist in closed form.

2.3 Statistical Inference for Gibbs point processes

2.31 Classical approaches

The estimation of parameters for Gibbs processes and statistical inference for them is complicated by the fact that the probability density for the process contains the normalizing constant (written as the function of the parameters of the process)

$$Z(\theta) = \sum_{n=0}^{\infty} \frac{e^{-\Lambda(F)}}{n!} \int_F \ldots \int_F g_\theta\left(\{x_1, x_2, \ldots, x_n\}\right) \Lambda(dx_1) \ldots \Lambda(dx_n)$$

In the case of a finite Gibbs process, where the number of points $n(\mathbf{x}) = n$ is fixed, the normalizing constant can be simplified to

$$Z(\theta) = \frac{e^{-\Lambda(F)}}{n!} \int_F \ldots \int_F g_\theta\left(\{x_1, x_2, \ldots, x_n\}\right) \Lambda(dx_1) \ldots \Lambda(dx_n) \qquad [47]$$

The choice of whether to use a stationary or a finite Gibbs process is not obvious (Illian et al. 2008, p. 160). Some methods of classical inference for Gibbs processes, such as the approximate and Monte Carlo maximum likelihood methods require fixed n. In contrast, the methods of Takacs-Fiksel estimation and maximum pseudo-likelihood estimation are applicable to both stationary and finite Gibbs processes where $n(\mathbf{x})$ is a random quantity that needs to be estimated in addition to the parameters of the interaction potential

functions. For the purposes of this discussion, finite Gibbs processes defined on the bounded region $F \subset \mathbb{R}^d$ are assumed.

A further simplifying assumption that is used to deal with the normalizing constant is to construct the Gibbs process density with respect to a Poisson process with mean of 1 point per unit area of F. This implies that the intensity measure Λ given in Eq. 47 satisfies

$$\Lambda(dx) = \lambda\ \mu_L(dx) = \mu_L(dx) \qquad [48]$$

so that the normalizing constant may be evaluated as a Riemann integral.

Despite these simplifying assumptions, the normalizing constant of a Gibbs process remains an analytically intractable function of the parameters of the process due to the multiplicity of the integral given in Eq. 47. Hence parameter estimation through exact maximum likelihood techniques or methods of moments is impossible. To overcome this, four general approaches have been proposed in the literature using classical statistical techniques. The first approach encompasses a variety of techniques to approximate the normalizing constant in order to allow maximum likelihood estimation of the Gibbs process parameters. These include Poisson approximations (Saunders and Funk 1977; Penttinen 1984; Ripley 1988, p. 56-59), and virial or Mayer cluster expansions (Ogata and Tanemura 1981, 1984; Ripley 1988, p.59-62; Diggle et al. 1994). These approximate maximum likelihood methods rely upon so called "sparse-data" assumptions (Saunders and Funk 1977), which assume that the area of the observation window F is proportional to the square of the number of points of the Gibbs process, e.g.

n = 10 trees in 100m^2 fixed area plot F. Saunders and Funk (1977) described a Poisson approximation for the Strauss process (Strauss 1975) given by the pair potential

$$\phi(r) = \begin{Bmatrix} \theta & \text{for } r \le R \\ 0 & \text{for } r > R \end{Bmatrix} \qquad [49]$$

where the interaction distance R is assumed to be small relative to the size of F and $\theta > 0$ is a scalar parameter. Saunders and Funk proved that under the sparse data assumptions, the sufficient statistic of the Strauss process given by

$$Y_n(r) = \sum_{i<j} I_{[0,R]}(r) \qquad [50]$$

approximately follows a Poisson distribution with mean $(0.5)n(n\text{-}1)\pi r^2$, when F is scaled to the unit square. This allows for an approximation of the normalizing constant of the Strauss process since

$$Z(\theta) \approx \mathbb{E}\left[\exp\left(\theta \sum_{i<j} I_{[0,R]}(r) \right) \right] \qquad [51]$$

becomes the moment generating function of $Y_n(r)$. Penttinen (1984) extended this approach to any pairwise interaction point processes with a finite interaction distance so that

$$Z(\theta) \approx \exp\left[0.5n(n-1)\left(\frac{\pi R^2}{\mu_L(F)} \right)\left(\left(\frac{2}{R^2} \right)\int_0^R r\exp\left(-\phi(r)\right)dr - 1 \right)\right] \quad [52]$$

Penttinen (1984) noted that one implication of the sparse data assumption is that short interpoint distances are independent, a feature true in the limit for purely repulsive pair potentials.

An alternate sparse data approximation for the normalizing constant of pairwise interaction processes is the Mayer cluster expansion proposed by Ogata and Tanemura (1981, 1984), which gives that:

$$Z(\theta) \approx \mu_L(F)^n \left(1 - \frac{a(\theta)}{\mu_L(F)} \right)^{n(n-1)/2} \quad [53]$$

where

$$a(\theta) = \int_0^\infty \left(1 - \exp\left(-\phi(r)\right)\right) b(r)\, dr$$

and $b(r) = 2\pi r$ for pairwise interaction point processes defined on $F \subseteq \mathbb{R}^2$. Ilian et al. (2008, p. 163) remark that both the Mayer cluster expansion and Poisson approximations of the normalizing constant when used with maximum likelihood estimation can give very poor parameter estimates if the data are not sparse or the interaction distance is large

relative to region F. Gates and Westcott (1986) demonstrate that the Mayer cluster expansion can be a very poor approximation to the normalizing constant whenever tight clusters of 3 or more points can frequently occur in the given point configuration.

The second approach is to employ Monte Carlo methods to approximate the normalizing constant within maximum likelihood estimation. A number of variants have been developed including 1) the use of Monte Carlo samples to estimate polynomial functions that approximate virial expansion approximations of the normalizing constant (Ogata and Tanemura 1984), 2) the use of Monte Carlo samples from the Gibbs density using a fixed parameter vector ψ to calculate the gradient and the Hessian of the log-likelihood function combined with the Newton-Raphson algorithm to iteratively update ψ and resample from the Gibbs density until the obtained maximum likelihood estimates of θ converge to a common value (Penttinen 1984), 3) the use of stochastic approximation methods (Moyeed and Baddeley 1991; Gu and Zhu 2001), and 4) the Monte Carlo Likelihood (MCL) of Geyer and Thompson (1992). Of these, the MCL method of Geyer and Thompson (1992) has received the most discussion in the literature (Geyer 1996, 1999; Van Lieshout 2000; Møller and Waagepetersen 2003, 2004; Illian et al. 2008). The MCL approach is an importance sampling procedure that attempts to maximize the log likelihood ratio of a spatial point process using a fixed parameter vector. Let ψ be this fixed parameter vector so that the log-likelihood ratio of is given by

$$\begin{aligned} l(\theta \mid \mathbf{x}, \psi) &= \log\left(\frac{f(\mathbf{x} \mid \theta)}{f(\mathbf{x} \mid \psi)}\right) \\ &= \log\left(f(\mathbf{x} \mid \theta)\right) - \log\left(f(\mathbf{x} \mid \psi)\right) \end{aligned} \qquad [54]$$

Now define $h_\theta(\mathbf{x}) = \frac{e^{-\Lambda(F)}}{n!} g_\theta(\mathbf{x})$ so that the normalizing constant of the Gibbs density (Eq. 72) may be written as

$$\begin{aligned} Z(\theta) &= \frac{e^{-\Lambda(F)}}{n!} \int_{F^n} g_\theta(\mathbf{x}) \Lambda(d\mathbf{x}) \\ &= \int_{F^n} \frac{e^{-\Lambda(F)}}{n!} g_\theta(\mathbf{x}) \Lambda(d\mathbf{x}) \\ &= \int_{F^n} h_\theta(\mathbf{x}) \Lambda(d\mathbf{x}) \end{aligned}$$

Consequently the log-likelihood ratio in Eq. 54 can be expressed as

$$\begin{aligned} l(\theta \mid \mathbf{x}, \psi) &= \log\left(f(\mathbf{x} \mid \theta)\right) - \log\left(f(\mathbf{x} \mid \psi)\right) \\ &= \log\left(\frac{h_\theta(\mathbf{x})}{Z(\theta)}\right) - \log\left(\frac{h_\psi(\mathbf{x})}{Z(\psi)}\right) \\ &= \log\left(h_\theta(\mathbf{x})\right) - \log\left(Z(\theta)\right) - \log\left(h_\psi(\mathbf{x})\right) + \log\left(Z(\psi)\right) \\ &= \log\left(\frac{h_\theta(\mathbf{x})}{h_\psi(\mathbf{x})}\right) - \log\left(\frac{Z(\theta)}{Z(\psi)}\right) \\ &= \log\left(\frac{e^{-\Lambda(F)}/n!}{e^{-\Lambda(F)}/n!} \frac{g_\theta(\mathbf{x})}{g_\psi(\mathbf{x})}\right) - \log\left(\frac{Z(\theta)}{Z(\psi)}\right) \\ &= \log\left(\frac{g_\theta(\mathbf{x})}{g_\psi(\mathbf{x})}\right) - \log\left(\frac{Z(\theta)}{Z(\psi)}\right) \end{aligned} \qquad [55]$$

As noted by Geyer (1999), the ratio of normalizing constants in Eq. 55 can be expressed as

$$
\begin{aligned}
\frac{Z(\theta)}{Z(\psi)} &= \frac{1}{Z(\psi)} \int_{F^n} h_\theta(\mathbf{x}) \Lambda(d\mathbf{x}) \\
&= \frac{1}{Z(\psi)} \int_{F^n} h_\theta(\mathbf{x}) \left(\frac{h_\psi(\mathbf{x})}{h_\psi(\mathbf{x})} \right) \Lambda(d\mathbf{x}) \\
&= \int_{F^n} \frac{h_\theta(\mathbf{x})}{h_\psi(\mathbf{x})} \left(\frac{h_\psi(\mathbf{x})}{Z(\psi)} \right) \Lambda(d\mathbf{x}) \\
&= \int_{F^n} \frac{e^{-\Lambda(F)}/n!}{e^{-\Lambda(F)}/n!} \frac{g_\theta(\mathbf{x})}{g_\psi(\mathbf{x})} f(\mathbf{x} \mid \psi) \Lambda(d\mathbf{x}) \\
&= \int_{F^n} \frac{g_\theta(\mathbf{x})}{g_\psi(\mathbf{x})} P(d\mathbf{x}) \\
&= \mathbb{E}_{f(\mathbf{x}|\psi)} \left[\frac{g_\theta(\mathbf{x})}{g_\psi(\mathbf{x})} \right]
\end{aligned}
\qquad [56]
$$

where P is the probability distribution of the Gibbs point process. The expectation in Eq. 56 can be approximated by importance sampling with importance weights

$$
w(\mathbf{x}) = \frac{g_\theta(\mathbf{x}) / g_\psi(\mathbf{x})}{\sum_{i=1}^{N} g_\theta(\mathbf{X}_i) / g_\psi(\mathbf{X}_i)} \qquad [57]
$$

where $\mathbf{X}_1$, $\mathbf{X}_2$, …, $\mathbf{X}_N$ form a sample of size N from the Gibbs density with parameters ψ, so that

$$
\mathbb{E}_\psi \left[\frac{g_\theta(\mathbf{x})}{g_\psi(\mathbf{x})} \right] \approx \frac{1}{N} \sum_{i=1}^{N} g_\theta(\mathbf{X}_i) / g_\psi(\mathbf{X}_i) \qquad [58]
$$

and the log-likelihood ratio becomes

$$l(\theta \mid \mathbf{x}, \psi) \approx \log\left(\frac{g_\theta(\mathbf{x})}{g_\psi(\mathbf{x})}\right) - \log\left(\frac{1}{N}\sum_{i=1}^{N} g_\theta(\mathbf{X}_i)/g_\psi(\mathbf{X}_i)\right) \qquad [59]$$

To find the maximum likelihood estimator of the Gibbs density, Eq. 59 is maximized with respect to θ by using numerical methods such as the Newton-Raphson algorithm. While similar to the method of Penttinen (1984), Geyer and Thompson's method requires only one sample from the target density under the parameter vector ψ. Penttinen's method initially draws samples using ψ, but then updates ψ to ψ' by the Newton-Raphson algorithm. Samples from the target density are then re-drawn using ψ' and the gradient and Hessian of the log-likelihood is re-evaluated. This process is iteratively repeated until the obtained maximum likelihood estimates converge to a common value. Geyer and Thompson argue that it is difficult to prove what Penttinen's method converge to; Møller and Waagepetersen (2004) comment that the estimates obtained by Penttinen's method will not converge but will oscillate around the maximizing value due to Monte Carlo error.

A third classical approach to statistical inference for Gibbs processes is known as the Takacs-Fiksel method, after the works of R. Takacs (1983, 1986) and T. Fiksel (1984, 1988). Under the assumption of a stationary and isotropic Gibbs process, the Takacs – Fiksel method relies upon the Nguyen-Zessin formula (Nguyen and Zessin 1979) given in Eq. 29 and the relationship

$$\lambda \, \mathbb{E}_u^! \left[f\left(\Phi - \delta_u\right) \right] = \mathbb{E}\left[\lambda^*\left(u, \Phi\right) f\left(\Phi\right) \right] \qquad [60]$$

which holds for any non-negative measurable function f, where it is recalled from §2.1.3 that the λ^* is the conditional intensity of the point process Φ, λ is the (unconditional) intensity of the process Φ, the notation $\Phi - \delta_u$ refers to the point process Φ excluding the point located at u. Since the expectation on the left-hand side of Eq. 60 is with respect to the reduced Palm distribution of the process (Eq. 26) and the expectation on the right-hand side is with respect to the induced probability distribution of the point process, Eq. 60 can be equivalently expressed (Cressie 1993, p. 686) as

$$\lambda \int f\left(\Phi\right) P_u^!\left(d\Phi\right) = \int f\left(\Phi\right) \lambda^*\left(u, \Phi\right) P\left(d\Phi\right) \qquad [61]$$

The Takacs-Fiksel procedure seeks to estimate the parameters of the process by choosing a system of m test functions f_i for $i = 1,2, \ldots, m$ with m less than or equal to the dimension of the parameter vector θ and constructing estimators of the left- and right-hand sides of Eq. 61. Fiksel (1988) proposed for a pairwise interaction process to set the point u to be the origin (0) and the test function f_i to be

$$f_i\left(\Phi\right) = N\left(b\left(0, r_i\right)\right) \exp\left(-\log\left(\beta\right) + \sum_{\substack{i=1 \\ x_i \neq 0}}^{n(\mathbf{x})} \phi\left(\left\| x_i \right\|\right) \right) \qquad [62]$$

where $N(b(0, r_i))$ is the number points in the ball centered at the origin with radius r_i and $r_1 < r_2 < \ldots < r_m$. The estimated parameters of the Gibbs process are those that minimize the sum of squared differences between the estimators of the left- and right-hand sides of Eq. 61 (Fiksel 1984). Alternate test functions and their merits are discussed in (Illian et al. 2008, pp. 402-407); applications of the Takacs-Fiksel method to unmarked Gibbs processes for forestry data include the dissertations by Tomppo (1986) and Li (2004).

A fourth classical approach to statistical inference for Gibbs processes is known as maximum pseudolikelihood. The concept of maximum pseudolikelihood was first proposed by Besag (1975, 1977) who considered a finite collection of spatially referenced random variables, X_i, for $i = 1, 2, \ldots n$. Besag (1975) defined the pseudolikelihood function as the product of conditional probabilities of observing each random variable $X_i = x_i$, given the observation of the other random variables $X_j = x_j$ for all $i \neq j$. Besag (1977) applied this concept to the auto-Poisson distribution of counts of objects within cells of a spatial lattice. When the lattice cell area is taken successively smaller to the limit zero, Besag argued that resulting sequence of stochastic processes would converge almost surely to the Strauss process (Strauss 1975). This result was proven by Besag et al. (1982), who also demonstrated convergence for pairwise interaction processes through a different specification of the mean of the auto-Poisson distribution. Ripley (1988) presented the pseudolikelihood for a stochastic point process in terms of the Papangelou conditional intensity $\lambda^*(u; \mathbf{x})$ so that the pseudolikelihood of any stochastic point process, $PL(\theta, \mathbf{x})$, observed in a bounded region F of d-dimensional Euclidean space is given as

$$PL(\theta, \mathbf{x}) = \left(\prod_{i=1}^{n(\mathbf{x})} \lambda_\theta^*(x_i; \mathbf{x}) \right) \exp\left(-\int_F \lambda_\theta^*(u; \mathbf{x})\, du \right) \quad [63]$$

where θ is a vector of parameters defining the probability density of the point process. The parameter vector θ may then be solved for by maximizing $PL(\theta, \mathbf{x})$ with respect to θ, as the conditional intensity function depends upon θ[1]. One clever approach to solving the maximum pseudolikelihood equations is developed in the R-language package Spatstat (Baddeley and Turner 2005). In Spatstat, Baddeley and Turner (2000) adapted an approximation to the pseudolikelihood proposed by Berman and Turner (1992) that approximates the integral defined over the region F in Eq. 63 by numerical quadrature. In the case of a pairwise interaction process, a list of m integration points, $u_j, j = 1, 2, \ldots, m$ located within F is created so that

$$\int_F \lambda_\theta^*(u; \mathbf{x})\, du \approx \sum_{j=1}^{m} \lambda_\theta^*(u_j; \mathbf{x})\, w_j \quad [64]$$

where w_j are quadrature weights that sum up to the area of F. Then the pseudolikelihood can be approximated as

$$PL(\theta, \mathbf{x}) \approx \left(\prod_{i=1}^{n(\mathbf{x})} \lambda_\theta^*(x_i; \mathbf{x}) \right) \exp\left(-\sum_{j=1}^{m} \lambda_\theta^*(u_j; \mathbf{x})\, w_j \right) \quad [65]$$

[1] The subscript θ is added to the conditional intensity, chemical activity, and interaction functions to emphasize the dependency of these functions on θ.

Taking the natural logarithm of both sides of Eq. 65 and defining the list of integration points u_j to include the data points x_i, $i = 1, 2, \ldots, n$ gives

$$\begin{aligned} \log PL(\theta, \mathbf{x}) &\approx \left(\sum_{i=1}^{n(\mathbf{x})} \log\left(\lambda_\theta^*(x_i; \mathbf{x})\right) \right) - \sum_{j=1}^{m} \lambda_\theta^*(u_j; \mathbf{x})\, w_j \\ &= \sum_{j=1}^{m} \left[y_j \log\left(\lambda_\theta^*(u_j)\right) - \lambda_\theta^*(u_j) \right] w_j \end{aligned} \qquad [66]$$

as shown by Baddeley and Turner (2000, p. 288), where $y_j = z_j / w_j$, with w_j being the quadrature weights and z_j is an indicator variable that is one when the integration point is a data point and zero when it is a "dummy" point, e.g.

$$z_j = \begin{Bmatrix} 1, \text{ if } u_j \in \mathbf{x} \\ 0, \text{ if } u_j \notin \mathbf{x} \end{Bmatrix} \qquad [67]$$

As pointed out by Baddeley and Turner (2000), this formulation is equivalent to a weighted Poisson regression, treating the y_j as independent Poisson random variables with means $\mu_j = \lambda^*_\theta(u_j)$ and weights w_j. Consequently standard software for fitting generalized linear models can be used to maximize the pseudolikelihood and estimate θ for any Gibbs point process provided that the conditional intensity function is log-linear with respect to θ, that is

$$\log\left(\lambda_\theta^*(u; \mathbf{x})\right) = \theta^{\mathrm{T}} S(u, \mathbf{x}) \qquad [68]$$

where $S(u, \mathbf{x})$ is a vector of spatial covariates which depends upon the formulation of the model of the point process in question. In Spatstat, Eq. 66 is maximized through a call to the function glm ( ) specifying a log-link function. The weights w_j are obtained through the Dirichlet tessellation[2] of F using the list of integration points that includes the data and a number of "dummy" points. While in principle, any d-dimensional numerical quadrature scheme may be used to approximate the pseudolikelihood, Berman and Turner (1992) remark that maximization of the pseudolikelihood is less computationally intensive using a Dirichlet tessellation than using quadrature schemes based on product rules which divide F into a grid of small rectangles with vertices at the data and dummy points.

Few comparison studies for Gibbs processes using classical parameter estimation methods are available (Diggle et al. 1994; Mateu and Montes 2001). Diggle et al. compared 1) Ogata and Tanemura's (1984) approximate maximum likelihood estimation using virial expansions, 2) maximum pseudolikelihood estimation, and 3) the Takacs-Fiksel method. In turn, Mateu and Montes (2001) compared 1) Ogata and Tanemura's (1981) approximate maximum likelihood using Mayer cluster expansions, 2) a sparse data approximation to the normalizing constant as in Eq. 52, 3) Ogata and Tanemura's (1984) approximate maximum likelihood estimation using virial expansions, 4) Geyer and Thompson's (1992) Monte Carlo maximum likelihood method (MCL), 5) Penttinen's (1984) Monte Carlo Newton–Raphson (MCNR) maximum likelihood method and 6) maximum likelihood using stochastic approximation (Moyeed and Baddeley 1991). Both Diggle et al. and Mateu and Montes considered the parameter estimation

[2] The Dirichlet tessellation is also known as a Voronoi diagram, where the region F is divided into mutually disjoint convex polygons centered on the integration points (Okabe et al. 1992)

procedures with the following edge-correction procedures: 1) the periodic (toroidal) edge-effect corrections, 2) Ripley's (1977) isotropic edge-effect correction, and 3) the free boundary condition, i.e. no edge-effect correction. While Mateu and Montes' (2001) study was applied to the Strauss process (Strauss 1975), Diggle et al. looked at two other pair potential models, one named the "Diggle" potential from Diggle (1986) which has the form

$$\phi(r) = \begin{cases} -\ln\left(1-\left(1-r^2/\theta^2\right)^2\right) & \text{for } r \le \theta \\ 0 \text{ for } r > \theta \end{cases} \qquad [69]$$

and the other called the "very soft-core potential" from (Ogata and Tanemura 1984), which has the form

$$\phi(r) = -\ln\left(1-\exp\left(-r^2/\theta^2\right)\right) \qquad [70]$$

The results of the simulation studies by Diggle et al. (1994) and Mateu and Montes (2001) provide some insight into the various estimation methods for Gibbs point processes. Approximate maximum likelihood methods that rely upon sparse data assumptions, such as cluster expansions and virial expansions, showed substantial bias in both studies when strong inhibitory spatial interaction as present. Mateu and Montes (2001) found that when the number of points was conditioned upon, MCNR and maximum likelihood provided unbiased parameter estimates, provided that the interaction distance was no more than 10% of the length of the side of the square bounding window.

When the number of points was not conditioned upon, MCL and MCNR provided better parameter estimates in terms of bias and variance than estimates obtained from maximum likelihood using stochastic approximations. Maximum pseudolikelihood and the Takacs-Fiksel were not tried by Mateu and Montes (2001), however Diggle et al. (1994) found that the performance of these methods depended upon the strength of the inhibitory point to point interaction. When this interaction was weak, maximum pseudolikelihood performed well in terms of bias and variance; when strong, maximum pseudolikelihood estimates showed "significant bias" (p.116). The Takacs-Fiksel method did not perform as well as maximum pseudolikelihood or approximate maximum likelihood using virial expansions when the interaction was weak, though it did give unbiased parameter estimates under the periodic edge-effect correction. Both the simulation studies of Diggle et al. (1994) and Mateu and Montes (2001) emphasized that the quality of the parameter estimates depends significantly upon the edge-effect correction used when the inhibitory interaction is strong. The periodic edge-effect correction provided the least bias, whereas Ripley's isotropic edge effect correction showed negative bias. When the spatial interaction is weak, Diggle et al. (1994) found little differences in biases between edge-effect corrections for a particular estimation method. Overall, any edge-effect correction provided better results (less bias) than using the free boundary condition, regardless of the method used.

In addition to providing point estimates of the parameters of a spatial point process, it is also desirable to have knowledge of the variability of that point estimate through interval estimation. For point processes, interval estimation for classical parameter estimates is very limited due to the lack of available asymptotic theory. In

particular such asymptotic theory would need to describe the behavior of the parameter estimators as the observed sampling window increases to infinite size (Møller and Waagepetersen 2004, p. 150). For exponential family pairwise interacting Gibbs point processes, Jensen (1993) demonstrated asymptotic normality for exact maximum likelihood estimators. Restrictions are placed on the pair potential models so that they are of finite range and are Ruelle – stable (Ruelle 1969), which is attained by purely repulsive potentials and potentials with a hard-core parameter. While of theoretical interest, it is not clear how Jensen's results can be put to practical use. Both the vector of first derivatives and the Hessian matrix of second derivatives of the log-likelihood of the Gibbs distribution need to be evaluated to invoke asymptotic normality of the maximum likelihood estimators, but this implies differentiating the normalizing constant which itself is intractable. Furthermore, Geyer and Møller (1994) and Møller and Waagepetersen (2007) remark that the conditions needed to obtain asymptotic normality of MLE by Jensen demand weak spatial interaction that is nearly Poisson. Under similar restrictions of Ruelle – stability and finite range pair potentials, maximum pseudolikelihood estimates of the parameters of exponential family pairwise interaction processes are shown to be asymptotically normally distributed (Jensen and Künsch 1994).

Outside of these restrictive circumstances, asymptotic theory for classical estimators of Gibbs point process parameters does not yet exist. The primary alternative to relying on asymptotics for classical inference on Gibbs point processes is to use the parametric bootstrap (Møller and Waagepetersen 2007; Illian et al. 2008). The parametric bootstrap works by simulating the Gibbs point process repeatedly using estimated values of the parameters of the process, which are obtained from maximum likelihood,

maximum pseudolikelihood, or other means. These simulated point patterns are then treated as observed data, from which new estimates of the parameter values are obtained. The variance of the estimated parameters can then be taken to estimate the variance of the estimator itself (Illian et al. 2008, p. 455). As pointed out by Geyer and Møller (1994), there is no theory to support the parametric bootstrap in absence of a central limit theorem for the MLE of a Gibbs point process; Bognar (2005) questions in general the soundness of using the parametric bootstrap for interval estimation for Gibbs point processes.

2.31 Bayesian approaches

In the classical approaches to statistical inference, the data are assumed to be a random sample from a probability distribution that is specified by unknown, fixed parameter values. In the Bayesian setting, the sampling distribution of the data is viewed as a conditional distribution given parameter values that are also considered to be random values. The parameters are believed to follow a probability distribution known as the prior distribution, which is specified by the researcher in advance. Inference about the parameters is obtained from the posterior distribution, which can be viewed as the prior distribution updated by the knowledge of the observed data (Casella and Berger 2002, p. 324). The posterior distribution is obtained through the application of Bayes' Theorem

$$f(\theta \mid \mathbf{x}) = \frac{f(\mathbf{x} \mid \theta) f(\theta)}{\int f(\mathbf{x} \mid \theta) f(\theta) d\theta} \quad [71]$$

where $f(\mathbf{x} \mid \theta)$ is the likelihood and $f(\theta)$ is the prior distribution of the parameter vector θ. In addition to providing point estimates of θ through the posterior mean and mode, Bayesian inference can provide an interval estimate for θ with a direct probability interpretation; that is, real numbers a and b such that there is an α% probability that the true value of θ lies between a and b. Formally, this Bayesian credible interval can be expressed as

$$\mathbb{P}(a \leq \theta \leq b) = \int_{[a,b]} f(\theta \mid \mathbf{x}) d\theta \quad [72]$$

Despite the advantage of being able to simply construct interval estimates for the parameters of Gibbs point processes, few authors have attempted to do so (Berthelsen and Møller 2006). The first attempt at Bayesian inference by Heikkinen and Penttinen (1999) considered "the full posterior analysis to be too demanding with the currently available machinery". Instead analysis was restricted to obtaining the posterior mode for a pairwise interaction process specified by the multiscale potential (Penttinen 1984). To obtain the posterior mode, Heikkinen and Penttinen did not attempt to sample from the posterior distribution but rather used a Monte Carlo Marquardt algorithm, constructed analogously to the Monte Carlo Newton Raphson algorithm of Penttinen (1984), to iteratively find the posterior mode given an initial estimate of the MLE of the parameters of the Gibbs process. More recent efforts at Bayesian inference of Gibbs point processes include the

works of Bognar (2004, 2005, 2008) and Berthelsen and Møller (2003, 2006, 2008) which have utilized Markov Chain Monte Carlo (MCMC) methodology to draw samples from the posterior distribution given a specified Gibbs likelihood. To draw the samples from the posterior in the context of Metropolis-Hastings MCMC methods, it becomes necessary to evaluate ratios of the likelihood (e.g. Eq. 54) which requires that a *ratio* of normalizing constants (e.g. Eq. 55) must be evaluated. Perhaps the most important difference in the methodology of Bognar and Berthelsen and Møller is how the authors choose to evaluate the normalizing constant ratio. Bognar used importance sampling methodology whereas Berthelsen and Møller used path sampling combined with perfect simulation (2003) and also an auxiliary variable technique (2006, 2008) to address this technical challenge. Since an understanding of MCMC methodology is necessary to appreciate the ideas of Bognar and Berthelsen and Møller, discussion of these papers is deferred to Sec. 3.7, where detailed discussion of Bognar's importance sampling method and Berthelsen and Møller's auxiliary variable method are found in Sec. 3.72 and 3.73, respectively.

2.4 Markov Chain Monte Carlo

2.41 General principles

Given a random variable X defined on a probability space $(\Omega, \mathcal{F}, \mu)$ with probability distribution $\mu(A) = \mu\left(X^{-1}(A)\right)$ for any $A \in \mathcal{F}$, it is frequently desired to evaluate the expected value of a real function of X, say $g(X)$, defined as (Billingsley 1995, p.274):

$$\mathbb{E}\left(g(X)\right) = \int_{-\infty}^{\infty} g(x)\,\mu(dx) \qquad [73]$$

If the random variable X is continuous with probability density function $f(X)$, then the more familiar Riemann integration notation may be used (Casella and Berger 2002, p. 55):

$$\mathbb{E}\left(g(X)\right) = \int_{-\infty}^{\infty} g(x)\,f(x)\,dx \qquad [74]$$

Because it is quite possible that the integral given by Eq. 74 does not exist in closed form, a natural question arises as to how this integral can be evaluated. Monte Carlo integration

is a technique that approximates the expectation of $g(X)$ by drawing m samples X_t from $f(X)$ so that (Gilks et al. 1996, p.4)

$$\mathbb{E}\left(g\left(X\right)\right) \approx \frac{1}{m}\sum_{t=1}^{m} g\left(X_t\right) \qquad [75]$$

Ideally, the m samples would be independently and identically distributed (i.i.d.) from $f(X)$ so that the large of law numbers would apply to make the error in this approximation as small as desired. Because $f(X)$ may be complex, independent samples from $f(X)$ are often not feasible. Instead dependent samples may be used provided that the samples occur throughout the support of $f(X)$ (Gilks et al. 1996, p. 4).

Markov Chain Monte Carlo (MCMC) describes a collection of techniques to obtain dependent samples from $f(X)$ by setting up a Markov chain whose probability distribution will converge to $f(X)$ as the chain is allowed to run infinitely long. Formally, a Markov chain is a discrete-time stochastic process $X_0, X_1, X_2, \ldots$ that has the property that the state of the chain at time t, X_t, only depends upon the previous state of the chain, X_{t-1}. Mathematically, this is expressed as

$$\mathbb{P}\left(X_t \in A \mid X_0, X_1, \ldots, X_{t-1}\right) = \mathbb{P}\left(X_t \in A \mid X_{t-1}\right) \qquad [76]$$

for any set $A \in \mathcal{F}$, where $P\left(x, A\right) = \mathbb{P}\left(X_t \in A \mid X_{t-1} = x\right)$ is a conditional probability distribution that is referred to as the transition kernel of the chain (Roberts 1996; Robert

and Casella 2004). For illustration, suppose that the state space of the Markov chain is finite e.g. $\Omega = \{x_1, x_2, \ldots, x_k\}$, then the transition kernel P is a k by k matrix (referred to as the transition matrix) such that $P_{ij} = \mathbb{P}\left(X_t = x_j \mid X_{t-1} = x_i\right)$ with the properties that $P_{ij} \geq 0$ and $\sum_{j=1}^{k} P_{ij} = 1$ for all $x_i, x_j \in \Omega$ (Häggström 2002). In a finite state-space Markov chain, the i^{th} row of the transition matrix gives the probabilities of transitioning from the i^{th} state to all k possible states of the chain. In the context of simulating stochastic point processes and performing Bayesian inference for them, the state spaces will consist of sets of real-valued point configurations in $\mathbb{R}^2$ and sets of real-valued parameter vectors, which require an infinite state space theory for Markov chains. While only a few key principles of continuous state space Markov chains are provided here, detailed and extensive theory is available in the literature (Meyn and Tweedie 1993).

To obtain samples from a target distribution $f(X)$ through a Markov chain, it is essential that there exists an invariant σ-finite measure π, defined for a Markov chain with transition kernel P that satisfies

$$\pi(B) = \int_{\Omega} P(x, B)\,\pi(dx) \qquad [77]$$

for all Borel sets B of Ω (Robert and Casella 2004, p. 223). If this measure is also a probability measure, then π is referred to as the stationary distribution of the chain. Eq. 77 implies that if $X_t \sim \pi$, then consequently $X_{t+1} \sim \pi$, and all subsequent states of the chain X_{t+2} X_{t+3}, … will also be distributed by the stationary distribution. The key idea of MCMC is that a Markov chain is constructed with a stationary distribution that is the

target distribution μ from which samples are desired. Furthermore, as $t \to \infty$, the distribution of the random variables X_t from the Markov chain will converge in total variation to the stationary distribution, regardless of what the initial state X_0 is specified to be. In order for this to happen, the Markov chain must be Harris recurrent, ψ-irreducible, and aperiodic, conditions which loosely speaking, allow the chain to freely communicate from one state to another.

To understand recurrence, irreducibility, and aperiodicity, the notion of the transition kernel needs to be expanded. Consider the distribution of the Markov chain after the n^{th}-transition $P^n(x, A)$ defined as (Robert and Casella 2004, p. 210)

$$P^n(x, A) = \int_{\Omega} P^{n-1}(y, A) P(x, dy) \qquad [78]$$

In a finite state space setting, the n^{th}-transition kernel can be obtained by simple matrix multiplication, that is given $\Omega = \{x_1, x_2, \ \ldots \ , x_k\}$ and the k by k transition matrix P, then $P^n = P \cdot P \cdot \ \ldots \ \cdot P$. The n^{th}-transition matrix has the interpretation that if P_{ij} represents the probability of being in state j after one transition given starting in state i, then $P^n{}_{ij}$ represents the probability of being in state j after n transitions given starting in state i. Similarly then $P^n(x, A)$ can be interpreted as giving the probability of $X_n \in A$ given $X_0 = x$ for any $A \in \mathcal{F}$.

The property of irreducibility is fundamental to Markov chains in that it implies that all states can communicate with each other. Intuitively, this notion might be expressed through the probability that given the chain currently is in state x, that there is a

non-zero probability of reaching some set A in a finite number of transitions. Mathematically this can be stated as saying that there exists an integer n such that $P^n(x, A) > 0$ for all $x \in \Omega$ and Borel sets A. Unfortunately it is possible that $P^n(x, A) = 0$ for some x and A, so an auxiliary measure ψ is introduced with the restriction placed on the definition above so that $P^n(x, A) > 0$ for all $x \in \Omega$ and Borel sets A where $\psi(A) > 0$. To emphasize the dependence upon the measure ψ, the property of irreducibility is referred to as ψ-irreducibility (Robert and Casella 2004, p. 213).

Recurrence is the notion that the Markov chain can revisit a set A an infinite number of times. More specifically, let η_A be the number of times the Markov chain passes through the set A, that is $\eta_A = \sum_{t=1}^{\infty} I_A(X_t)$ (Robert and Casella 2004, p. 212). A ψ-irreducibile Markov chain is defined to be recurrent (Robert and Casella 2004, p. 220) if for every Borel set A where $\psi(A) > 0$, $\mathbb{E}(\eta_A) = \infty$ for every $x \in A$; otherwise the chain is said to be transient. A stronger property of recurrence is Harris recurrence, where a set A is Harris recurrent if $\mathbb{P}(\eta_A = \infty \mid X_0 = x) = 1$ for every $x \in A$. When a Markov chain is said to be Harris recurrent, the chain is ψ-irreducible and for every set A that $\psi(A) > 0$, A must be a Harris recurrent set (Robert and Casella 2004, p. 222). To put it another way, recurrence means that the average number of visits of the chain to A is infinite, whereas Harris recurrence states that, given the chain starts at some arbitrary value x, the chain will visit A an infinite number of times with probability one.

A somewhat different notion of recurrence that is frequently specified for Markov chains is positive recurrence. Whereas Harris recurrence is a property that is primarily used to banish measure theoretic pathologies incurred by null sets (Geyer 1999), positive

recurrence is a property that implies the existence of the stationary distribution π. Specifically, a ψ-irreducible Markov chain is positive recurrent when there exists a σ-finite invariant probability measure π for the chain (the stationary distribution) that satisfies Eq. 77 (Robert and Casella 2004). A sufficient, though not necessary condition to establish the existence of the stationary distribution is through the concept of time-reversibility, or detailed balance. Informally, detailed balance is a condition on the Markov chain which implies that the probability of moving forward in time from $X_{t-1} = x$ to $X_t = y$ is the same as the probability of moving backward from y to x. Formally, a Markov chain with transition kernel P satisfies detailed balance if there exists a function f such that

$$P(y,x)f(y) = P(x,y)f(x) \qquad [79]$$

Furthermore, when f is the probability distribution π, detailed balance implies (Robert and Casella 2004, p. 230)

$$
\begin{aligned}
\int_{\Omega} P(x,B)\pi(dx) &= \int_{\Omega} P(x,B)\pi(x)dx \\
&= \int_{\Omega}\int_{B} P(x,y)\pi(x)dydx \\
&= \int_{\Omega}\int_{B} P(y,x)\pi(y)dydx \quad \text{(detailed balance)} \\
&= \int_{B}\int_{\Omega} P(y,x)\pi(y)dxdy \quad \text{(Fubini's Theorem)} \\
&= \int_{B} \pi(y)\int_{\Omega} P(y,x)dxdy \\
&= \int_{B} \pi(y)dy \quad \text{(since } \int_{\Omega} P(y,x)dx = 1\text{)} \\
&= \int_{B} \pi(dy) \\
&= \pi(B)
\end{aligned}
$$

for any measurable B, so that π is the stationary distribution of the chain as required in Eq. 77.

In order to establish that that the distribution of the Markov chain will in fact converge to the stationary distribution, the chain must have the property of aperiodicity. Intuitively, aperiodicity of a chain means that the chain could start at $X_t = x$ and return to it in only one transition. In contrast, a periodic chain is one that needs d transitions to return to x, for integer d greater than one. In the case of irreducible Markov chains on finite state spaces with transition matrix P, the period of the chain is the greatest common divider of $\left\{n \geq 1 : P_{xx}^{n}(t) > 0\right\}$. The chain is then defined to be aperiodic when the period of the chain is one for all states $x \in \Omega$ (Häggström 2002, p. 25). The notion of aperiodicity can of course be extended to infinite state spaces, but it is less intuitive and requires the concepts of minorizing measures and small sets (Robert and Casella 2004, p. 214-217). When aperiodicity is combined with the properties of Harris recurrence (which

implies ψ-irreducibility) and positive recurrence, the chain is said to be ergodic (Tierney 1994; Møller and Waagepetersen 2004).

With the various properties of irreducibility, recurrence, and aperiodicity in hand, the theoretical basis for the convergence of the limiting distribution of the Markov chain (that is the sequence X_t as $t \to \infty$) to the stationary distribution π can be established. To provide meaning to the concept of convergence of a sequence of probability distributions, the total variation norm is defined as

$$\left\| \mu - \nu \right\|_{TV} = \sup_{A \subseteq \Omega} \left| \mu(A) - \nu(A) \right| \qquad [80]$$

for any two probability measures μ and ν defined on Ω (Møller and Waagepetersen 2004, p. 121). In the case of a countable state space, it can be shown (Robert and Casella 2004, p.234) that for a positive recurrent, irreducible, and aperiodic Markov chain that for every initial state $X_0 = x$ that

$$\lim_{n \to \infty} \left\| P^n(x, \cdot) - \pi \right\|_{TV} = 0 \qquad [81]$$

In the case of an uncountably infinite state space, Tierney (1994) shows that a positive recurrent, π-irreducible, and aperiodic Markov chain with stationary distribution π will satisfy Eq. 81 for almost every initial state $X_0 = x$; when the chain is Harris positive recurrent then Eq. 81 will hold for every initial state $X_0 = x$. Robert and Casella (2004, p. 234) present the theorem that, given a Markov chain that is Harris recurrent and aperiodic with stationary distribution π,

$$\lim_{n\to\infty}\left\|\mathbb{E}\left[P^n(x,\cdot)\right]-\pi\right\|_{TV}=0 \qquad [82]$$

where $\mathbb{E}\left[P^n(x,\cdot)\right]=\int P^n(x,\cdot)\,\mu(dx)$ for every initial distribution μ An important consequence of this theorem is that the expected value of any bounded function $g(X_n)$ will converge to the expected value of $g(X)$ under π, given any initial distribution of the chain μ, i.e. (Robert and Casella 2004, p. 235)

$$\lim_{n\to\infty}\left|\mathbb{E}_{\mu}\left[g(X_n)\right]-\mathbb{E}^{\pi}\left[g(X)\right]\right|=0 \qquad [83]$$

This theorem does not necessarily justify the approximation given by the Monte Carlo integration in Eq. 75 when the samples $X_t = x_t$ are generated by an ergodic Markov chain however. The marginal distribution of X_n represents the probabilistic behavior of the chain at some given point in time n, whereas the m samples x_t represent the realization of a particular path of the Markov chain. To insure that the approximation in Eq. 75 is valid, the "Ergodic Theorem" is needed which gives the convergence of partial sums of the form $S_n(g)=\frac{1}{n}\sum_{i=1}^{n}g(X_i)$ to their expectation under the invariant measure π.

Specifically, let g and h be integrable functions. If a Markov chain has an invariant measure π and is Harris recurrent, then

$$\lim_{n\to\infty}\frac{S_n(g)}{S_n(h)}=\frac{\int g(x)\pi(dx)}{\int h(x)\pi(dx)}=\frac{\mathbb{E}^{\pi}\left[g(x)\right]}{\mathbb{E}^{\pi}\left[h(x)\right]} \qquad [84]$$

provided that $\mathbb{E}^{\pi}\left[h(x)\right] \neq 0$ (Robert and Casella 2004, p. 241-242). In effect, Eq. 84 says that provided *n* is large enough, the ergodic average $S_n(g)$ will approximate its expectation under the stationary distribution. In practice, the transition kernel of a Markov chain is constructed so that a desired stationary distribution exists. The chain is initially allowed to proceed a certain number of iterations that are discarded, called the burn-in. After burn-in, the transition kernel of the chain is considered to be "close enough" to $\pi(X)$ so that realized states of the Markov chain obtained after burn-in are used to form a dependent sample for use in Eq.75 (Gilks et al. 1996). There is however, no automatic way to determine what the length of the burn-in period should be. While different diagnostics have been proposed to address this issue (Cowles and Carlin 1996), they are better characterized as demonstrating clear lack of convergence of a Markov chain, rather than demonstrating that the chain has in fact converged to the stationary distribution. Some authors (Casella et al. 2001) have argued that the major flaw of MCMC is that its convergence is only valid asymptotically, and a poor choice of burn-in would lead to serious biases in the inference drawn from the output of Markov chains.

2.42 The Metropolis-Hastings algorithm

The Metropolis-Hastings algorithm, first proposed by Metropolis et al. (1953), and later generalized by Hastings (1970), is one of the most well-known algorithmic implementations of MCMC due to its flexibility and wide applicability for sampling from

probability distributions. In its most basic form, the Metropolis-Hastings algorithm is the the following (Robert and Casella 2004, p. 270):

1. Given a probability distribution μ with a density $f(X)$ function having connected support $\mathcal{E}$, e.g. $\mathcal{E} = \mathbb{R}^d$, choose the initial state $x^{(0)} \in \mathcal{E}$ such that $f\left(x^{(0)}\right) > 0$ and set $t = 0$.
2. Generate a sample x^* from a desired proposal distribution q whose support includes $\mathcal{E}$ so that $X^* \sim q\left(x^* \mid x^{(t)}\right)$.
3. Assign the next state of the Markov chain according to the decision rule

$$X^{(t+1)} = \begin{cases} X^*, \text{ with probability } \alpha\left(x^{(t)}, x^*\right) \\ X^t, \text{ with probability } 1 - \alpha\left(x^{(t)}, x^*\right) \end{cases}$$

   where the acceptance probability $\alpha\left(x^{(t)}, x^*\right)$ is defined as

$$\alpha\left(x^{(t)}, x^*\right) = \min\left\{1, \frac{f\left(x^*\right)}{f\left(x^{(t)}\right)} \frac{q\left(x^{(t)} \mid x^*\right)}{q\left(x^* \mid x^{(t)}\right)}\right\} \qquad [85]$$

4. Increment t and repeat until $t = T$, the chosen number of total iterations for the Markov chain.

The Metropolis-Hastings algorithm is quite flexible in that theoretically any proposal distribution could be used for the target density f subject to the relatively minor conditions regarding the support of the proposal. In practice, the choice of the proposal is

very important and needs to be selected with care so that the Markov chain draws samples throughout the support of *f* quickly, i.e. the chain mixes well.

By design the Metropolis-Hastings algorithm is constructed so that *f* is the stationary distribution of the chain, and the proof follows from the fact that the algorithm satisfies detailed balance. Specifically, assume without loss of generality that the Hastings ratio

$$\frac{f\left(x^{*}\right)}{f\left(x^{(t)}\right)}\frac{q\left(x^{(t)} \mid x*\right)}{q\left(x* \mid x^{(t)}\right)} > 1$$

so that the acceptance probabilities $\alpha\left(x^{(t)}, x^{*}\right) = 1$ and

$$\alpha\left(x^{*}, x^{(t)}\right) = \frac{f\left(x^{(t)}\right)}{f\left(x^{*}\right)}\frac{q\left(x^{*} \mid x^{(t)}\right)}{q\left(x^{(t)} \mid x^{*}\right)}$$

The transition kernel of the Metropolis-Hastings algorithm is (Robert and Casella 2004, p. 272)

$$P(x,y) = \alpha(x,y)\,q(y \mid x) + \left(1 - \int \alpha(x,y)\,q(y \mid x)\,dy\right)\delta_{x}(y) \qquad [86]$$

where δ_{x} is the point mass in *x*.

Letting $x = x^{(t)}$ and $y = x^{*}$,

$$
\begin{aligned}
P\left(x^{(t)},x^{*}\right)f\left(x^{(t)}\right) &= \left[\alpha\left(x^{(t)},x^{*}\right)q\left(x^{*}\mid x^{(t)}\right)+\left(1-\int\alpha\left(x^{(t)},x^{*}\right)q\left(x^{*}\mid x^{(t)}\right)dx^{*}\right)\delta_{x^{(t)}}\left(x^{*}\right)\right]f\left(x^{(t)}\right)\\
&= \left[q\left(x^{*}\mid x^{(t)}\right)+\left(1-\int q\left(x^{*}\mid x^{(t)}\right)dx^{*}\right)\delta_{x^{(t)}}\left(x^{*}\right)\right]f\left(x^{(t)}\right)\\
&= \left[q\left(x^{*}\mid x^{(t)}\right)+\left(1-1\right)\delta_{x^{(t)}}\left(x^{*}\right)\right]f\left(x^{(t)}\right)\\
&= \left[q\left(x^{*}\mid x^{(t)}\right)+0\right]f\left(x^{(t)}\right)\\
&\quad q\left(x^{*}\mid x^{(t)}\right)f\left(x^{(t)}\right)
\end{aligned}
$$

Likewise,

$$
\begin{aligned}
P\left(x^{*},x^{(t)}\right)f\left(x^{*}\right) &= \left[\alpha\left(x^{*},x^{(t)}\right)q\left(x^{(t)}\mid x^{*}\right)+\left(1-\int\alpha\left(x^{*},x^{(t)}\right)q\left(x^{(t)}\mid x^{*}\right)dx^{(t)}\right)\delta_{x^{*}}\left(x^{(t)}\right)\right]f\left(x^{*}\right)\\
&= \frac{f\left(x^{(t)}\right)}{f\left(x^{*}\right)}\frac{q\left(x^{*}\mid x^{(t)}\right)}{q\left(x^{(t)}\mid x^{*}\right)}q\left(x^{(t)}\mid x^{*}\right)f\left(x^{*}\right)\\
&\qquad +\left(1-\int\frac{f\left(x^{(t)}\right)}{f\left(x^{*}\right)}\frac{q\left(x^{*}\mid x^{(t)}\right)}{q\left(x^{(t)}\mid x^{*}\right)}q\left(x^{(t)}\mid x^{*}\right)dx^{(t)}\right)\delta_{x^{*}}\left(x^{(t)}\right)f\left(x^{*}\right)\\
&= f\left(x^{(t)}\right)q\left(x^{*}\mid x^{(t)}\right)+\left(1-\int\frac{f\left(x^{(t)}\right)}{f\left(x^{*}\right)}q\left(x^{*}\mid x^{(t)}\right)dx^{(t)}\right)\delta_{x^{*}}\left(x^{(t)}\right)f\left(x^{*}\right)\\
&= f\left(x^{(t)}\right)q\left(x^{*}\mid x^{(t)}\right)+\left(f\left(x^{*}\right)-\int f\left(x^{*}\right)\frac{f\left(x^{(t)}\right)}{f\left(x^{*}\right)}q\left(x^{*}\mid x^{(t)}\right)dx^{(t)}\right)\delta_{x^{*}}\left(x^{(t)}\right)\\
&= f\left(x^{(t)}\right)q\left(x^{*}\mid x^{(t)}\right)+\left(f\left(x^{*}\right)-\int f\left(x^{(t)}\right)q\left(x^{*}\mid x^{(t)}\right)dx^{(t)}\right)\delta_{x^{*}}\left(x^{(t)}\right)\\
&= f\left(x^{(t)}\right)q\left(x^{*}\mid x^{(t)}\right)+\left(f\left(x^{*}\right)-\int f\left(x^{*},x^{(t)}\right)dx^{(t)}\right)\delta_{x^{*}}\left(x^{(t)}\right)\\
&= f\left(x^{(t)}\right)q\left(x^{*}\mid x^{(t)}\right)+\left(f\left(x^{*}\right)-f\left(x^{*}\right)\right)\delta_{x^{*}}\left(x^{(t)}\right)\\
&= f\left(x^{(t)}\right)q\left(x^{*}\mid x^{(t)}\right)
\end{aligned}
$$

which implies that $P\left(x^{(t)}, x^{*}\right) f\left(x^{(t)}\right) = P\left(x^{*}, x^{(t)}\right) f\left(x^{*}\right)$ and detailed balance holds. In addition, it is well known that the Metropolis-Hastings algorithm is *f*-irreducible, Harris recurrent, and aperiodic (Robert and Casella 2004, p. 272-273). Consequently the Ergodic Theorem applies and satisfies the convergence of ergodic averages such as Eq. 75 to their true expectation under the stationary distribution *f*.

There are many ways to implement a Metropolis-Hastings algorithm in practice. Single-component Metropolis-Hastings (Gilks et al. 1996), where not all the components of *X* are updated in one block, the Gibbs sampler (Geman and Geman 1984; Gelfand and Smith 1990) which relies upon the specification of the full conditional distributions of *f*; and the independence sampler (Tierney 1994), which uses proposal distributions that do not depend upon the current value of *X* are just a few of the better known variants of the Metropolis-Hastings algorithm.

3. Methods

3.1 Data

For the analysis of tree – tree interactions through stochastic Gibbs processes, data were obtained from five 80 x 80m (0.64 ha) stem-mapped plots that form the terrestrial gradient tree growth survey (Clark et al. 1998), at the Coweeta Hydrologic Laboratory's Long Term Ecological Research (LTER) site located in western North Carolina (35.04ºN, 83.45ºW). Plots were established along elevation and moisture gradients in two watersheds forested by multi-aged, mixed hardwood stands, typical of forest ecosystems in the southern Appalachian Mountains.

All trees greater than 1m tall were stem-mapped on the gradient plots (Clark et al. 1998). Data recorded for each tree record include the tree tag number, species code, Cartetsian x- and y- coordinates, and diameter at breast height (dbh). Tree dbh data were collected on the gradient plots in 1993, 1995, 1996, 1998, and 2002; however, not all trees were measured at each survey due to changes in plot design. The 1998 re-measurement data were used for fitting and simulation of Gibbs point processes.

In addition to using *observed* data from Coweeta, many analyses were performed using *simulated* data from a specific Gibbs point process model with pre-specified parameter values. A key advantage in using simulated rather than observed data is that a check can be provided on the statistical inference for the model. For instance suppose a configuration **x*** is generated from a point process with parameters θ^*, so that $\mathbf{x}^* \sim f\left(\mathbf{x} \mid \theta^*\right)$. If Bayesian inference is performed on the point process model using non-

informative (uniform) priors, then the posterior distribution $f(\theta \mid \mathbf{x})$ should be concentrated on θ^*. If θ^* is at the extremes of the sampled values of θ for the posterior or θ^* is not sampled at all, then clearly the inferential procedure is producing poor results. For the research here, simulated plots were generated from a variety of models to test the Bayesian inference using Metropolis-Hastings algorithms. "Plot" sizes ranged from 10 x 10m to 30 x 30m, with parameter values chosen throughout the support of the Gibbs posterior densities.

3.2 Exploratory Data Analysis

Prior to fitting and simulation of Gibbs point process models, several preprocessing steps were carried out to ensure quality and to provide summaries of spatial structure. Data preprocessing included the following initial six steps: 1) reading in the raw data, 2) eliminating any trees from the database whose stem locations fall outside the plot boundaries, 3) eliminating any trees that were not alive in the 1998 remeasurement, 4) eliminating any trees with 1998 DBH $\leq$ 10 cm, 5) creating unique tree ID numbers from the tag numbers used on the trees, and 6) comparing the 1998 DBH data against other remeasurements to ensure the DBH data are free from data entry errors and are consistent across the remeasurements.

Data for the terrestrial gradient tree growth survey is publicly available from the Coweeta LTER website, http://coweeta.ecology.uga.edu. After downloading the data, tree locations were edited to eliminate multiple tree records at the same location. The existence of multiple tree records with the same coordinate location is a consequence of how stem locations were measured in the field, since it is possible that a tree (viewed as an organism) can have multiple stems coming from a common root system. If the stems separate below 1.3m (4.5 feet), standard diameter measurement protocols (Avery and Burkhart 2002) would treat this tree as having two measurable stems with tree records thus reflecting separate (distinguishable) stems rather than individual organisms. When stem-mapping in the field, it is possible that a tree with multiple stems will have the piths of the two or more stems so close to each other at ground level that their locations are practically indistinguishable, and both stems are given the same coordinate location.

While commonly encountered in forest inventory, the presence of multiple stems with the same location is undesirable for stochastic point process modeling. To keep the analysis on simple point processes, multiple tree records with the same coordinate location were modified as follows: 1) the tree with the largest DBH was kept at the recorded location of the group, 2) the second largest DBH tree had its *Y*-coordinate (using Cartesian coordinates) increased by its diameter, while unchanging its *X*-coordinate and 3) the next largest DBH tree had its *X*-coordinate increased by its DBH (Fig. 3). In practice this idea could be extended to larger groups of trees, but in the raw data from Coweeta there were no more than three trees at any given location. All modified coordinates were then checked to ensure that all trees had unique locations and were not outside the plot boundaries.

After data preprocessing, the spatial structure of the data was explored through two summary functions:

1) The *K*-function (Ripley 1976, 1977)

For a stationary and isotropic process the *K*-function is defined from the reduced second moment measure as shown (Eq. 35):

$$K(r) = \frac{2\pi}{\lambda^2} \int_0^r \rho_2(t) t \, dt$$

where r is the pairwise distance between points and λ is the intensity of the point process. An intuitive interpretation for the *K*-function is that the expected number of extra events

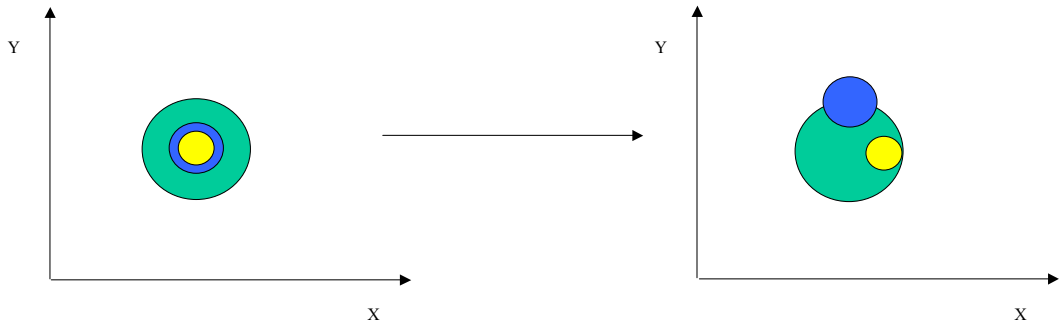

Figure 3. Stem location modification procedure for three hypothetical trees with the same coordinate location. Circle area represents tree basal area.

within r distance of any arbitrary event is equal to the product $\lambda K(r)$ (Ripley 1977). Under the assumptions of stationarity and isotropy, Ripley (1988) and Haase (1995) give an unbiased estimator for $K(r)$ as

$$\widehat{K}(r) = \frac{|F|}{n^2} \sum\sum_{i \neq j} w_{ij}^{-1} I_r(u_{ij}) \qquad [87]$$

where $|F|$ is the area of the observation window, n is the number of trees located in F, $I_r(u_{ij})$ is an indicator function which equals one when the distance between a pair of trees u_{ij} is less than or equal to the value r and equals zero otherwise, and w_{ij} is an edge correction factor. For this analysis the isotropic edge correction factor (Ripley 1988) is used which weights the indicator function I_r by the proportion of the circumference of a circle of radius u_{ij} centered at tree i that falls within the boundaries of the plot. To aid in

the interpretation of the *K*-function, the *L*-function, proposed by J. Besag in the discussion of Ripley's (1977) paper is used, where $\hat{L}(r) = \sqrt{\hat{K}(\mathrm{r})/\pi}$, so that a plot of $\hat{L}(r) - r$ under complete spatial randomness will equal zero for all *r*.

2) The pair correlation function (Stoyan and Stoyan 1994, 1996)

The pair correlation function, $g(r)$, given in Eq. 18, was proposed as a normalization of the second-order product density of a homogeneous point process:

$$g(r) = \frac{\rho_2(x_1, x_2)}{\lambda^2} = \frac{\rho_2(\|x_1 - x_2\|)}{\lambda^2} = \frac{\rho_2(r)}{\lambda^2}$$

Penttinnen et al. (1992) and Stoyan and Penttinen (2000) provide a heuristic to interpret the pair correlation function. If dx_1 and dx_2 represent the areas of infinitesimally small circles centered at x_1 and x_2 separated at the distance *r*, then the probability that both circles each contain exactly one event of a homogeneous point process, denoted $p_2(x_1, x_2)$, can be defined in terms of the pair correlation function as

$$p_2(x_1, x_2) = \rho_2(x_1, x_2)\, dx_1 dx_2 = \lambda^2 g(r)\, dx_1 dx_2 \qquad [88]$$

The pair correlation function of Eq. 88 is estimated by (Stoyan and Stoyan 1994)

$$\hat{g}(r) = \frac{1}{2\pi r \hat{\lambda}^2} \sum\sum_{i \neq j} \frac{k(r - u_{ij})}{w_{ij}} \qquad [89]$$

where $k(\cdot)$ is a kernel density estimator, $\hat{\lambda} = n/|F|$ is the observed intensity of the point process, u_{ij} is the distance between trees i and j, and w_{ij} is the isotropic edge correction factor. Following the recommendations of Stoyan and Stoyan (1994), the Epanečnikov kernel was used

$$k(x) = \begin{cases} \frac{3}{4h}\left(1 - \frac{x^2}{h^2}\right), & \text{for } -h \le x \le h \\ 0, & \text{otherwise} \end{cases} \qquad [90]$$

with bandwidth $h = 0.15/\sqrt{\hat{\lambda}}$.

3.3 Gibbs point process models

3.31 Pairwise interaction models

Competitive interactions between trees for light suggest that over time positions of surviving trees would be farther apart than a purely random pattern (Metsaranta and Lieffers 2008). The competitive influence of trees would decrease with distance from the stem location, suggesting a natural application of repulsive pair potential models where the potential decreases in some smooth fashion over distance, such as the model of Fiksel (1984). On the other hand, the ability of many Appalachian hardwood species to reproduce by stump sprouts suggests that at very small distances, clustering is also possible.

To capture these effects in the context of Gibbs pairwise interaction point processes, pair potential models were proposed that attempt to simulate repulsive competitive effects alone, or attempt to simulate short range clustering due to stump sprouting combined with repulsive competitive effects at larger distances. These models all incorporate finite interaction and hard-core distances, which are denoted R and r_0, respectively, and are expressed as functions of the distance between points $r_{ij} = \|x_i - x_j\|$.

The pair potential models for the Gibbs pairwise interaction processes are as follows (Acronyms in parentheses):

1) Fiksel potential (F), as published by Fiksel (1984)

$$\phi_2\left(r_{ij}\right)=\left\{\begin{array}{l}\infty,\ \text{for } r_{ij}\le r_0\\ \theta_1\exp\left(-\theta_2 r_{ij}\right),\ \text{for } r_0<r_{ij}\le R\\ 0,\ \text{for } r_{ij}>R\end{array}\right\},\ \theta_2>0 \qquad [91]$$

2) Hard-core modified exponential pair potential (HCME)

$$\phi_2\left(r_{ij}\right)=\left\{\begin{array}{l}\infty,\ \text{for } r_{ij}\le r_0\\ \dfrac{\theta_1}{r_{ij}^2+1}+\theta_2\exp\left(-\theta_3 r_{ij}\right),\ \text{for } r_0<r_{ij}\le R\\ 0,\ \text{for } r_{ij}>R\end{array}\right\},\ \theta_3>0 \qquad [92]$$

3) Hard-core serpentine pair potential (HCSP)

$$\phi_2\left(r_{ij}\right)=\left\{\begin{array}{l}\infty,\ \text{for } r_{ij}\le r_0\\ \theta_1\left(\dfrac{r_{ij}+\theta_2}{r_{ij}^3+1}\right),\ \text{for } r_0<r_{ij}\le R\\ 0,\ \text{for } r_{ij}>R\end{array}\right\} \qquad [93]$$

where θ_1, θ_2, and θ_3 are parameters.

The pair potential model published by Fiksel (1984) was designed to be purely repulsive, with the degree of the spatial interaction decreasing exponentially with distance. For the purposes here, the potential is also allowed to be attractive by letting the parameter θ_1 take on negative values. This gives a pair potential that is either purely

repulsive or purely attractive but which cannot be attractive and repulsive at different distances (Fig. 4). The HCME potential modifies Fiksel's model by adding the product of $1/(r^2 + 1)$ and the parameter θ_1 to Fiksel's potential. This term influences the behavior of the potential at small distances such that the potential can be attractive (i.e. negative). Whenever the parameter θ_1 is negative and large in absolute value relative to θ_2, the pair potential will be attractive throughout its interaction range. Unlike the Fiksel potential, the HCME potential can be both attractive at small distances and repulsive at larger distances, which occur when θ_1 is positive and θ_2 is negative, (Fig. 5).

The hard-core serpentine pair potential (HCSP) was inspired by Newton's serpentine function, given by $f(x) = abx/(x^2 + a^2)$ (Weisstein 2007). The HCSP model has much the same flexibility of the HCME model but does so with fewer parameters. It can be purely repulsive when both θ_1 and θ_2 are positive, and when θ_1 is positive and θ_2 is negative and small in absolute value relative to θ_1 the potential is attractive at small distances and repulsive at larger distances. When θ_1 is negative and θ_2 is positive, the potential can be purely attractive throughout its interaction range (Fig. 6 – 9).

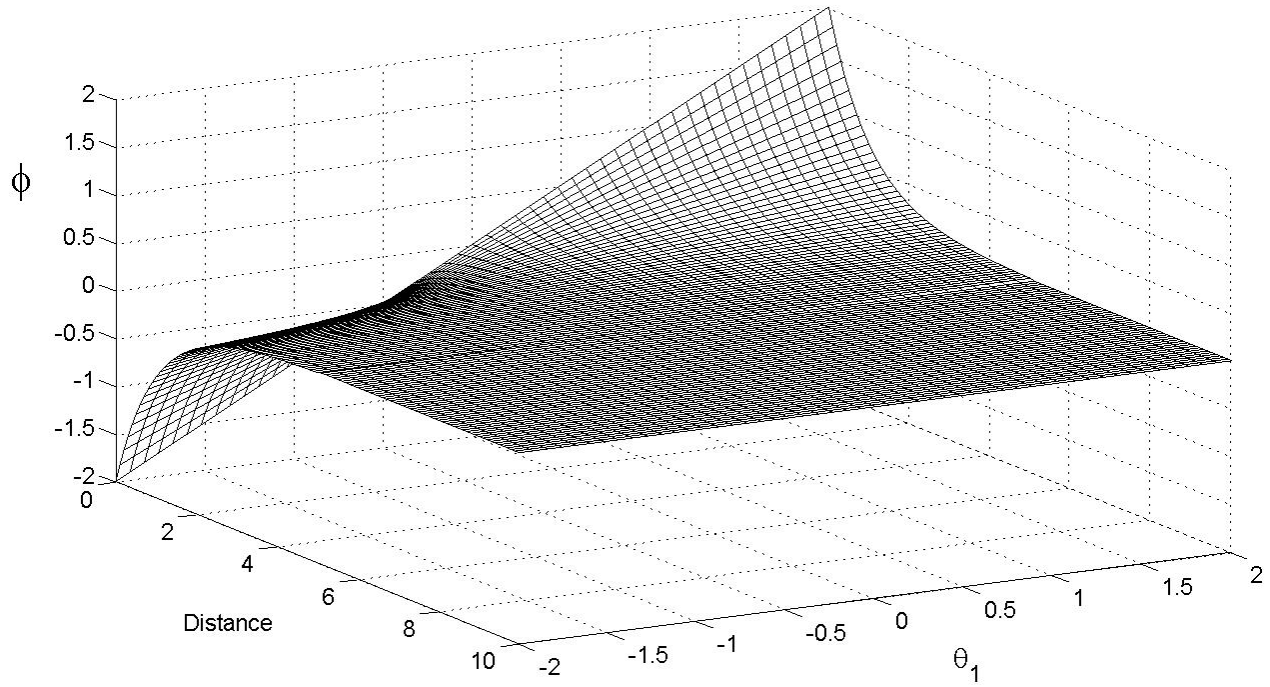

Figure 4. Fiksel potential with $\theta_2 = 1$, $r_0 = 0$, and $R = 5$.

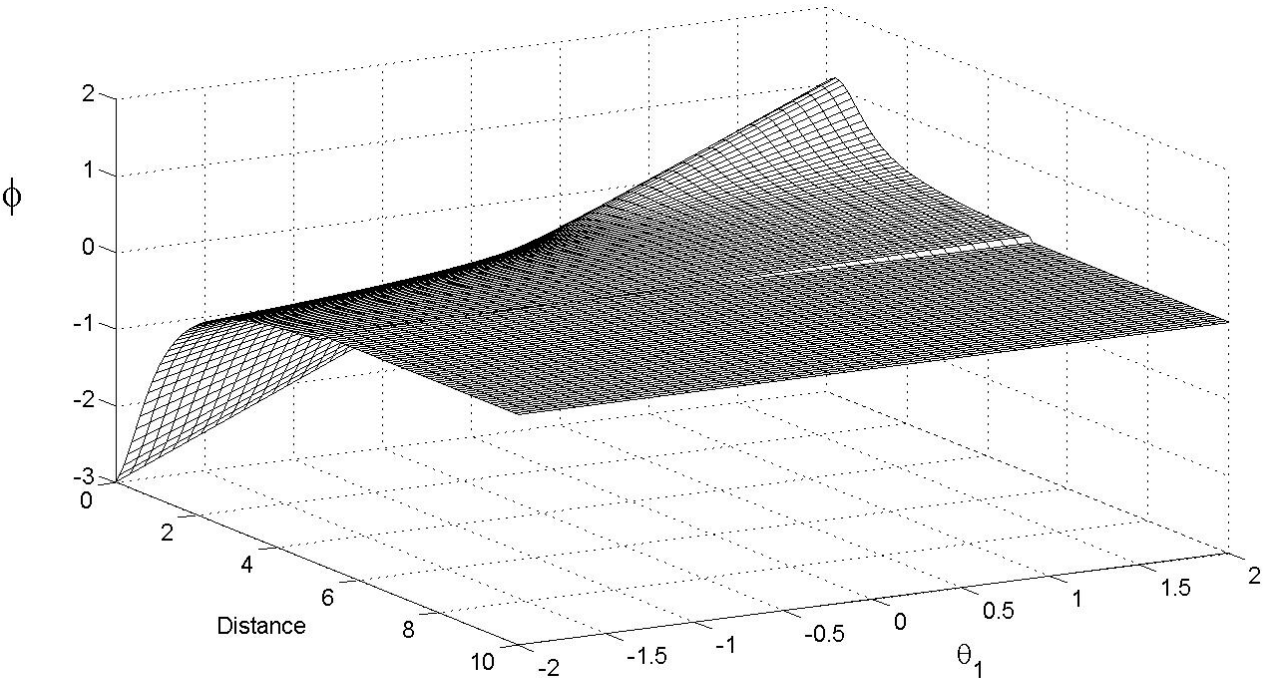

Figure 5. Hard-core modified exponetial potential with $\theta_2 = -1$, $\theta_3 = 1$, $r_0 = 0$, and $R = 5$.

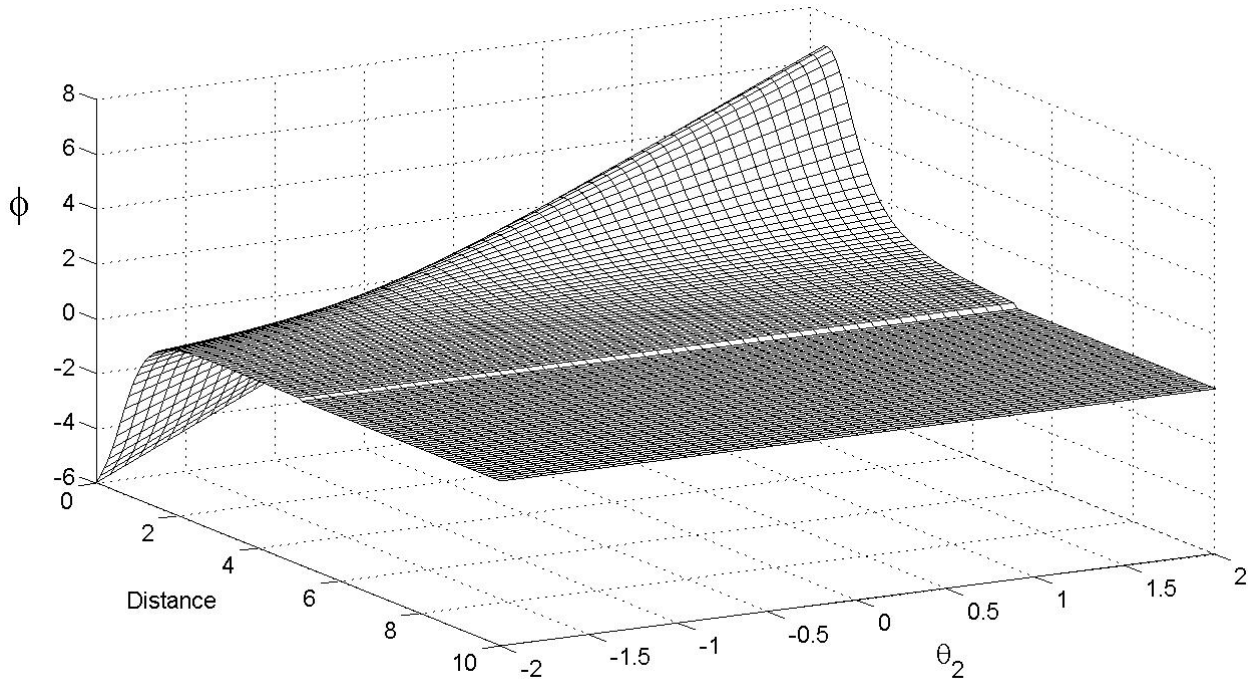

Figure 6. Hard-core serpentine potential with $\theta_1 = 2$, $r_0 = 0$, and $R = 5$.

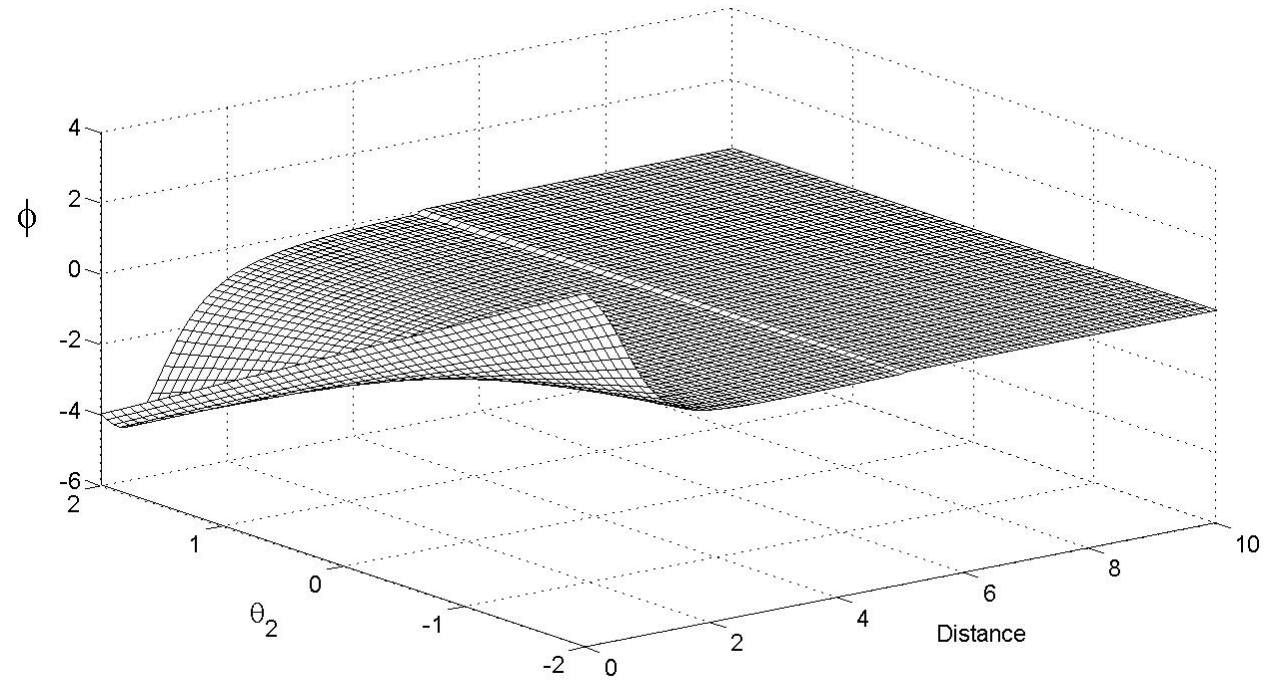

Figure 7. Hard-core serpentine potential with $\theta_1 = -2$, $r_0 = 0$, and $R = 5$.

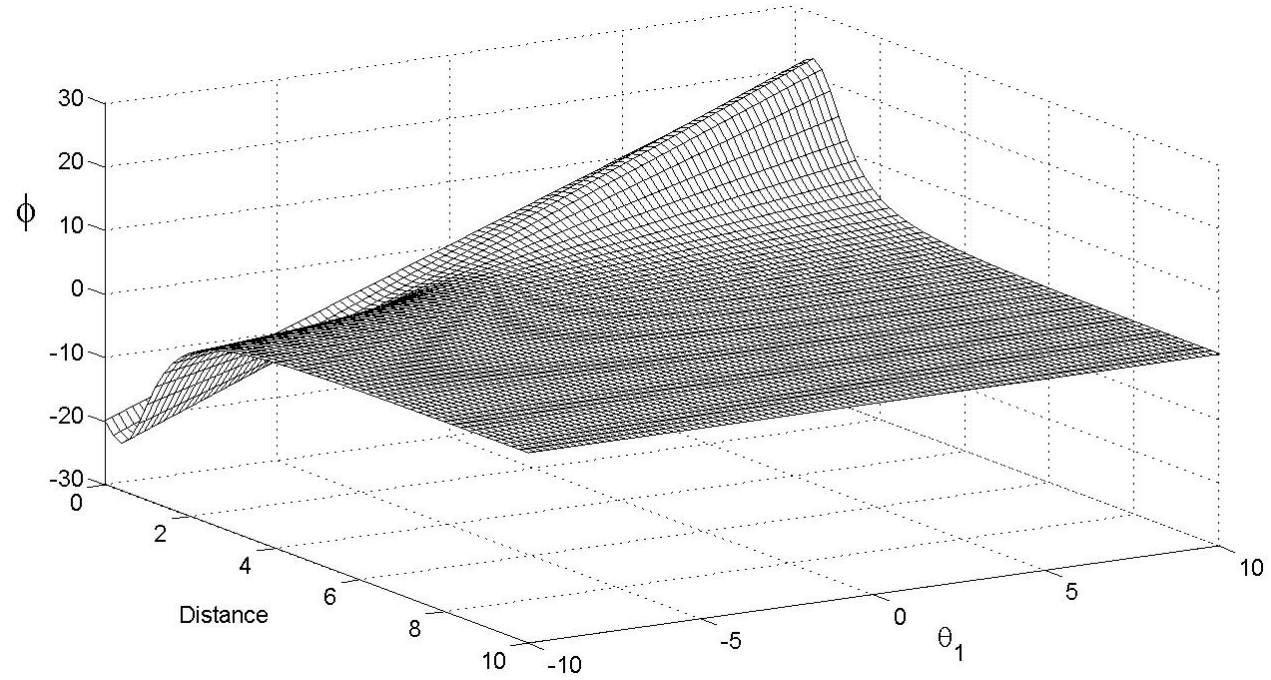

Figure 8. Hard-core serpentine potential with $\theta_2 = 2$, $r_0 = 0$, and $R = 5$.

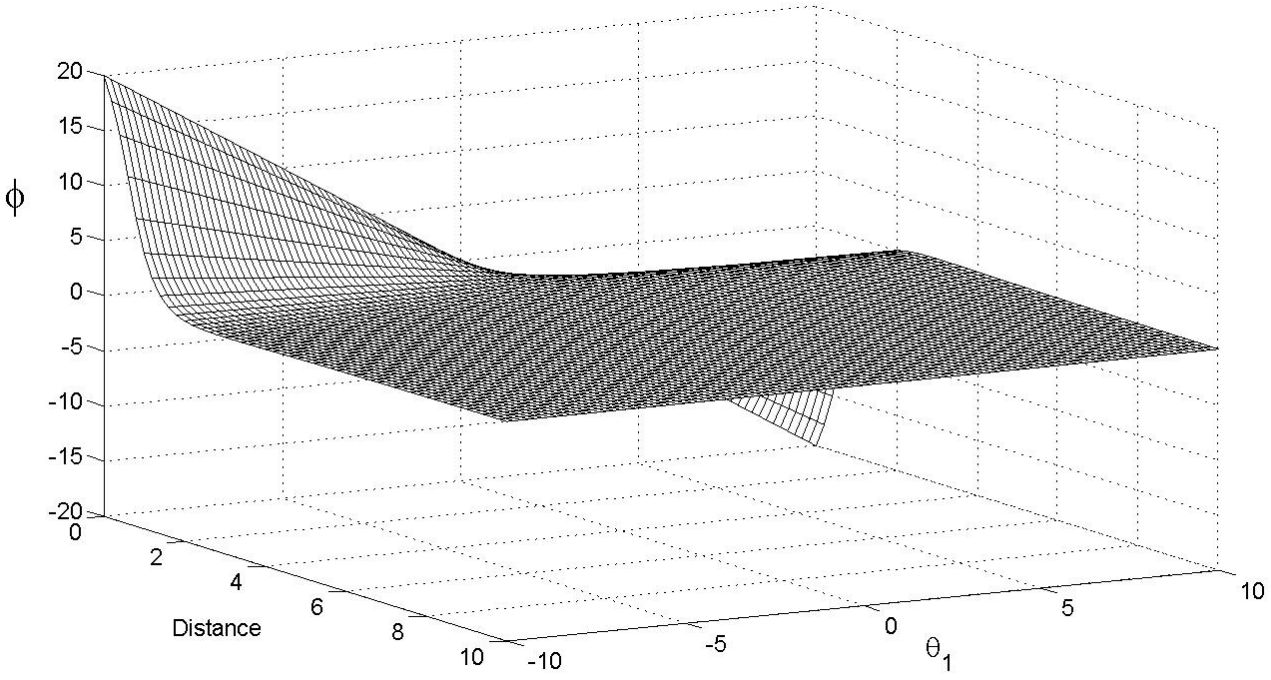

Figure 9. Hard-core serpentine potential with $\theta_2 = -2$, $r_0 = 0$, and $R = 5$.

3.32 Triplets models

An alternative method to allow attractive interactions in Gibbs and Markov point processes incorporates a third-order potential interaction. As discussed by Geyer (1999), a "triplets" model allows for attractive pairwise interactions by also requiring repulsive third-order interactions that create a limit on the number points that may be arbitrarily close to a given point. Mathematically, a triplets process with constant intensity function may be expressed as

$$
\begin{aligned}
f(\mathbf{x}) &= \alpha \prod_{i=1}^{n(\mathbf{x})} b(x_i) \prod_{i<j} h_2(x_i, x_j) \prod_{i<j<k} h_3(x_i, x_j, x_k) \\
&= \alpha \prod_{i=1}^{n(\mathbf{x})} \beta \prod_{i<j} \exp\left[-\phi_2\left(\|x_i - x_j\|\right)\right] \prod_{i<j<k} \exp\left[-\phi_3\left(\|x_i - x_j\|,\ \|x_i - x_k\|,\ \|x_j - x_k\|\right)\right] \\
&= \alpha\beta^{n(\mathbf{x})} \prod_{i<j} \exp\left[-\phi_2(r_{ij})\right] \prod_{i<j<k} \exp\left[-\phi_3(r_{ij}, r_{ik}, r_{jk})\right] \\
&= \alpha\beta^{n(\mathbf{x})} \exp\left[\sum_{i<j} -\phi_2(r_{ij})\right] \exp\left[\sum_{i<j<k} -\phi_3(r_{ij}, r_{ik}, r_{jk})\right] \qquad [94]
\end{aligned}
$$

where α is a normalizing constant, β is the intensity parameter, and ϕ_2 and ϕ_3 are the 2nd- and 3rd-order interaction potentials, respectively. Triplets processes have received little attention (Mateu and Montes 2001) since their introduction to the literature by Fiksel (1988) and Geyer (1999). One possible explanation for this is that triplets models are viewed as difficult to interpret (Møller and Waagepetersen 2007).

Despite this concern, triplets models are explored here as a viable means of incorporating attraction of points in a Gibbs point process. The 2nd- and 3rd-order potential models for the triplets processes studied here are as follows (Acronyms in parentheses):

1) Exponential triplets (EXPTP)

2nd-order potential:

$$\phi_2\left(r_{ij}\right)=\begin{Bmatrix}\infty,\ \text{for } r_{ij}\le r_0\\ \theta_1\exp\left(-r_{ij}\right),\ \text{for } r_0<r_{ij}\le R\\ 0,\ \text{for } r_{ij}>R\end{Bmatrix} \qquad [95]$$

3rd-order potential:

$$\phi_3\left(r_{ij},r_{ik},r_{jk}\right)=\begin{Bmatrix}\theta_2\exp\left(-\dfrac{\left(r_{ij}+r_{ik}+r_{jk}\right)}{3}\right)\ \text{for } r_{ij},r_{ik},r_{jk}\le R\\ 0\ \text{ for any } r_{ij},r_{ik},r_{jk}>R\end{Bmatrix},\ \theta_2>0 \qquad [96]$$

2) Geyer triplets (GTP), modified from Geyer (1999)

2nd-order potential:

$$\phi_2\left(r_{ij}\right)=\begin{Bmatrix}\infty,\ \text{for } r_{ij}\le r_0\\ \theta_1,\ \text{for } r_0<r_{ij}\le R\\ 0,\ \text{for } r_{ij}>R\end{Bmatrix} \qquad [97]$$

3rd-order potential:

$$\phi_3\left(r_{ij}, r_{ik}, r_{jk}\right) = \left\{ \begin{matrix} \theta_2 \ \text{ for } \ 0 \le r_{ij}, r_{ik}, r_{jk} \le R \\ 0 \ \text{ for any } \ r_{ij}, r_{ik}, r_{jk} > R \end{matrix} \right\}, \theta_2 > 0 \qquad [98]$$

3) Serpentine triplets (SPTP)

2nd-order potential:

$$\phi_2\left(r_{ij}\right) = \left\{ \begin{matrix} \infty \ \text{ for } \ r_{ij} \le r_0 \\ \theta_1 \left(\dfrac{r_{ij} + \theta_2}{r_{ij}^3 + 1} \right) \ \text{ for } \ r_0 < \ r_{ij} \le R \\ 0 \ \text{ for } \ r_{ij} > R \end{matrix} \right\} \qquad [99]$$

3rd-order potential:

$$\phi_3\left(r_{ij}, r_{ik}, r_{jk}\right) = \left\{ \begin{matrix} \theta_3 \exp\left(-\dfrac{\left(r_{ij} + r_{ik} + r_{jk}\right)}{3} \right) \ \text{ for } \ r_{ij}, r_{ik}, r_{jk} \le R \\ 0 \ \text{ for any } \ r_{ij}, r_{ik}, r_{jk} > R \end{matrix} \right\}, \theta_3 > 0 \qquad [100]$$

where θ_1, θ_2, and θ_3 are parameters. Aside from Geyer's (1999) model, the triplets processes proposed here all model the repulsive 3rd-order interaction as the exponential decay of the average of the three pairwise distances. All models require that 3rd-order interactions are non-zero only for triplets composed of mutual neighbors, where each pairwise distance is less than the value of the parameter R.

The Geyer triplets process is modified slightly from the "triplets" model published by Geyer (1999) in that a non-zero hard-core parameter (r_0) has been added. Also, the potentials shown for the GTP process reparameterizes the potentials given by Geyer (1999). Geyer expressed the triplets density as

$$f(\mathbf{x}) = \frac{1}{Z(\theta)} \exp\left[n(\mathbf{x})\theta_1 + s(\mathbf{x})\theta_2 + t(\mathbf{x})\theta_3 \right] \qquad [101]$$

where $n(\mathbf{x})$ is the number of points in the configuration $\mathbf{x}$, $s(\mathbf{x})$ is the number of pairs of points within the interaction distance R, and $t(\mathbf{x})$ is the number of triples of points that are composed only of mutual neighbors. Given that no pair of points can be closer in distance than the hard-core parameter, it is straightforward to see that for the GTP model specified above that

$$\sum_{i<j} -\phi_2(r_{ij}) = \sum_{i<j} -\theta_1 I_{[r_0,R]}(r_{ij}) = -\theta_1 \sum_{i<j} I_{[r_0,R]}(r_{ij}) = -\theta_1 s(\mathbf{x})$$

and

$$\begin{aligned} \sum_{i<j<k} -\phi_3(r_{ij}, r_{ik}, r_{jk}) &= \sum_{i<j<k} -\theta_2 I_{[0,R]}(r_{ij}) I_{[0,R]}(r_{ik}) I_{[0,R]}(r_{jk}) \\ &= -\theta_2 \sum_{i<j<k} I_{[0,R]}(r_{ij}) I_{[0,R]}(r_{ik}) I_{[0,R]}(r_{jk}) \\ &= -\theta_2 t(\mathbf{x}) \end{aligned}$$

so that the parameters in Geyer's publication can be related to the parameters for the GTP model by $\theta_1 = \ln(\beta)$, $\theta_2 = -\theta_1$, and $\theta_3 = -\theta_2$. Provided that the hard-core and interaction distances are fixed, Eq. 101 demonstrates that the GTP density is a member of an exponential family (Casella and Berger 2002, p.217). The EXPTP model is also an exponential family density with canonical parameters $(\ln(\beta), -\theta_1, -\theta_2)$, provided that the hard-core and interaction distances are fixed. The SPTP model on the other hand, does not belong to an exponential family unless the parameter θ_2 is fixed to some constant in addition to having fixed hard-core and interaction distances. In the language of Baddeley and Turner (2000), the parameter θ_2 for this model is an "irregular" parameter.

3.4 Edge effect correction

The Gibbs point models proposed account for interaction with trees that are outside the plot boundaries by means of an isotropic edge correction factor. The idea of adding an edge-correction weight to the potential function of a Gibbs or Markov point process was first introduced by Diggle et al. (1994) who proposed modifying the product of the interaction function of a pairwise interaction process

$$\prod_{i<j} h\left(x_i, x_j\right) = \prod_{i<j} \exp\left[-\phi\left(\left\|x_i - x_j\right\|\right)\right] = \exp\left[-\sum_{i<j} \phi\left(\left\|x_i - x_j\right\|\right)\right] \qquad [102]$$

with an edge-corrected version given by

$$\exp\left[-\frac{1}{2}\sum_{i\neq j} w_{ij}^{-1}\phi\left(\left\|x_i - x_j\right\|\right)\right] \qquad [103]$$

where the weight w_{ij} is the proportion of the circumference of the circle centered at x_i with radius that falls within the plot. Now since the interaction between points is symmetric, that is $\phi\left(\left\|x_i - x_j\right\|\right) = \phi\left(\left\|x_j - x_i\right\|\right)$, by expanding the sum in Eq.102 and combining like terms it is straightforward to see that

$$\exp\left[-\frac{1}{2}\sum_{i\neq j} w_{ij}^{-1}\phi\left(\left\|x_i - x_j\right\|\right)\right] = \exp\left[-\sum_{i<j}\left(\frac{w_{ij}^{-1} + w_{ji}^{-1}}{2}\right)\phi\left(\left\|x_i - x_j\right\|\right)\right] \qquad [104]$$

so that the edge correction factor is an average of the reciprocal of the weights centered at point x_i and point x_j. Therefore let

$$e_{ij} = \frac{w_{ij}^{-1} + w_{ji}^{-1}}{2} \qquad [105]$$

denote this isotropic edge correction weight so that the pdf for Gibbs processes with 2nd-order interaction is given

$$f(\mathbf{x}) = \alpha\beta^{n(\mathbf{x})} \exp\left[-\sum_{i<j} e_{ij}\phi_2(\mathrm{r}) \right] \qquad [106]$$

and the pdf for Gibbs processes with 3rd-order interaction is given by

$$f(\mathbf{x}) = \alpha\beta^{n(\mathbf{x})} \exp\left[-\sum_{i<j} e_{ij}\phi_2(\mathrm{r}) \right] \exp\left[-\sum_{i<j<k} \left(\frac{e_{ij} + e_{ik} + e_{jk}}{3} \right) \phi_3(\mathrm{r}_1, \mathrm{r}_2, \mathrm{r}_3) \right] \qquad [107]$$

where in the case of the 3rd-order interactions, the edge correction factor is the arithmetic average of the correction factors for the three pairwise interactions formed from any given triple of points. The computations used here for the edge correction weights w_{ij} for rectangular plots follow Haase (1995):

1. Calculate from the point x_i the distance to the nearest edge e_1, the distance to the second nearest edge e_2, and the distance to point x_j, $r = \left\| x_i - x_j \right\|$

2. If $r \leq e_1$ then no edge correction is needed and $w_{ij} = 1$.

3. Otherwise if $r > e_1$ and $r \leq e_2$ (Case A of Haase), set

$$w_{ij} = 1 - \cos^{-1}\left(e_1 / r\right) / \pi$$

4. Otherwise calculate the distance to the nearest corner, c. If $r \geq c$, $r > e_1$, and $r > e_2$ (Case B of Haase), set.

$$w_{ij} = 1 - \left[\cos^{-1}\left(e_1 / r\right) + \cos^{-1}\left(e_2 / r\right) + \pi/2\right] / 2\pi$$

5. Otherwise $r < c$, $r > e_1$, and $r > e_2$ (Case C of Haase). Set

$$w_{ij} = 1 - \left[2\cos^{-1}\left(e_1 / r\right) + 2\cos^{-1}\left(e_2 / r\right)\right] / 2\pi$$

3.5 Stability of Gibbs point process models

As described in §2.21, the interaction potentials of a Gibbs point process must follow certain requirements so that the normalizing constant of the process is finite and the density of the process is well defined. These requirements are known as stability conditions and can be expressed in terms of the Gibbs point process density. Specifically, denote the Gibbs point process Φ that maps from the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to the measurable space $(\mathbb{N}, \mathcal{N})$ as described in §2.21. Let $f(\mathbf{x}) = \alpha g(\mathbf{x})$ denote the probability density function of the process, with normalizing constant α and finite point configuration $\mathbf{x} = \{x_1, x_2, ..., x_n\}$. Stability conditions for f are as follows (Kendall and Møller 2000; Møller and Waagepetersen 2003):

1) Ruelle stability

$f(\mathbf{x}) \leq \alpha K^{n(\mathbf{x})}$ for positive constant K and all point configurations $\mathbf{x} \in \mathbb{N}$

2) Local stability

$f(\mathbf{x}) \geq \frac{f(\mathbf{x} \cup \{u\})}{K}$ for positive constant K, all point configurations $\mathbf{x} \in \mathbb{N}$, and any point $\{u\} \in F \subseteq \Omega$ with $\{u\} \neq \{x_i\}$

Ruelle stability implies that $g(\mathbf{x}) \leq K^{n(\mathbf{x})}$, which occurs when the potentials of the Gibbs process are purely repulsive. To illustrate, consider a pairwise interaction process where

$$g(\mathbf{x}) = \prod_{i=1}^{n(\mathbf{x})} b(x_i) \prod_{i<j} h(x_i, x_j)$$
$$= \beta^{n(\mathbf{x})} \exp\left[\sum_{i<j} -\phi_2 \left(\left\| x_i - x_j \right\| \right) \right]$$

When the process is purely repulsive, $\phi_2 \left(\left\| x_i - x_j \right\| \right) \geq 0$ for all pairs (x_i, x_j), and $\sum_{i<j} -\phi_2 \left(\left\| x_i - x_j \right\| \right) \leq 0$ so that $g(\mathbf{x}) \leq \beta^{n(\mathbf{x})} \exp(0) = \beta^{n(\mathbf{x})}$, and hence the process is Ruelle stable with $K = \beta$. In addition to the purely repulsive case, Ruelle (1969, p. 37) proved that any Gibbs process with a pair potential incorporating a non-zero hard-core parameter will be stable.

Local stability is considered a "stronger" requirement on the density in that local stability implies Ruelle stability but not vice versa (Kendall and Møller 2000). Local stability can be expressed in terms of the conditional intensity function of the Gibbs process through

$$\lambda^*(u; \mathbf{x}) = \frac{f(\mathbf{x} \cup \{u\})}{f(\mathbf{x})} \leq K$$

implying that local stability is satisfied when the conditional intensity function is uniformly bounded.

These stability conditions have quite different implications for Gibbs point processes. Ruelle stability is a necessary condition for the Gibbs density to exist, meaning that $g(\mathbf{x})$ is integrable with respect to a Poisson process with intensity measure Λ. Ruelle

(1969) describes potentials that violate this requirement as "catastrophic". Some examples of catastrophic potentials given by Ruelle include purely attractive pair potentials and pair potentials that are finitely repulsive at short distances and attractive at larger distances. Perhaps the best known example of a catastrophic potential was the model proposed by Strauss (1975), which specified a random number of points and the pair potential $\phi_2\left(\left\|x_i - x_j\right\|\right) = -\theta_1 I_{[0,R]}\left(\left\|x_i - x_j\right\|\right)$, where R is an interaction distance. Strauss then allowed the possibility that $\theta_1 > 0$ so that the potential could be purely attractive. It was unfortunate that Strauss entitled his paper "A model for clustering" since the purely attractive pair potential is catastrophic and was proven by Kelly and Ripley (1976) and Gates and Westcott (1986) to yield an un-normalizable density that is not at all a valid model for clustering.

Local stability is often required of point process models for it proves the positivity condition of Gibbs point processes (Baddeley 2001) and it allows for straightforward proof of geometric ergodicity of Markov Chain Monte Carlo (MCMC) algorithms that simulate the process (Geyer 1999). As described in §2.4, ergodicity is necessary for a Markov chain to converge to the desired stationary distribution. Geometric ergodicity implies that convergence of the Markov chain will be at a geometrically fast rate. The ability to prove geometric ergodicity of MCMC samplers is of obvious theoretical benefit when simulating a Gibbs point process density. The condition of local stability has particularly been exploited to create coupling constructions for "perfect simulation" of Gibbs point processes (Häggström et al. 1999; Kendall and Møller 2000; Berthelsen and Møller 2002; Møller and Waagepetersen 2004; Van Lieshout and Stoica 2006).

Since it is a objective of this research to model local clustering in tree stems that result from stump sprouting, it is essential that the proposed Gibbs point processes can be attractive but still satisfy stability conditions. All of the proposed pairwise interaction models (F, HCME, and HCSP) incorporate a non-zero hard-core parameter to ensure that the Gibbs densities using these pair potentials are Ruelle stable (Ruelle 1969, p. 37). The incorporation of a hard-core parameter allows these models to be locally stable as well. As an example consider the Fiksel model that is attractive when $\theta_1 < 0$ and repulsive when $\theta_1 > 0$. From Eq. 46 and incorporating the edge correction factor e_{ui}

$$\lambda^*(u;\mathbf{x}) = b(u)\prod_{\substack{i=1 \\ x_i \neq u}}^{n(\mathbf{x})} h(u,x_i)$$
$$= \beta \prod_{\substack{i=1 \\ x_i \neq u}}^{n(\mathbf{x})} \exp\left(-e_{ui}\phi\left(\|u-x_i\|\right)\right)$$

Now if any x_i are within r_0 distance of u (a violation of the hard-core distance), the potential ϕ is infinite and the conditional intensity will have an upper bound at 0. Assume that this is not the case, so

$$\lambda^*(u;\mathbf{x}) = \beta \prod_{\substack{i=1 \\ x_i \neq u}}^{n(\mathbf{x})} \exp\left[-e_{ui}\theta_1 \exp\left(-\theta_2 \|u-x_i\|\right) I_{[r_0,R]}\left(\|u-x_i\|\right)\right]$$

Since taking the natural logarithm of the conditional intensity is a monotonic transformation, the upper bound can be found through

$$\ln\left[\lambda^*(u;\mathbf{x})\right]=\ln(\beta)+\ln\left[\prod_{\substack{i=1\\x_i\neq u}}^{n(\mathbf{x})}\exp\left[-e_{ui}\theta_1\exp\left(-\theta_2\|u-x_i\|\right)I_{[r_0,R]}\left(\|u-x_i\|\right)\right]\right] \quad [108]$$

$$=\ln(\beta)+\sum_{\substack{i=1\\x_i\neq u}}^{n(\mathbf{x})}-e_{ui}\theta_1\exp\left(-\theta_2\|u-x_i\|\right)I_{[r_0,R]}\left(\|u-x_i\|\right)$$

Notice that if any point's distance from u exceeded the interaction distance R, then its contribution to the sum $\sum_{\substack{i=1\\x_i\neq u}}^{n(\mathbf{x})}-e_{ui}\theta_1\exp\left(-\theta_2\|u-x_i\|\right)I_{[r_0,R]}\left(\|u-x_i\|\right)$ must be zero, so Eq. 108 can be simplified further to

$$\ln\left[\lambda^*(u;\mathbf{x})\right]=\ln(\beta)+\sum_{\substack{i=1\\x_i\neq u}}^{n_R(\mathbf{x})}-e_{ui}\theta_1\exp\left(-\theta_2\|u-x_i\|\right)$$

where the sum is being taken only over the $n_R(\mathbf{x})$ points in $\mathbf{x}$ that interact with u. In the purely repulsive case ($\theta_1 > 0$), the sum over the $n_R(\mathbf{x})$ points is negative, so the log-conditional intensity has an upper bound at $\ln(\beta)$. In the purely attractive case of $\theta_1 < 0$, it is observed that since $\exp\left(-\theta_2 r_0\right) \geq \exp\left(-\theta_2\|u-x_i\|\right)$ for all x_i

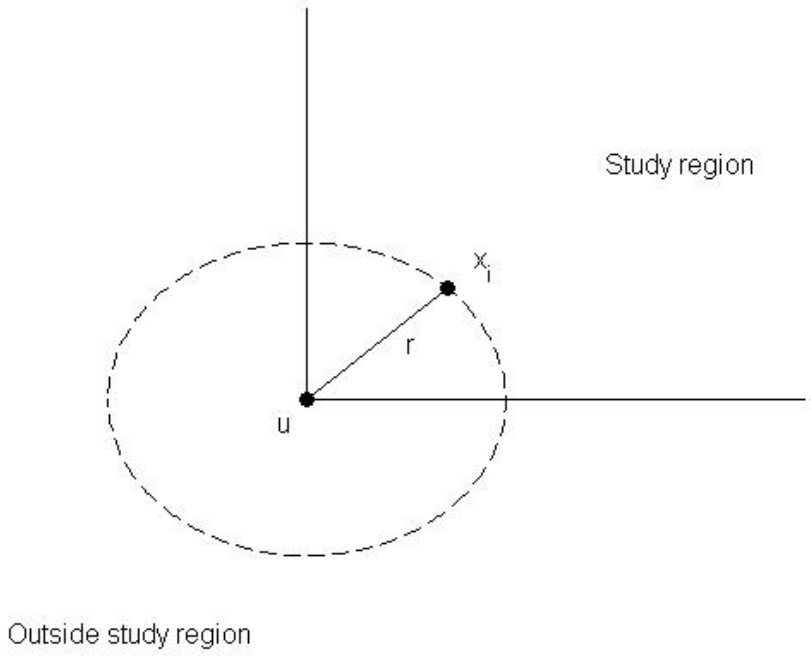

Figure 10. Geometry of maximal edge correction weight w_{ui}^{-1}. Note: $r = \|u - x_i\|$

$$\begin{aligned}\ln\left[\lambda^*(u;\mathbf{x})\right] &\le \ln(\beta) + \sum_{i=1}^{n_R(\mathbf{x})} -e_{ui}\theta_1\exp(-\theta_2 r_0) \\ &= \ln(\beta) - \theta_1\exp(-\theta_2 r_0)\sum_{i=1}^{n_R(\mathbf{x})} e_{ui} \\ &\le \ln(\beta) - \theta_1\exp(-\theta_2 r_0)4n_R(\mathbf{x})\end{aligned}$$

where the last inequality follows from the fact that the minimum proportion of the circumference of a circle centered at *u* that lies within the study region is ¼, which occurs when *u* is located directly on a corner (Fig. 10). In other words, the bounds on the edge correction weights $1/4 \le w_{ui} \le 1$ imply $1 \le w_{ui}^{-1} \le 4$ and $1 \le e_{ui} \le 4$, so that

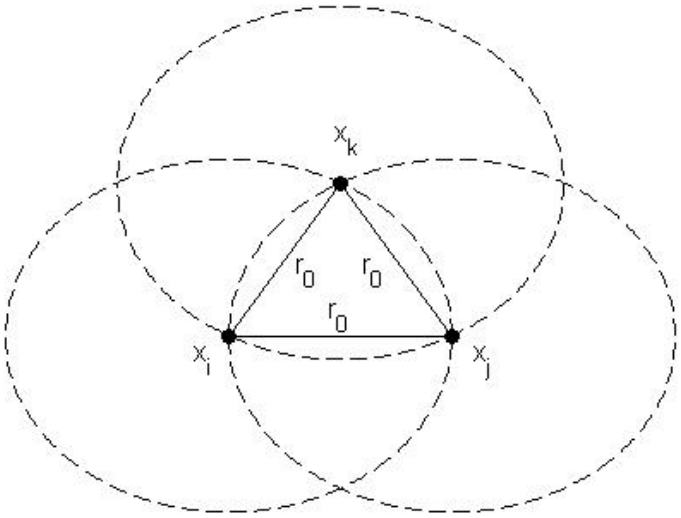

Figure 11. Geometry of three points (x_i, x_j, x_k) mutually r_0 distance apart.

$$\sum_{i=1}^{n_R(\mathbf{x})} e_{ui} \le \sum_{i=1}^{n_R(\mathbf{x})} 4 = 4n_R(\mathbf{x})$$

Thus the bound on the log-conditional intensity increases linearly with the number of points in **x** that interact with the point u If the hard-core distance was zero, then a random number of points $n(\mathbf{x})$ would cause $n_R(\mathbf{x})$ and the log-conditional intensity to increase without limit to infinity. With a non-zero hard-core distance, there is an upper bound on $n_R(\mathbf{x})$, as all of these points must be mutually r_0 distance apart from each other and u, but collectively are within R distance of u. For every pair of points x_i, x_j separated by r_0 distance, the closest that a third point x_k can be to the pair and be r_0 distance from both x_i and x_j will occur when all three points form an equilateral triangle (Fig. 11). This implies

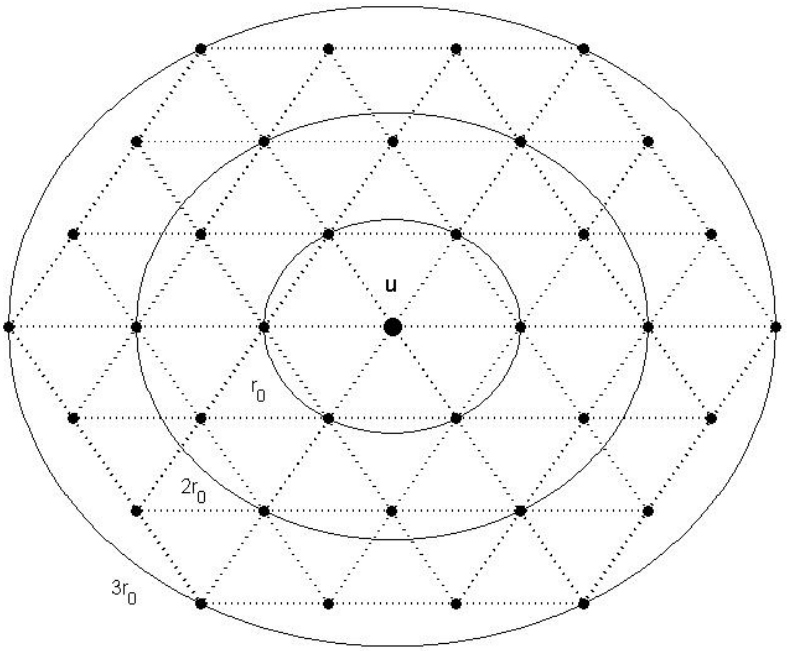

Figure 12. Geometry of the maximum number of points mutually r_0 distance apart from each other and the point u at increasing distances from u.

that for any given point u, there can be no more than six points that are all mutually r_0 distance apart, as these points will form a regular hexagon that is inscribed within the circle centered at u with radius r_0. With increasing distance from u, successively larger hexagons will prescribe the upper bound on the number of points that are mutually r_0 distance apart from each other and u, but collectively are within R distance of u (Fig. 12). So for any given r_0 and R,

$$n_R(\mathbf{x}) \le 6\left(\frac{R}{r_0}\right)$$

and the upper bound on the log-conditional intensity is achieved as

$$\ln\left[\lambda^*(u;\mathbf{x})\right] \le \ln(\beta) - \theta_1 \exp(-\theta_2 r_0) 24\left(\frac{R}{r_0}\right) \qquad [109]$$

Notice that for the foregoing argument to hold, the attractive pair potential must have both a non-zero hard core distance and a finite range of interaction, which rules out the attractive Strauss process (hard-core distance = 0) and infinite range potentials like the soft core and very soft core models(Ogata and Tanemura 1984) defined throughout $\mathbb{R}^2$.

The proof of local stability for the HCME and HCSP models follows closely the logic of the foregoing example for the Fiksel model. In the HCME and HCSP cases, the distance that will maximize $-\phi(\|u - x_i\|)$ will depend upon the values of the parameters of the potential. When these models are purely attractive, as is the case of the Fiksel model when $\theta_1 < 0$, this maximum will occur at $\|u - x_i\| = r_0$ and the foregoing arguments will be nearly identical. This will also be true when the models are attractive at short distances and repulsive at larger distances, such as the HCME model (Fig. 5) and the HCSP model (Fig. 6). When this maximum occurs at some distance in between the hard-core and the interaction distance, as is shown for the HCSP model (Fig. 8), then the bound on the log-conditional intensity will occur when a maximum number of points that can interact with *u* are all given this maximal value of $-\phi(\|u - x_i\|)$. Since $\phi(\|u - x_i\|) < -\infty$ for both the HCME and HCSP potential models with finite parameter

values, this sum will have an upper limit and the log-conditional intensity will consequently have an upper bound.

For the triplets models proposed here, it is sufficient to demonstrate their local stability only, since local stability implies Ruelle stability (Geyer 1999). To prove local stability, an upper bound is needed for the conditional intensity:

$$\lambda^*(u;\mathbf{x}) = b(u)\prod_{\substack{i=1\\ x_i\neq u}}^{n(\mathbf{x})} h(u,x_i)\prod_{\substack{i<j\\ x_i,x_j\neq u}} h(u,x_i,x_j)$$

$$=\beta\prod_{\substack{i=1\\ x_i\neq u}}^{n(\mathbf{x})} \exp\left(-e_{ui}\phi_2\left(\|u-x_i\|\right)\right)\prod_{\substack{i<j\\ x_i,x_j\neq u}} \exp\left(-e_{uij}\phi_3\left(\|u-x_i\|,\|u-x_j\|,\|x_i-x_j\|\right)\right)$$

where the edge correction term for the 3rd-order interaction is $e_{uij} = \left(\frac{e_{ui}+e_{uj}+e_{ij}}{3}\right)$. As before, using the natural logarithm of the conditional intensity will not change the upper bound, so this bound can be found through

$$\ln\left[\lambda^*(u;\mathbf{x})\right] = \ln\left(b(u)\right) + \sum_{\substack{i=1\\ x_i\neq u}}^{n(\mathbf{x})} -e_{ui}\phi_2\left(\|u-x_i\|\right) + \sum_{\substack{i<j\\ x_i,x_j\neq u}} -e_{uij}\phi_3\left(\|u-x_i\|,\|u-x_j\|,\|x_i-x_j\|\right)$$

To illustrate the local stability of the proposed triplets models, consider the exponential triplets process with $b(u) = \beta$, 2nd-order potential

$$\phi_2\left(r_{ij}\right)=\begin{cases}\infty, \text{ for } r_{ij}\le r_0\\ \theta_1\exp\left(-r_{ij}\right), \text{ for } r_0<r_{ij}\le R\\ 0, \text{ for } r_{ij}>R\end{cases},$$

and 3rd-order potential

$$\phi_3\left(r_{ij},r_{ik},r_{jk}\right)=\begin{cases}\theta_2\exp\left(-\dfrac{\left(r_{ij}+r_{ik}+r_{jk}\right)}{3}\right) \text{ for } r_{ij},r_{ik},r_{jk}\le R\\ 0 \text{ for any } r_{ij},r_{ik},r_{jk}>R\end{cases}, \theta_2>0$$

In the EXPTP and all triplets models, the 3rd-order potential is constructed to be purely repulsive, that is $\phi_3\left(r_{ij},r_{ik},r_{jk}\right)\ge 0$ for all triples of points. If the 2nd-order potential is also repulsive, which occurs when $\theta_1 > 0$, then

$$\sum_{\substack{i=1\\ x_i\ne u}}^{n(\mathbf{x})}-e_{ui}\phi_2\left(\left\|u-x_i\right\|\right)+\sum_{\substack{i<j\\ x_i,x_j\ne u}}-e_{uij}\phi_3\left(\left\|u-x_i\right\|,\left\|u-x_j\right\|,\left\|x_i-x_j\right\|\right)\le 0$$

and $\ln\left[\lambda^*\left(u;\mathbf{x}\right)\right]\le\ln\left(\beta\right)$. Hence it is only the case when the 2nd-order potential is partially or totally attractive will the log-conditional intensity be greater than ln(β), which is now assumed.

As in the case of Gibbs processes with only pairwise interactions, the violation of the hard-core distance between u and any point x_i would cause the second-order potential to be infinite for that pairwise distance. Consequently the sum

$$\sum_{\substack{i=1 \\ x_i \neq u}}^{n(\mathbf{x})} -e_{ui}\phi_2\left(\|u - x_i\|\right)$$

must be negative infinity so that $\ln\left[\lambda^*\left(u;\mathbf{x}\right)\right] = -\infty$ and the conditional intensity has an upper bound at zero. Let's assume that this is not the case, so for the EXPTP model

$$\ln\left[\lambda^*\left(u;\mathbf{x}\right)\right] = \ln\left(\beta\right) + \sum_{\substack{i=1 \\ x_i \neq u}}^{n(\mathbf{x})} -e_{ui}\theta_1 \exp\left(-\|u - x_i\|\right) I_{[r_0,R]}\left(\|u - x_i\|\right) + \sum_{\substack{i<j \\ x_i, x_j \neq u}} -e_{uij}\theta_2 \exp\left(-\frac{\left(\|u - x_i\| + \|u - x_j\| + \|x_i - x_j\|\right)}{3}\right) I^*\left(u, x_i, x_j\right)$$

where

$$I^*\left(u, x_i, x_j\right) = I_{[0,R]}\left(\|u - x_i\|\right) I_{[0,R]}\left(\|u - x_j\|\right) I_{[0,R]}\left(\|x_i - x_j\|\right)$$

As in the case of Gibbs processes with only pairwise interactions, any pairwise distance that exceeds the interaction distance R will cause the pair potential to be zero. In a triplets process, if any of the three pairwise distances that forms a triple exceeds the

interaction distance R, then the third-order potential for that triple is zero. Therefore the log-conditional intensity function for the EXPTP process can be simplified to

$$\ln\left[\lambda^*(u;\mathbf{x})\right]=\ln(\beta)+\sum_{\substack{i=1\\ x_i\neq u}}^{n_R(\mathbf{x})}-e_{ui}\theta_1\exp\left(-\|u-x_i\|\right)$$
$$+\sum_{\substack{i<j^*\\ x_i,x_j\neq u}}-e_{uij}\theta_2\exp\left(-\frac{\left(\|u-x_i\|+\|u-x_j\|+\|x_i-x_j\|\right)}{3}\right)$$

where the sum of the pair potentials is being taken only over the $n_R(\mathbf{x})$ points in $\mathbf{x}$ that interact with u, and the double sum of the triple potentials is being taken over only those pairs of points in $\mathbf{x}$ that are within the interaction distance of each other and the point u. In the purely attractive case of $\theta_1 < 0$, it is observed that $\exp(-r_0)\geq\exp\left(-\|u-x_i\|\right)$ for all x_i implies $-\theta_1\exp(-r_0)\geq-\theta_1\exp\left(-\|u-x_i\|\right)$ and since

$$\exp\left(-\frac{(R+R+R)}{3}\right)\leq\exp\left(-\frac{\left(\|u-x_i\|+\|u-x_j\|+\|x_i-x_j\|\right)}{3}\right)$$

for all x_i and x_j implies

$$-\theta_2\exp\left(-\frac{(R+R+R)}{3}\right)\geq-\theta_2\exp\left(-\frac{\left(\|u-x_i\|+\|u-x_j\|+\|x_i-x_j\|\right)}{3}\right)$$

for $\theta_2 > 0$, then the log-conditional intensity satisfies

$$\ln\left[\lambda^*(u;\mathbf{x})\right] \le \ln(\beta) + \sum_{\substack{i=1 \\ x_i \neq u}}^{n_R(\mathbf{x})} -e_{ui}\theta_1 \exp(-r_0) + \sum_{\substack{i<j^* \\ x_i, x_j \neq u}} -e_{uij}\theta_2 \exp(-R)$$

$$= \ln(\beta) - \theta_1 \exp(-r_0) \sum_{\substack{i=1 \\ x_i \neq u}}^{n_R(\mathbf{x})} e_{ui} - \theta_2 \exp(-R) \sum_{\substack{i<j^* \\ x_i, x_j \neq u}} e_{uij}$$

$$\le \ln(\beta) - \theta_1 \exp(-r_0) 4 n_R(\mathbf{x}) - \theta_2 \exp(-R) 4 \left(\frac{n_R(\mathbf{x})\left(n_R(\mathbf{x}) - 1\right)}{2} \right)$$

$$= \ln(\beta) - \theta_1 \exp(-r_0) 4 n_R(\mathbf{x}) - \theta_2 \exp(-R) 2 \left(n_R(\mathbf{x})^2 - n_R(\mathbf{x}) \right)$$

$$= -2\theta_2 \exp(-R) n_R(\mathbf{x})^2 + \left[2\theta_2 \exp(-R) - 4\theta_1 \exp(-r_0) \right] n_R(\mathbf{x}) + \ln(\beta)$$

Now let $a = -2\theta_2 \exp(-R)$, $b = \left[2\theta_2 \exp(-R) - 4\theta_1 \exp(-r_0) \right]$, and $c = \ln(\beta)$ so that

$$\ln\left[\lambda^*(u;\mathbf{x})\right] \le a n_R(\mathbf{x})^2 + b n_R(\mathbf{x}) + c \qquad [110]$$

Since $\theta_2 > 0$ and $a < 0$, the upper bound of the log-conditional intensity function is a negative definite quadratic form. It is then a simple exercise in calculus to show that this quadratic form has a maximum at $-\frac{b^2}{4a} + c$ at the critical value $n_R(\mathbf{x})^* = \frac{-b}{2a}$.

The idea behind the proof that the EXPTP process is locally stable is not novel. In presenting the "triplets process" (here the GTP model), Geyer (1999) laid out a proof of

local stability that is essentially the same. Gibbs processes with repulsive 3rd-order interactions will be locally stable since the number of triples involving the point u will eventually overwhelm the number of pairs involving u, no matter how attractive the pair potential is, so long as the pair potential has a finite lower bound and both the pair and 3rd-order potentials have a finite range of interaction. Additionally, the proofs that the triplets and pairwise Gibbs models presented here are locally stable do not depend upon the use here of isotropic edge correction factors, which only act to increase the upper bound of the log-conditional intensity by a multiplicative factor.

3.6 Gibbs Point Process Simulation

To simulate Gibbs point processes with given parameter values, a Metropolis-Hastings algorithm (Geyer and Møller 1994; Geyer 1999) was used that updates configurations from a point process in one of three ways: 1) by proposing to displace a point location, 2) by adding one new point to the configuration, or 3) by removing a point from the configuration. Although these updates are called shift, birth, and death proposals respectively, the Metropolis-Hastings algorithm of Geyer and Møller (1994) should not be confused with spatial birth and death processes (Preston 1977) which can simulate Gibbs point processes using continuous time Markov chains. The shift-birth-death Metropolis-Hastings algorithm utilizes discrete time Markov chains defined on the state space of the set of finite point configurations for which the Gibbs point process density f is positive, i.e. $\left\{\mathbf{x} \in \mathbb{N}: f\left(\mathbf{x}\right) > 0\right\}$ as in Sec 2.12. Given some initial configuration of n points, $X_0 = \mathbf{x}^{(0)}$, the shift-birth-death Metropolis-Hastings algorithm updates the configuration to a new state of the Markov chain, X_1 through the following procedure:

1) Define the following probabilities ($0 \leq p \leq 1$):

Probability of a shift of a point : p_s

Probability of a birth (addition of a point to the configuration) : p_b

Probability of a death (deletion of a point to the configuration) : $p_d = 1 - p_b$

2) Define the following Hastings ratios:

$$r_s\left(\mathbf{x}_s^*, \mathbf{x}^{(t)}\right) = \frac{f\left(\mathbf{x}_s^*\right) q_s\left(\mathbf{x}^{(t)}\right)}{f\left(\mathbf{x}^{(t)}\right) q_s\left(\mathbf{x}_s^*\right)} \qquad [111]$$

$$r_b\left(\mathbf{x}_b^*, \mathbf{x}^{(t)}\right) = \frac{f\left(\mathbf{x}_b^*\right) p_d \; q_d\left(\mathbf{x}_b^*\right)}{f\left(\mathbf{x}^{(t)}\right) p_b \; q_b\left(\mathbf{x}^{(t)}\right)} \qquad [112]$$

$$r_d\left(\mathbf{x}_d^*, \mathbf{x}^{(t)}\right) = \frac{f\left(\mathbf{x}_d^*\right) p_b \; q_b\left(\mathbf{x}_d^*\right)}{f\left(\mathbf{x}^{(t)}\right) p_d \; q_d\left(\mathbf{x}^{(t)}\right)} \qquad [113]$$

where

$f(\cdot)$ is the Gibbs point process density,

$q_s(\cdot)$ is a proposal density of point-shifts,

$q_b(\cdot)$ is a proposal density of point-births, and

$q_d(\cdot)$ is a proposal density of point-deaths

3) Initialize $t = 0$ and $X_0 = \mathbf{x}^{(0)}$

4) Generate a random number, U_s, distributed as Uniform (0,1)

5) If $U_s \leq p_s$ then do the following:

5a) Generate a random number, I, distributed as Uniform (1, n)

5b) Generate a random number, R^s, distributed as Uniform (0,1)

5c) Given $I = i$, draw a point $\xi \in F$ and propose

$$\mathbf{x}_s^* = \left(x_1, x_2, \ldots, x_{i-1}, \xi, x_{i+1}, \ldots, x_n\right)$$

5d) If $R^s \leq r_s\left(x_s^*, x^{(t)}\right)$, set $X_{t+1} = \mathbf{x}_s^*$, otherwise set $X_{t+1} = X_t = \mathbf{x}^{(t)}$

6) Otherwise, if $U_s > p_s$ then do the following:

6a) Generate a random number, U_b, distributed as Uniform (0,1)

6b) Generate a random number, R^b, distributed as Uniform (0,1)

6c) If $U_b \leq p_b$ then do the following:

i) Draw a point $\xi \in F$ and propose $\mathbf{x}_b^* = \mathbf{x}^{(t)} \cup \xi$

ii) If $R^b \leq r_b\left(x_b^*, x^{(t)}\right)$ set $X_{t+1} = \mathbf{x}_b^*$, otherwise set $X_{t+1} = X_t = \mathbf{x}^{(t)}$

6d) Otherwise if $U_b > p_b$ then do the following:

i) If $\mathbf{x} = \emptyset$ then set $X_{t+1} = X_t = \mathbf{x}^{(t)}$

ii) If $\mathbf{x} \neq \emptyset$, then select ξ uniformly from $\mathbf{x}^{(t)}$ and propose $\mathbf{x}_d^* = \mathbf{x}^{(t)} \setminus \xi$. If $R^d \leq r_d\left(x_d^*, x^{(t)}\right)$, set $X_{t+1} = \mathbf{x}_d^*$, otherwise set $X_{t+1} = X_t = \mathbf{x}^{(t)}$

7) Update t, $\mathbf{x}$ and repeat steps 4 – 6 until $t = T$, the chosen number of total iterations for the Markov chain.

To summarize, the algorithm first determines if a shift in the location of a point should occur. If yes, then the shift proceeds if a randomly generated number is less than the Hastings ratio for a shift, otherwise the configuration remains the same. If a point shift did not occur, then the algorithm determines if a birth (an addition of a point to the configuration) should occur. If yes, then the birth proceeds if a randomly generated number is less than the Hastings ratio for a birth; otherwise the configuration remains the same. If neither a point shift nor a birth occurs, then a death (a removal of a point from the configuration) will occur if a randomly generated number is less than the Hastings ratio for a death; otherwise the configuration remains the same. The key quantities to specify in the algorithm are the probabilities for point shifts, births, and deaths, along with the proposal densities for point shifts, births and deaths. Here the choice for the probabilities for point-shifts, births, and deaths were chosen so that a shift, birth, or death was equally likely, i.e. $p_s = 1/3$ and conditional upon no point shift, $p_b = p_d = 1/2$ (Møller

and Waagepetersen 2004). Following the suggestion of Geyer and Møller (1994) and Geyer (1999), the proposal densities used here are as follows:

$$q_s(\mathbf{x}) = \frac{1}{n(\mathbf{x})}\frac{1}{|F|} \quad [114]$$

$$q_b(\mathbf{x}) = \frac{1}{|F|} \quad [115]$$

$$q_d(\mathbf{x}) = \frac{1}{n(\mathbf{x})} \quad [116]$$

where $n(\mathbf{x})$ is the random number of points in the configuration $\mathbf{x}$. Certainly other proposal densities can be incorporated with the shift-birth-death Metropolis-Hastings algorithm. Allen and Tildesley (1987, p. 118-122), for example, describe a procedure where a point is shifted only within a local neighborhood of its pre-existing location. As pointed out by Van Lieshout (2000, p. 85), such a neighborhood would have to be constructed so that the Markov chain would be irreducible and aperiodic; choosing to shift a point anywhere in F assures these properties.

Given the proposal for point shifts and using the fact that throughout the algorithm the parameter vector for the Gibbs density is fixed, the Hastings ratios for a point shift may be simplified to the following:

$$r_s\left(\mathbf{x}_s^*, \mathbf{x}^{(t)}\right) = \frac{\frac{g_\theta\left(\mathbf{x}_s^*\right)}{z(\theta)}\frac{1}{n\left(\mathbf{x}^{(t)}\right)}\frac{1}{|F|}}{\frac{g_\theta\left(\mathbf{x}^{(t)}\right)}{z(\theta)}\frac{1}{n\left(\mathbf{x}_s^*\right)}\frac{1}{|F|}} = \frac{g_\theta\left(\mathbf{x}_s^*\right)}{g_\theta\left(\mathbf{x}^{(t)}\right)}$$

The Hastings ratio only depends upon the ratio of un-normalized Gibbs densities, since the normalizing constants cancel and a displacement of a point does not change the total number of points in the configuration. Similarly the Hastings ratios for births and deaths can also be simplified to the following:

$$r_b\left(\mathbf{x}_b^*, \mathbf{x}^{(t)}\right) = \frac{\frac{g_\theta\left(\mathbf{x}_b^*\right)}{z(\theta)} p_d \frac{1}{n\left(\mathbf{x}_b^*\right)}}{\frac{g_\theta\left(\mathbf{x}^{(t)}\right)}{z(\theta)} p_b \frac{1}{|F|}} \qquad [117]$$

$$= \frac{g_\theta\left(\mathbf{x}_b^*\right) p_d |F|}{g_\theta\left(\mathbf{x}^{(t)}\right) p_b \; n\left(\mathbf{x}_b^*\right)} = \frac{g_\theta(\mathbf{x} \cup \xi)}{g_\theta(\mathbf{x})} \frac{p_d |F|}{p_b \; (n(\mathbf{x}) + 1)}$$

$$r_d\left(\mathbf{x}_d^*, \mathbf{x}^{(t)}\right) = \frac{\frac{g_\theta\left(\mathbf{x}_d^*\right)}{z(\theta)} p_b \frac{1}{|F|}}{\frac{g_\theta\left(\mathbf{x}^{(t)}\right)}{z(\theta)} p_d \frac{1}{n\left(\mathbf{x}^{(t)}\right)}} \qquad [118]$$

$$= \frac{g_\theta\left(\mathbf{x}_d^*\right) p_b \; n\left(\mathbf{x}^{(t)}\right)}{g_\theta\left(\mathbf{x}^{(t)}\right) p_d \; |F|} = \frac{g_\theta(\mathbf{x} \setminus \xi)}{g_\theta(\mathbf{x})} \frac{p_b \; n(\mathbf{x})}{p_d \; |F|}$$

where $\mathbf{x} = \mathbf{x}^{(t)}$. Notice the presence of the important conditional intensity function (see Eqs. 30, 45) in the Hastings ratio for births. As pointed out in Sec. 3.5, the local stability of point process models implies that this conditional intensity function is bounded. When local stability holds, the shift-birth-death Metropolis-Hastings algorithm can be proven to be *v*-uniformly ergodic (Møller and Waagepetersen 2004, p. 132); additionally Geyer

(1999) proved that the algorithm without the shift proposal is geometrically ergodic for locally stable point processes.

There are two features to this algorithm that are important to note. One is that it is an example of a reversible jump Metropolis-Hastings algorithm that can propose moves across subspaces Ω_k with different dimension (Green 1995). Green's requirement for detailed balance to hold when proposing moves across subspaces of different dimension (e.g. $\Omega = \bigcup_{k \in \mathcal{K}} \Omega_k$) is the existence of a symmetric measure ξ_m on $\Omega \times \Omega$ so that $\pi(dx) Q_m(x, dy)$ has a density $f_m(x, dy)$ with respect to ξ_m, where $\pi(dx)$ is the target probability distribution and $Q_m(x, dy)$ is the proposal distribution for transitions of type m for all $x, y \in \Omega$ The standard metropolis-Hastings acceptance probability (Eq. 85) is replaced with the acceptance probability.

$$\alpha_m(x, y) = \min\left\{1, \frac{f_m(y, x)}{f_m(x, y)}\right\} \qquad [119]$$

Geyer (1999) demonstrated that for a spatial point process with density f (and un-normalized density g) with respect to the finite Poisson process with intensity measure $\Lambda(dx)$, the symmetric measure ξ_m may be taken to be Λ^{m+1} and the Metropolis-Hastings-Green ratio is the following:

$$\frac{f_m(y,x)}{f_m(x,y)} = \frac{\frac{\pi(dy)Q_m(y,dx)}{\xi_m(dy,dx)}}{\frac{\pi(dx)Q_m(x,dy)}{\xi_m(dx,dy)}}$$

$$= \frac{\pi(dy)Q_m(y,dx)}{\pi(dx)Q_m(x,dy)}$$

$$= \frac{\frac{g(y)}{(m+1)!}}{\frac{g(x)}{m!\Lambda(F)}}$$

$$= \frac{g(y)}{g(x)}\frac{\Lambda(F)}{(m+1)}$$

Since the Gibbs processes used here are constructed with respect to a Poisson process with unit intensity, e.g. $\Lambda(F) = \lambda\ \mu_L(F) = 1\ |F| = |F|$, then the Metropolis-Hastings ratio for a birth proposal (Eq. 117) is the Metropolis-Hastings-Green ratio above with $y = \mathbf{x} \cup \xi$, $x = \mathbf{x}$, $m = n(\mathbf{x})$, and $p_b = p_d = ½$. Similarly, the Metropolis-Hastings ratio for a death proposal (Eq. 118) can be seen to be the reciprocal of the Metropolis-Hastings-Green ratio above.

A second important feature is that the algorithm is constructed by mixing transition kernels together. The construction of a transition kernel P that is a linear combination of two kernels P_0 and P_1 which both satisfy detailed balance with respect to stationary distribution π implies that P will also satisfy detailed balance (Geyer and Møller 1994). Specifically the transition kernel of the shift-birth-death Metropolis-Hastings algorithm can be expressed as

$$P(A,\mathbf{x}) = p_s P_0(A,\mathbf{x}) + (1-p_s)P_1(A,\mathbf{x}) \qquad [120]$$

for all $x \in \mathbb{N}$, and $A \in \mathcal{N}$, where p_s is the probability of a point shift, P_0 is the Metropolis-Hastings transition kernel associated with point shifts and P_1 is the transition kernel that gives point births and deaths. The forms of the transition kernels P_0 and P_1 are somewhat mathematically involved (Geyer and Møller 1994, p. 361-362) and so are passed over here. Geyer and Møller note that the kernel P_0 is the same transition kernel as originally proposed by Metropolis et al. (1953), for which detailed balance has already been well established. The kernel P_1, which represents the transitions associated with point births and deaths, satisfies detailed balance through the existence of the symmetric measure ξ_m (Geyer 1999). Prior to the publication of reversible jump Metropolis-Hastings theory by Green (1995), Geyer and Møller (1994) showed that detailed balance is satisfied for P_1 by the condition:

$$\begin{aligned}\left[1 - p_b(\mathbf{x} \cup \xi)\right] q_d(\mathbf{x} \cup \xi)\, \alpha_d(\mathbf{x}, \mathbf{x} \cup \xi)\, f(\mathbf{x} \cup \xi) \\ = p_b(\mathbf{x})\, q_b(\mathbf{x})\, \alpha_b(\mathbf{x} \cup \xi, \mathbf{x})\, f(\mathbf{x})\end{aligned} \qquad [121]$$

where $\alpha_b(\ ,\ )$ and $\alpha_d(\ ,\ )$ are the acceptance probabilities for the birth and death updates, that is $\alpha_b(\mathbf{x} \cup \xi, \mathbf{x}) = \min\{1, r_b(\mathbf{x} \cup \xi, \mathbf{x})\}$, $\alpha_d(\mathbf{x}, \mathbf{x} \cup \xi) = \min\{1, r_d(\mathbf{x}, \mathbf{x} \cup \xi)\}$, and where the probability of birth is bounded between 0 and 1.

3.7 Bayesian Inference for Gibbs Point Processes

3.71 Inference with Metropolis-Hastings Algorithms

To perform Bayesian inference for Gibbs point processes, the Metropolis-Hastings algorithm is used to generate samples from the Gibbs posterior distribution as follows:

1. Given a fixed point configuration $\mathbf{x} \in \mathbb{N}$, let the posterior distribution μ on the space $\left(\Theta^k, \mathcal{F}, \mu\right)$ have density function

$$f\left(\theta \mid \mathbf{x}\right) = \frac{f\left(\theta\right) f\left(\mathbf{x} \mid \theta\right)}{\int f\left(\theta\right) f\left(\mathbf{x} \mid \theta\right) \mu\left(d\theta\right)} \qquad [122]$$

   with support $\theta \in \Theta^k \subseteq \mathbb{R}^k$, so that θ is a vector of real numbers of length *k*. Choose the initial state $\theta^{(0)} \in \Theta^k$ such that $f\left(\theta^{(0)} \mid \mathbf{x}\right) > 0$ and set $t = 0$.

2. Generate a sample θ^* from a desired proposal distribution *q* whose support includes Θ^k so that $X^* \sim q\left(\theta^* \mid \theta^{(t)}\right)$.

3. Assign the next state of the Markov chain according to the decision rule

$$X^{(t+1)} = \begin{Bmatrix} X^*, \text{ with probability } \alpha\left(\theta^{(t)}, \theta^*\right) \\ X^t, \text{ with probability } 1 - \alpha\left(\theta^{(t)}, \theta^*\right) \end{Bmatrix}$$

where the acceptance probability $\alpha\left(\theta^{(t)}, \theta^{*}\right)$ is defined as

$$\alpha\left(\theta^{(t)}, \theta^{*}\right) = \min\left\{1, \frac{f\left(\theta^{*} \mid \mathbf{x}\right)}{f\left(\theta^{(t)} \mid \mathbf{x}\right)} \frac{q\left(\theta^{(t)} \mid \theta *\right)}{q\left(\theta * \mid \theta^{(t)}\right)}\right\} \qquad [123]$$

4. Increment t and repeat until $t = T$

Using the definition of the posterior density, the Metropolis-Hastings ratio in Eq. 123 becomes

$$r\left(\theta^{*}, \theta^{(t)}\right) = \frac{f\left(\theta^{*} \mid \mathbf{x}\right)}{f\left(\theta^{(t)} \mid \mathbf{x}\right)} \frac{q\left(\theta^{(t)} \mid \theta *\right)}{q\left(\theta * \mid \theta^{(t)}\right)}$$

$$= \frac{\frac{f\left(\theta^{*}\right) f\left(\mathbf{x} \mid \theta^{*}\right)}{\int f(\theta) f(\mathbf{x} \mid \theta) \mu(d\theta)}}{\frac{f\left(\theta^{(t)}\right) f\left(\mathbf{x} \mid \theta^{(t)}\right)}{\int f(\theta) f(\mathbf{x} \mid \theta) \mu(d\theta)}} \frac{q\left(\theta^{(t)} \mid \theta *\right)}{q\left(\theta * \mid \theta^{(t)}\right)}$$

$$= \frac{f\left(\theta^{*}\right)}{f\left(\theta^{(t)}\right)} \frac{f\left(\mathbf{x} \mid \theta^{*}\right)}{f\left(\mathbf{x} \mid \theta^{(t)}\right)} \frac{q\left(\theta^{(t)} \mid \theta *\right)}{q\left(\theta * \mid \theta^{(t)}\right)}$$

where the ratio of normalizing constants of the posterior density cancels in the Metropolis-Hastings ratio. Expressing the likelihood of the Gibbs point process f in terms of its unnormalized density g and normalizing constant Z gives

$$r\left(\theta^{*}, \theta^{(t)}\right) = \frac{f\left(\theta^{*}\right)}{f\left(\theta^{(t)}\right)} \frac{\frac{g\left(\mathbf{x} \mid \theta^{*}\right)}{Z\left(\theta^{*}\right)}}{\frac{g\left(\mathbf{x} \mid \theta^{(t)}\right)}{Z\left(\theta^{(t)}\right)}} \frac{q\left(\theta^{(t)} \mid \theta *\right)}{q\left(\theta * \mid \theta^{(t)}\right)}$$

$$r\left(\theta^{*},\theta^{(t)}\right)=\frac{f\left(\theta^{*}\right)}{f\left(\theta^{(t)}\right)}\frac{g\left(\mathbf{x}\mid\theta^{*}\right)}{g\left(\mathbf{x}\mid\theta^{(t)}\right)}\frac{Z\left(\theta^{(t)}\right)}{Z\left(\theta^{*}\right)}\frac{q\left(\theta^{(t)}\mid\theta*\right)}{q\left(\theta*\mid\theta^{(t)}\right)} \qquad [124]$$

The Metropolis-Hastings ratio contains the ratio of intractable normalizing constants $Z\left(\theta^{(t)}\right)/Z\left(\theta^{*}\right)$ which must be estimated at *every* iteration of the Metropolis Hastings algorithm. The accurate and efficient estimation of this ratio is a critical issue for Bayesian inference of Gibbs point processes and two different methods will be used to address it. The first will use importance sampling to estimate the ratio by drawing samples from the Gibbs density using a value θ' that is "close" to both $\theta*$ and $\theta^{(t)}$ as discussed by Bognar (2004). The second approach, discussed in Sec. 3.73, uses an auxiliary variable whose proposal distribution is the target Gibbs density so that the normalizing constants in Eq. 124 can be cancelled (Møller et al. 2006).

Aside from the computation of the normalizing constant ratio, the application of the Metropolis-Hastings algorithm is straightforward. Following Møller et al. (2006), the proposal distribution q is the product of k independent normal distributions, that is

$$\begin{aligned} q\left(\theta*\mid\theta^{(t)}\right) &= \prod_{i=1}^{k} q_i\left(\theta_i^{*}\mid\theta_i^{(t)}\right) \\ &= \prod_{i=1}^{k} N_i\left(\theta_i^{*}\mid\mu_i=\theta_i^{(t)},\sigma_i^{2}\right) \end{aligned}$$

and

$$q\left(\theta^{(t)} \mid \theta*\right) = \prod_{i=1}^{k} q_i\left(\theta_i^{(t)} \mid \theta_i^*\right)$$
$$= \prod_{i=1}^{k} N_i\left(\theta_i^{(t)} \mid \mu_i = \theta_i^*, \sigma_i^2\right)$$

where μ_i is the mean and $\sigma^2{}_i$ is the variance of the i^{th} normal distribution. By construction, the means of $q_i\left(\theta_i^* \mid \theta_i^{(t)}\right)$ and $q_i\left(\theta_i^{(t)} \mid \theta_i^*\right)$ mirror each other and the variance $\sigma^2{}_i$ is a tuning parameter that will need to be adjusted so that the chain mixes well. The advantage of this proposal mechanism is that it is simple and easy to compute; in particular, the ratio of these proposal distributions will equal one, since for any $i = 1, 2, \ldots, k$,

$$\frac{q_i\left(\theta^{(t)} \mid \theta*\right)}{q_i\left(\theta* \mid \theta^{(t)}\right)} = \frac{N_i\left(\theta_i^{(t)} \mid \mu_i^{(t)} = \theta_i^*, \sigma_i^2\right)}{N_i\left(\theta_i^* \mid \mu_i^* = \theta_i^{(t)}, \sigma_i^2\right)}$$
$$= \frac{\frac{1}{\sqrt{2\pi\sigma_i}} \exp\left[-\frac{1}{2\sigma_i^2}\left(\theta_i^{(t)} - \mu_i^{(t)}\right)^2\right]}{\frac{1}{\sqrt{2\pi\sigma_i}} \exp\left[-\frac{1}{2\sigma_i^2}\left(\theta_i^* - \mu_i^*\right)^2\right]}$$

$$= \frac{\exp\left[-\frac{1}{2\sigma_i^2}\left(\theta_i^{(t)} - \theta_i^*\right)^2\right]}{\exp\left[-\frac{1}{2\sigma_i^2}\left(\theta_i^* - \theta_i^{(t)}\right)^2\right]}$$

$$= \frac{\exp\left[-\frac{1}{2\sigma_i^2}\left(\left(\theta_i^{(t)}\right)^2 - 2\theta_i^{(t)}\theta_i^* + \left(\theta_i^*\right)^2\right)\right]}{\exp\left[-\frac{1}{2\sigma_i^2}\left(\left(\theta_i^*\right)^2 - 2\theta_i^{(t)}\theta_i^* + \left(\theta_i^{(t)}\right)^2\right)\right]}$$

$$= 1$$

The use of independent normal proposals implies that the Markov chain associated with the algorithm is a random walk (Robert and Casella 2004, p. 287-291). Despite its simplicity, random walk Metropolis-Hastings algorithms are not uniformly ergodic (Robert and Casella 2004, p. 288) and consequently may be slow to converge to the posterior distribution of interest. Given the likely complexity of the surface of the posterior distribution however, it is not clear *apriori* what a better strategy would be to propose the parameter vector θ.

The prior distributions used here are non-informative proper priors, specifically independent uniform priors so that

$$\begin{aligned} f(\theta) &= \prod_{i=1}^{k} f_i(\theta_i) \\ &= \prod_{i=1}^{k} U_i(\theta_i \mid l_i, u_i) \\ &= \prod_{i=1}^{k} \frac{1}{u_i - l_i} \end{aligned}$$

The parameters l_i, u_i were chosen so that the parameter space Θ^k of the Gibbs posterior density would be large (Table 1). Put another way, by choosing l_i to be extremely small and u_i to be extremely large relative to "typical" values of θ_i, the uniform priors will have practically no influence on the realized path of the Markov chain of the Metropolis-Hastings algorithm. Since the uniform priors only depend upon the value of the parameters and not on the value of θ_i, the ratio of priors in Eq. 124 can be simplified to

Table 1. Values of parameters of uniform priors (l, u) for the triplets models.

Exponential Triplets Process (EXPTP)			
Parameter of point process	β	θ_1	θ_2
Uniform – l	0.030	-100	0
Uniform – u	1,000	100	1,000

Geyer Triplets Process (GTP)			
Parameter of point process	β	θ_1	θ_2
Uniform – l	0.030	-100	0
Uniform – u	1,000	100	1,000

Serpentine Triplets Process (SPTP)				
Parameter of point process	β	θ_1	θ_2	θ_3
Uniform – l	0.030	-15	-15	0
Uniform – u	1,000	15	15	1,000

$$\frac{f\left(\theta^*\right)}{f\left(\theta^{(t)}\right)} = I_{\Theta^k}\left(\theta^*\right)$$

This implies that if the proposed parameter vector is feasible (e.g. $\theta^* \in \Theta^k$), then the ratio of priors is one. Otherwise the proposed parameter vector is automatically rejected and the Markov chain remains at the current state $\theta^{(t)}$. When combined with the independent normal parameter proposal distributions, the Metropolis-Hastings ratio takes on the simple form:

$$r\left(\theta^{*},\theta^{(t)}\right)=I_{\Theta^{k}}\left(\theta^{*}\right)\frac{g\left(\mathbf{x}\mid\theta^{*}\right)}{g\left(\mathbf{x}\mid\theta^{(t)}\right)}\frac{Z\left(\theta^{(t)}\right)}{Z\left(\theta^{*}\right)} \qquad [125]$$

To simplify the analysis, both the hard-core and interaction distance parameters were fixed throughout the Metropolis-Hastings simulations of Gibbs posterior densities. The modification of stem-coordinate data for multiple trees with the same location (see Sec 3.2), implied that after data processing no trees were closer than 0.1m apart, as tree locations were mapped to the nearest 0.1m. The hard-core parameter for all Gibbs point process models was set to 0.09, ensuring that the Gibbs point process models would treat the observed data as feasible configurations, i.e. $f\left(\mathbf{x}\mid\theta\right)>0$, for all observed data **x**. For the interaction distance parameter *R*, no consensus was found in the literature to guide this choice. Without any definitive interaction distance to use for the Coweeta data, different interaction distances within the range of 4 to 10m were tested with the Gibbs point processes.

3.72 Estimation of ratios of normalizing constants

In the Metropolis-Hastings scheme to sample from a posterior distribution given the Gibbs point process likelihood, Eqs. 124, 125 indicate the need to estimate the ratio of normalizing constants given by

$$\frac{Z\left(\theta^{(t)}\right)}{Z\left(\theta^{*}\right)}$$

where $\theta^{(t)}$ is the current state of Markov chain (i.e. the current value of the parameter vector of the posterior density) and θ^* is the proposed parameter vector at iteration $t + 1$. Chen and Shao (1997) reviewed methods to estimate the ratio of analytically intractable normalizing constants including importance sampling (Geweke 1989), bridge sampling (Meng and Wong 1996), and path sampling (Gelman and Meng 1994, 1998). Chen and Shao (1997) proposed the method of ratio importance sampling which they argued would be optimal in terms of minimizing relative asymptotic mean square error when compared with other existing estimators. Ratio importance sampling applies the methodology of importance sampling in the following way. Define

$$f\left(\mathbf{x} \mid \theta^{(t)}\right) = \frac{g\left(\mathbf{x} \mid \theta^{(t)}\right)}{Z\left(\theta^{(t)}\right)} = \frac{g_1(\mathbf{x})}{Z_1}$$

$$f\left(\mathbf{x} \mid \theta^{*}\right) = \frac{g\left(\mathbf{x} \mid \theta^{*}\right)}{Z\left(\theta^{*}\right)} = \frac{g_2(\mathbf{x})}{Z_2}$$

$$f\left(\mathbf{x} \mid \psi\right) = \frac{g\left(\mathbf{x} \mid \psi\right)}{Z\left(\psi\right)} = \frac{g_\psi(\mathbf{x})}{Z_\psi}$$

for all $\left\{\mathbf{x} \in \mathbb{N}: f(\mathbf{x}) > 0\right\}$, i.e. the set of finite point configurations for which the Gibbs point process density is positive. As in Sec. 2.21, only finite Gibbs processes defined on a bounded region $F \subset \mathbb{R}^{\mathrm{d}}$ are considered, and $f\left(\mathbf{x} \mid \theta^{(t)}\right)$ and $f\left(\mathbf{x} \mid \theta^{*}\right)$ are Gibbs point process densities with respect to the Poisson process with intensity measure Λ. The normalizing constants of these densities are defined through Eq. 47. For now suppose

there exists a point process on $\mathbb{N}$ with σ-finite probability distribution P that is absolutely continuous with respect to Lebesgue measure so that the density $f(\mathbf{x}|\psi)$ exists. It is straightforward to verify the ratio importance sampling identity

$$\begin{aligned}
\frac{\mathbb{E}_{f(\mathbf{x}|\psi)}\left[g_1(\mathbf{x})/f(\mathbf{x}|\psi)\right]}{\mathbb{E}_{f(\mathbf{x}|\psi)}\left[g_2(\mathbf{x})/f(\mathbf{x}|\psi)\right]} &= \frac{\int_{F^n}\left[g_1(\mathbf{x})/f(\mathbf{x}|\psi)\right]P(d\mathbf{x})}{\int_{F^n}\left[g_2(\mathbf{x})/f(\mathbf{x}|\psi)\right]P(d\mathbf{x})} \\
&= \frac{\int_{F^n}\left[g_1(\mathbf{x})/f(\mathbf{x}|\psi)\right]f(\mathbf{x}|\psi)\mu_L(d\mathbf{x})}{\int_{F^n}\left[g_2(\mathbf{x})/f(\mathbf{x}|\psi)\right]f(\mathbf{x}|\psi)\mu_L(d\mathbf{x})} \\
&= \frac{\lambda\frac{e^{-\Lambda(F)}}{n!}\int_{F^n}g_1(\mathbf{x})\mu_L(d\mathbf{x})}{\lambda\frac{e^{-\Lambda(F)}}{n!}\int_{F^n}g_2(\mathbf{x})\mu_L(d\mathbf{x})} \\
&= \frac{\frac{e^{-\Lambda(F)}}{n!}\int_{F^n}g_1(\mathbf{x})\lambda\mu_L(d\mathbf{x})}{\frac{e^{-\Lambda(F)}}{n!}\int_{F^n}g_2(\mathbf{x})\lambda\mu_L(d\mathbf{x})} \\
&= \frac{\frac{e^{-\Lambda(F)}}{n!}\int_{F^n}g_1(\mathbf{x})\Lambda(d\mathbf{x})}{\frac{e^{-\Lambda(F)}}{n!}\int_{F^n}g_2(\mathbf{x})\Lambda(d\mathbf{x})} \\
&= \frac{Z_1}{Z_2}
\end{aligned}$$

Consequently, the ratio of normalizing constants can be approximated by

$$\frac{\mathbb{E}_{f(\mathbf{x}|\psi)}\left[g_1(\mathbf{x})/f(\mathbf{x}|\psi)\right]}{\mathbb{E}_{f(\mathbf{x}|\psi)}\left[g_2(\mathbf{x})/f(\mathbf{x}|\psi)\right]} \approx \frac{\frac{1}{N}\sum_{i=1}^{N} g_1(\mathbf{x}_i)/f(\mathbf{x}_i|\psi)}{\frac{1}{N}\sum_{i=1}^{N} g_2(\mathbf{x}_i)/f(\mathbf{x}_i|\psi)} \qquad [126]$$

where $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$ form a sample of size N from the density f with parameters ψ. As with all importance sampling estimators, the accuracy of the approximation in Eq. 126 depends critically upon the choice of the density $f(\mathbf{x}|\psi)$. Chen and Shao (1997) proposed to choose the optimal importance sampling density f_{opt} by

$$f_{opt}(\mathbf{x}|\psi) = \frac{\left|g_1(\mathbf{x})/Z_1 - g_2(\mathbf{x})/Z_2\right|}{\int_{F^n} \left|g_1(\mathbf{x}')/Z_1 - g_2(\mathbf{x}')/Z_2\right| P(d\mathbf{x}')} \qquad [127]$$

Unfortunately, the optimal density in Eq. 127 depends upon the unknown normalizing constants Z_1 and Z_2 and is not useful in practice. To circumvent this, Chen and Shao proposed the following two-stage sampling scheme:

1. Draw $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$ as a sample of size N from the density $f(\mathbf{x}|\psi)$.
2. Define $\tau_N = \frac{\frac{1}{N}\sum_{i=1}^{N} g_1(\mathbf{x}_i)/f(\mathbf{x}_i|\psi)}{\frac{1}{N}\sum_{i=1}^{N} g_2(\mathbf{x}_i)/f(\mathbf{x}_i|\psi)}$

3. Conditional upon τ_N and for all $\mathbf{y} \in \mathbb{N}$ define the density

$$\Psi_N(\mathbf{y} \mid \tau_N) = \frac{\left| g_1(\mathbf{y}) - \tau_N g_2(\mathbf{y}) \right|}{\int_{F^n} \left| g_1(\mathbf{y}') - \tau_N g_2(\mathbf{y}') \right| P(d\mathbf{y}')}$$

4. Draw $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_M$ as a sample of size M from the density $\Psi_N(\mathbf{y} \mid \tau_N)$.

5. Estimate the ratio of normalizing constants as

$$\frac{Z_1}{Z_2} \approx \frac{\frac{1}{M} \sum_{i=1}^{M} g_1(\mathbf{y}_i) / \Psi_N(\mathbf{y}_i \mid \tau_N)}{\frac{1}{M} \sum_{i=1}^{M} g_2(\mathbf{y}_i) / \Psi_N(\mathbf{y}_i \mid \tau_N)}$$

As pointed out by Bognar (2002), in working with Gibbs point processes this two-stage sampling scheme will be very computationally expensive to carry out. To draw the sample of point configurations $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$ from $f(\mathbf{x} \mid \psi)$, typically an MCMC algorithm will be necessary, which must be run for a sufficiently large number of iterations to sample past the burn-in period. Given these samples, a new density $\Psi_N(\mathbf{y} \mid \tau_N)$ is created and a sample $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_M$ must be drawn, which will also require an MCMC algorithm running for a sufficiently large number of iterations. Consequently, the "optimal" ratio sampling estimator will require two MCMC algorithms to be generated at *every* iteration of the Metropolis Hastings algorithm that is drawing samples from the Gibbs posterior distribution in order to estimate the normalizing constant ratio.

As an alternative to the optimal density in Eq. 127, Bognar (2002, 2005) proposed using a Gibbs point process density where the parameter ψ is "close" to the values of $\theta^{(t)}$ and θ^*. This implies that the ratio importance sampling identity becomes

$$\begin{aligned}
\frac{\mathbb{E}_{f(\mathbf{x}|\psi)}\left[g_1(\mathbf{x})/f(\mathbf{x}|\psi)\right]}{\mathbb{E}_{f(\mathbf{x}|\psi)}\left[g_2(\mathbf{x})/f(\mathbf{x}|\psi)\right]} &= \frac{\int_{F^n}\left[g_1(\mathbf{x})/f(\mathbf{x}|\psi)\right]P(d\mathbf{x})}{\int_{F^n}\left[g_2(\mathbf{x})/f(\mathbf{x}|\psi)\right]P(d\mathbf{x})} \\
&= \frac{\int_{F^n}\left[g_1(\mathbf{x})\Big/\frac{g_\psi(\mathbf{x})}{Z_\psi}\right]P(d\mathbf{x})}{\int_{F^n}\left[g_2(\mathbf{x})\Big/\frac{g_\psi(\mathbf{x})}{Z_\psi}\right]P(d\mathbf{x})} \\
&= \frac{Z_\psi\int_{F^n}\left[g_1(\mathbf{x})/g_\psi(\mathbf{x})\right]P(d\mathbf{x})}{Z_\psi\int_{F^n}\left[g_2(\mathbf{x})/g_\psi(\mathbf{x})\right]P(d\mathbf{x})} \\
&= \frac{\int_{F^n}\left[g_1(\mathbf{x})/g_\psi(\mathbf{x})\right]P(d\mathbf{x})}{\int_{F^n}\left[g_2(\mathbf{x})/g_\psi(\mathbf{x})\right]P(d\mathbf{x})}
\end{aligned} \qquad [128]$$

Consequently the ratio of normalizing constants can be approximated through

$$\frac{Z_1}{Z_2} \approx \frac{\frac{1}{N}\sum_{i=1}^{N} g_1(\mathbf{x}_i)/g(\mathbf{x}_i|\psi)}{\frac{1}{N}\sum_{i=1}^{N} g_2(\mathbf{x}_i)/g(\mathbf{x}_i|\psi)} \qquad [129]$$

where the configurations $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$ are a sample from the Gibbs density $f(\mathbf{x}|\psi)$. Bognar proposed two different ways of implementing this ratio sampling approach. The first approach "blankets" the parameter space Θ^k with a grid of possible values of $\psi_j, j = 1,2, \ldots, J$. For any given values of $\theta^{(t)}$ and θ^*, the closest ψ_j is chosen so that the importance samples are generated from $f(\mathbf{x}|\psi_j)$ (Bognar 2004). The advantage of this grid-based approach is that all of the importance samples needed to estimate ratios of normalizing constants can be generated in advance of the execution of the Metropolis-Hastings sampler for the Gibbs posterior distribution. This greatly improves the computational speed of the algorithm since once θ^* is proposed, the computer code that implements the algorithm need only determine the value of ψ_j and retrieve the corresponding importance samples from disk storage to compute the ratio of normalizing constants. As noted by Bognar (2004), a primary disadvantage of the method is that it is only feasible when the dimension of the parameter space (k) is small. With increasing k, the number of values of ψ_j needed to generate a sufficiently fine grid will become exceedingly large. Additionally, for each ψ_j generated on the grid, the N importance samples must be saved in RAM or on disk, which requires substantial computing resources. In the context of the triplets models proposed here, for each of the JN importance samples, either the sampled configurations or the matrices listing all of the pairs and triples of interacting points derived from the sampled configurations would need to be stored to evaluate the un-normalized densities in Eq. 128. Because these storage considerations proved to be too demanding given the computing resources

available, a modification of the implementation of ratio importance sampling described by Bognar (2005) is used:

1. Given the current state of the Markov chain $\theta^{(t)}$, propose the parameter vector θ^*.
2. Set ψ as the average of $\theta^{(t)}$ and θ^*, that is for each of the *k* parameters in the vectors $\theta^{(t)}$ and θ^*, set $\psi = \frac{\theta^* + \theta^{(t)}}{2}$.
3. Given ψ, draw the sample of point configurations $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$ from the Gibbs density $f(\mathbf{x} \mid \psi)$ using Geyer and Møller's (1994) birth-death-shift Metropolis-Hastings algorithm as described in Sec. 3.6.
4. Estimate the ratio of normalizing constants through Eq. 129, and use this estimate to find the value of the Metropolis-Hastings ratio $r\left(\theta^*, \theta^{(t)}\right)$ (Eq. 125) to decide whether or not to accept θ^*.

In Bognar's (2005) implementation, a systematic scan Metropolis-Hastings algorithm combined with ratio importance sampling for the normalizing constant ratio is used to sample the Gibbs posterior density. The systematic scan Metropolis Hastings algorithm updates one of the *k* parameters of θ at each iteration and systematically updates all of the parameters through *k* iterations. In the implementation here, all *k* parameters of θ are updated at once in one iteration. The reason that the systematic scan Metropolis-Hastings algorithm is not used here is that evaluation of the Gibbs likelihood for triplets models is computationally expensive. Notice that for the triplets models in Sec. 3.4, the Gibbs density contains the product:

$$\prod_{i<j<k} \exp\left[-\phi_3\left(\|x_i - x_j\|,\ \|x_i - x_k\|,\ \|x_j - x_k\|\right)\right]$$

For a configuration with *n* points, the potential function $\phi_3(\cdot,\cdot,\cdot)$ must be evaluated for each of the $\binom{n}{3}$ triples that can be formed from the configuration. For the full 80 x 80m Plot 318 from Coweeta for example, this would imply that there are a total of $\binom{349}{3} = 7,023,974$ triples to evaluate. In practice only the number of *interacting* triples, that is triples where all three points are within the interaction distance *R* of one another, needs to be determined. All triples not having mutual neighbors within a distance *R* must have zero potential. Nevertheless, determining which of the 7,023,974 triples are interacting and have non-zero potential is still a computational burden, and so the Metropolis-Hastings algorithm used here is constructed to keep the number of Metropolis-Hastings updates (and evaluations of the Gibbs likelihood for them) to a minimum.

Bognar (2005, p.11) pointed out that "Because the respective [normalizing constant] ratios … are *approximated* in the acceptance probability, their product will not equal one as it should. Hence, the likelihood function is not constant, which implies that the established theory for MCMC samplers does not apply." From a theoretical point of view, the error in the approximation given in Eq. 129 implies that detailed balance does not hold, though it may be approximately true. The degree of error in this approximation can have serious consequences for the behavior of the Metropolis-Hastings sampler of the Gibbs posterior density. This belies the need to check on the quality of the approximation

of the ratio of normalizing constants. One way to check on the quality of importance sampling estimates is by examining the variability of the obtained importance sampling weights. In the context of standard importance sampling (Robert and Casella 2004, p. 90-107), it is desired to approximate

$$\mathbb{E}_{f(\mathbf{x})}\left[h(\mathbf{x})\right] = \int_{\Omega} h(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \qquad [130]$$

that is, the expectation of a function h with respect to the density $f(\mathbf{x})$. When it is difficult to sample directly from f, importance sampling allows for the approximation of the expectation by drawing samples from an instrumental density $g(\mathbf{x})$ so that

$$\begin{aligned} \mathbb{E}_{f(\mathbf{x})}\left[h(\mathbf{x})\right] &= \int_{\Omega} h(\mathbf{x}) \frac{f(\mathbf{x})}{g(\mathbf{x})} g(\mathbf{x}) d\mathbf{x} \\ &\approx \frac{1}{N} \sum_{i=1}^{N} \frac{f(\mathbf{x}_i)}{g(\mathbf{x}_i)} h(\mathbf{x}_i) \end{aligned} \qquad [131]$$

Ideally, $g(\mathbf{x})$ is chosen so that the ratio of densities $f(\mathbf{x})/g(\mathbf{x})$ is bounded and $\mathrm{var}_{f(\mathbf{x})}\left[h(\mathbf{x})\right] < \infty$. If $g(\mathbf{x})$ is chosen so that the ratio of densities is *not* bounded, then the importance sampling weights $f(\mathbf{x}_i)/g(\mathbf{x}_i)$ can vary widely across the different the different samples $\mathbf{x}_i$, leading to instability in the importance sampling estimator (Geweke 1989).

Comparing the ratio importance sampling identity in Eq. 128 and the resulting approximation for the ratio of normalizing constants in Eq. 129 with Eq. 131, it can be seen that $h(\mathbf{x}) = 1$ and the importance sampling weights are $g_1(\mathbf{x}_i)/g(\mathbf{x}_i \mid \psi)$ and $g_2(\mathbf{x}_i)/g(\mathbf{x}_i \mid \psi)$, corresponding to the estimation of the normalizing constants Z_1 and Z_2, respectively. In implementing the Metropolis-Hastings algorithm to sample from the Gibbs posterior distribution, at each iteration of the Gibbs posterior sampler, the mean, variance and coefficient of variation (CV) of the importance sampling weights for both Z_1 and Z_2 were computed. A check on the quality of the importance sample estimates of the normalizing constant ratio was then obtained by plotting the histogram of the CVs of the importance sampling weights generated from all iterations.

During implementation of the importance sampling based estimates of the ratio of normalizing constants, one technical challenge that became apparent was that the numerical values of the unnormalized Gibbs density, $g(\mathbf{x})$, were extremely small ($< e^{-10}$) and led to numerical underflows. To solve this all un-normalized densities were transformed to their natural logarithm so that the Metropolis-Hastings ratios were also log-transformed, for example Eq. 144 becomes

$$\begin{aligned} \ln\left(r\left(\theta^*, \theta^{(t)}\right)\right) &= \ln\left(I_{\Theta^k}\left(\theta^*\right) \frac{g\left(\mathbf{x} \mid \theta^*\right)}{g\left(\mathbf{x} \mid \theta^{(t)}\right)} \frac{Z\left(\theta^{(t)}\right)}{Z\left(\theta^*\right)} \right) \\ &= \ln\left(I_{\Theta^k}\left(\theta^*\right)\right) + \ln\left(g\left(\mathbf{x} \mid \theta^*\right)\right) - \ln\left(g\left(\mathbf{x} \mid \theta^{(t)}\right)\right) + \ln\left(\frac{Z\left(\theta^{(t)}\right)}{Z\left(\theta^*\right)} \right) \end{aligned} \qquad [132]$$

To preserve the Metropolis-Hastings decision rule (Eq. 123), a random uniform number u is drawn on [0,1] and is log-transformed so that θ^* is accepted whenever

$$\ln(u) < \ln\left(r\left(\theta^*, \theta^{(t)}\right)\right) \qquad [133]$$

While this resolved the problem of underflows in computing ratios of unnormalized densities, to find the log-ratio of normalizing constants, the averages in Eq. 129 need to be computed in terms of ln(g(**x**)), not g(**x**). The computation of CV of the importance sampling weights needs to be computed in similar terms. These issues were resolved by the following scaling procedure:

1. Define the importance sampling weights $w_{1i} = \frac{g_1(\mathbf{x}_i)}{g(\mathbf{x}_i \mid \psi)}$ and $w_{2i} = \frac{g_2(\mathbf{x}_i)}{g(\mathbf{x}_i \mid \psi)}$ for $i = 1, 2, \ldots, N$.

2. Find the maximum importance sampling weights $w_1^* = \max_i \{w_{1i}\}$ and $w_2^* = \max_i \{w_{2i}\}$.

3. Define the scaled importance sampling weights $w_{1i}^s = \frac{w_{1i}}{w_1^*}$ and $w_{2i}^s = \frac{w_{2i}}{w_1^*}$

4. Calculate the scaled weights from the importance samples

$$w_{1i}^{s} = \exp\left(\ln\left(\frac{w_{1i}}{w_1^*} \right) \right) = \exp\left(\ln\left(w_{1i} \right) - \ln\left(w_1^* \right) \right) \text{ and}$$

$$w_{2i}^{s} = \exp\left(\ln\left(\frac{w_{2i}}{w_1^*} \right) \right) = \exp\left(\ln\left(w_{2i} \right) - \ln\left(w_2^* \right) \right)$$

5. Calculate the average importance sample weights

$$\overline{w_1} = \frac{1}{N}\sum_{i=1}^{N} w_{1i} = \frac{w_1^*}{N}\sum_{i=1}^{N} \frac{w_{1i}}{w_1^*} = \frac{w_1^*}{N}\sum_{i=1}^{N} w_{1i}^{s} = w_1^* \overline{w_1^s}$$

$$\overline{w_2} = \frac{1}{N}\sum_{i=1}^{N} w_{2i} = \frac{w_2^*}{N}\sum_{i=1}^{N} \frac{w_{2i}}{w_1^*} = \frac{w_2^*}{N}\sum_{i=1}^{N} w_{2i}^{s} = w_2^* \overline{w_2^s}$$

6. Calculate the standard deviation of the importance sampling weights

$$\begin{aligned}
\sqrt{\frac{1}{N-1}\sum\left(w_{1i} - \overline{w_1} \right)^2} &= \sqrt{\frac{1}{N-1}\sum\left(w_1^* w_{1i}^s - w_1^* \overline{w_1^s} \right)^2} \\
&= \sqrt{\frac{1}{N-1}\sum\left(w_1^* \right)^2 \left(w_{1i}^s - \overline{w_1^s} \right)^2} \\
&= w_1^* \sqrt{\frac{1}{N-1}\sum\left(w_{1i}^s - \overline{w_1^s} \right)^2}
\end{aligned}$$

$$\begin{aligned}
\sqrt{\frac{1}{N-1}\sum\left(w_{2i} - \overline{w_2} \right)^2} &= \sqrt{\frac{1}{N-1}\sum\left(w_2^* w_{2i}^s - w_2^* \overline{w_2^s} \right)^2} \\
&= \sqrt{\frac{1}{N-1}\sum\left(w_2^* \right)^2 \left(w_{2i}^s - \overline{w_2^s} \right)^2} \\
&= w_2^* \sqrt{\frac{1}{N-1}\sum\left(w_{2i}^s - \overline{w_2^s} \right)^2}
\end{aligned}$$

7. Finally, calculate the CV of the importance sampling weights

$$CV\left(w_1\right)=\frac{s_{w_1}}{\overline{w_1}}=\frac{w_1^*\sqrt{\frac{1}{N-1}\sum\left(w_{1i}^s-\overline{w_1^s}\right)^2}}{w_1^*\overline{w_1^s}}=\frac{\sqrt{\frac{1}{N-1}\sum\left(w_{1i}^s-\overline{w_1^s}\right)^2}}{\overline{w_1^s}}$$

$$CV\left(w_2\right)=\frac{s_{w_2}}{\overline{w_2}}=\frac{w_2^*\sqrt{\frac{1}{N-1}\sum\left(w_{2i}^s-\overline{w_2^s}\right)^2}}{w_2^*\overline{w_2^s}}=\frac{\sqrt{\frac{1}{N-1}\sum\left(w_{2i}^s-\overline{w_2^s}\right)^2}}{\overline{w_2^s}}$$

In essence the CV of the unscaled weights was found through the CV of the scaled weights, since the maximal importance sampling weight cancels out of the CV ratios. Moreover, step 4 implies that all of the computations use the logarithm of unnormalized densities, e.g.

$$\ln\left(w_{1i}\right)=\ln\left(\frac{g_1\left(\mathbf{x}_i\right)}{g\left(\mathbf{x}_i\,|\,\psi\right)}\right)=\ln\left(g_1\left(\mathbf{x}_i\right)\right)-\ln\left(g\left(\mathbf{x}_i\,|\,\psi\right)\right)$$

Provided that $\ln\left(w_{1i}\right)-\ln\left(w_1^*\right)$ and $\ln\left(w_{2i}\right)-\ln\left(w_2^*\right)$ are not too small across the importance samples, the computation of the CV of the importance sampling weights will not cause numerical underflows. Finally, for the computation of the normalizing constant ratio, it is observed that since

$$\ln\left(\frac{Z\left(\theta^{(t)}\right)}{Z\left(\theta^{*}\right)}\right) = \ln\left(Z\left(\theta^{(t)}\right)\right) - \ln\left(Z\left(\theta^{*}\right)\right)$$

$$\approx \ln\left(\frac{1}{N}\sum_{i=1}^{N} g_1\left(\mathbf{x}_i\right)/g\left(\mathbf{x}_i \mid \psi\right)\right) - \ln\left(\frac{1}{N}\sum_{i=1}^{N} g_2\left(\mathbf{x}_i\right)/g\left(\mathbf{x}_i \mid \psi\right)\right)$$

$$= \ln\left(\frac{1}{N}\sum_{i=1}^{N} w_{1i}\right) - \ln\left(\frac{1}{N}\sum_{i=1}^{N} w_{2i}\right)$$

$$= \ln\left(\frac{w_1^*}{N}\sum_{i=1}^{N} w_{1i}^s\right) - \ln\left(\frac{w_2^*}{N}\sum_{i=1}^{N} w_{2i}^s\right)$$

$$= \ln\left(w_1^*\right) + \ln\left(\sum_{i=1}^{N} w_{1i}^s\right) - \ln\left(w_2^*\right) - \ln\left(\sum_{i=1}^{N} w_{2i}^s\right)$$

the normalizing constant ratio can be computed directly from the scaled importance sampling weights and the logarithm of the maximal importance sampling weights, which again use the logarithm of unnormalized densities as needed.

3.73 Auxiliary variable method

An alternative to stochastic approximation of the normalizing constant ratio in Eq. 124 was recently proposed in Møller et al. (2006) and Berthelsen and Møller (2006). Their idea was to introduce an auxiliary variable whose proposal distribution took the same form as the likelihood function, allowing cancellation of the normalizing constants. Recalling from Sec. 3.71, the Gibbs posterior density is

$$f(\theta \mid \mathbf{x}) = \frac{f(\mathbf{x} \mid \theta) f(\theta)}{\int f(\theta) f(\mathbf{x} \mid \theta) \mu(d\theta)}$$

$$\propto f(\mathbf{x} \mid \theta) f(\theta)$$

Now introduce an auxiliary variable **y**, defined on the same state space as **x**, so that the joint density of θ and **y** is

$$f(\mathbf{y}, \theta \mid \mathbf{x}) \propto f(\mathbf{x}, \mathbf{y}, \theta) = f(\mathbf{y} \mid \theta, \mathbf{x}) f(\mathbf{x} \mid \theta) f(\theta) \quad [134]$$

Møller et al. propose the following Metropolis-Hastings algorithm:

1. Choose the initial configuration $\mathbf{y}^{(0)} \in \mathbb{N}$ and $\theta^{(0)} \in \Theta^k$ such that $f\left(\mathbf{y}^{(0)} \mid \theta^{(0)}, \mathbf{x}\right) > 0$ and $f\left(\theta^{(0)} \mid \mathbf{x}\right) > 0$ and set $t = 0$.
2. Generate a new parameter vector θ^* from a desired proposal distribution $q\left(\theta^* \mid \theta^{(t)}\right)$.
3. Conditional upon θ^*, propose a new configuration $\mathbf{y}^*$ from the proposal distribution $q\left(\mathbf{y}^* \mid \mathbf{y}^{(t)}, \theta^*, \theta^{(t)}\right)$
4. Accept the new parameter vector and configuration with probability

$$\alpha\left(\mathbf{y}^{(t)}, \theta^{(t)}, \mathbf{y}^*, \theta^*\right) = \min\left\{1, \frac{f\left(\mathbf{y}^* \mid \theta^*, \mathbf{x}\right)}{f\left(\mathbf{y}^{(t)} \mid \theta^{(t)}, \mathbf{x}\right)} \frac{f\left(\theta^* \mid \mathbf{x}\right)}{f\left(\theta^{(t)} \mid \mathbf{x}\right)} \frac{q\left(\mathbf{y}^{(t)} \mid \mathbf{y}^*, \theta^*, \theta^{(t)}\right)}{q\left(\mathbf{y}^* \mid \mathbf{y}^{(t)}, \theta^*, \theta^{(t)}\right)} \frac{q\left(\theta^{(t)} \mid \theta^*\right)}{q\left(\theta^* \mid \theta^{(t)}\right)}\right\} \quad [135]$$

5. Increment t and repeat until $t = T$

Since the proposal distribution for the auxiliary variable **y** is arbitrary, Møller et al. (2006) make the clever choice of using the likelihood itself, that is

$$q\left(\mathbf{y}^{(t)} \mid \mathbf{y}^{*}, \theta^{*}, \theta^{(t)}\right) = f\left(\mathbf{y}^{(t)} \mid \theta^{(t)}\right) = \frac{g\left(\mathbf{y}^{(t)} \mid \theta^{(t)}\right)}{Z\left(\theta^{(t)}\right)}$$

and

$$q\left(\mathbf{y}^{*} \mid \mathbf{y}^{(t)}, \theta^{*}, \theta^{(t)}\right) = f\left(\mathbf{y}^{*} \mid \theta^{*}\right) = \frac{g\left(\mathbf{y}^{*} \mid \theta^{*}\right)}{Z\left(\theta^{*}\right)}$$

so that the Metropolis Hastings ratio becomes

$$r\left(\mathbf{y}^{*}, \mathbf{y}^{(t)} \theta^{*}, \theta^{(t)}\right)$$

$$= \frac{f\left(\mathbf{y}^{*} \mid \theta^{*}, \mathbf{x}\right)}{f\left(\mathbf{y}^{(t)} \mid \theta^{(t)}, \mathbf{x}\right)} \frac{f\left(\theta^{*} \mid \mathbf{x}\right)}{f\left(\theta^{(t)} \mid \mathbf{x}\right)} \frac{q\left(\mathbf{y}^{(t)} \mid \mathbf{y}^{*}, \theta^{*}, \theta^{(t)}\right)}{q\left(\mathbf{y}^{*} \mid \mathbf{y}^{(t)}, \theta^{*}, \theta^{(t)}\right)} \frac{q\left(\theta^{(t)} \mid \theta^{*}\right)}{q\left(\theta^{*} \mid \theta^{(t)}\right)}$$

$$= \frac{f\left(\mathbf{y}^{*} \mid \theta^{*}, \mathbf{x}\right)}{f\left(\mathbf{y}^{(t)} \mid \theta^{(t)}, \mathbf{x}\right)} \frac{f\left(\mathbf{x} \mid \theta^{*}\right)}{f\left(\mathbf{x} \mid \theta^{(t)}\right)} \frac{f\left(\theta^{*}\right)}{f\left(\theta^{(t)}\right)} \frac{q\left(\mathbf{y}^{(t)} \mid \mathbf{y}^{*}, \theta^{*}, \theta^{(t)}\right)}{q\left(\mathbf{y}^{*} \mid \mathbf{y}^{(t)}, \theta^{*}, \theta^{(t)}\right)} \frac{q\left(\theta^{(t)} \mid \theta^{*}\right)}{q\left(\theta^{*} \mid \theta^{(t)}\right)}$$

$$= \frac{f\left(\mathbf{y}^{*} \mid \theta^{*}, \mathbf{x}\right)}{f\left(\mathbf{y}^{(t)} \mid \theta^{(t)}, \mathbf{x}\right)} \frac{\frac{g\left(\mathbf{x} \mid \theta^{*}\right)}{Z\left(\theta^{*}\right)}}{\frac{g\left(\mathbf{x} \mid \theta^{(t)}\right)}{Z\left(\theta^{(t)}\right)}} \frac{f\left(\theta^{*}\right)}{f\left(\theta^{(t)}\right)} \frac{\frac{g\left(\mathbf{y}^{(t)} \mid \theta^{(t)}\right)}{Z\left(\theta^{(t)}\right)}}{\frac{g\left(\mathbf{y}^{*} \mid \theta^{*}\right)}{Z\left(\theta^{*}\right)}} \frac{q\left(\theta^{(t)} \mid \theta^{*}\right)}{q\left(\theta^{*} \mid \theta^{(t)}\right)}$$

$$= \frac{f\left(\mathbf{y}^* \mid \theta^*, \mathbf{x}\right)}{f\left(\mathbf{y}^{(t)} \mid \theta^{(t)}, \mathbf{x}\right)} \frac{g\left(\mathbf{x} \mid \theta^*\right)}{g\left(\mathbf{x} \mid \theta^{(t)}\right)} \frac{f\left(\theta^*\right)}{f\left(\theta^{(t)}\right)} \frac{g\left(\mathbf{y}^{(t)} \mid \theta^{(t)}\right)}{g\left(\mathbf{y}^* \mid \theta^*\right)} \frac{q\left(\theta^{(t)} \mid \theta^*\right)}{q\left(\theta^* \mid \theta^{(t)}\right)} \qquad [136]$$

so that the normalizing constants in the likelihood cancel the normalizing constants of the proposal density for **y**. Møller et al. (2006) recommend that proposals $\mathbf{y}^*$ be perfect samples from the proposal distribution $f\left(\mathbf{y} \mid \theta\right)$ so that the posterior Gibbs distribution will be obtained without error from the marginalization of the joint density $f\left(\mathbf{y}, \theta \mid \mathbf{x}\right)$.

For point processes, drawing perfect samples has been made possible by the adaptation of Propp and Wilson's (1996) coupling from the past (CFTP) ideas to so-called dominated CFTP by Kendall (1998), and Kendall and Møller (2000); an alternative adaptation of CFTP for point processes is the clan of ancestors algorithm proposed by Ferrari et al (2002) and extended to marked point processes by Van Lieshout and Stoica (2006). Both dominated CFTP and the clan of ancestors algorithms are applicable for perfect simulation of locally stable point processes, whose conditional intensity function $\lambda^*\left(u; \mathbf{x}\right)$ is bounded by a positive constant *K*.

To illustrate the key principle behind CFTP, suppose for some finite state space Ω that a sequence of Markov chains is constructed where each Markov chain is represented as a random mapping, also known as a stochastic recursive sequence, such that $X_{t+1} = f\left(X_t, u_t\right)$ where *f* is a deterministic function and the u_t are i.i.d. random variables. Because the sequence of Markov chains are based on the same random mappings, any two chains that have the same state at time *t* will have identically the same sequence after time *t* regardless of what states the two chains were started at. Now suppose that the

sequence of Markov chains are initialized at all possible states of Ω at some point in the past, $-T$. As the chains proceed forward in time, paths realized by these chains will *coalesce* when they choose the same state. If all paths have coalesced by time 0, then the distribution of this Markov chain at time 0 is the same as if the sequence of Markov chains had been started at time $-\infty$, implying that this distribution is the stationary distribution of the chain and the state of the chain at time 0 is an exact sample from the stationary distribution (Robert and Casella 2004). Now if all paths have not coalesced by time 0, then the algorithm picks a time more distant in the past, say $-2T$, and proceeds forward in time while reusing the same random numbers u_t generated in evaluating the sequence of chains from time $-T$ to time 0. This process is repeated by successively choosing times more distant in the past until coalescence of all chains has been achieved.

Because point processes are defined on continuous (infinite) state spaces, Propp and Wilson's CFTP idea must be modified since there are an infinite number of states and Markov chains needed to create the sequence just described. As noted by Propp and Wilson (1996), if there exists a partial *ordering*, $\preceq$, on the state space Ω such that $X_t \preceq Y_t$ implies $f\left(X_t, u_t\right) \leq f\left(Y_t, u_t\right)$ for all t and $X_t, Y_t \in \Omega$ where stochastically maximal and minimal states of the state space are known, then the CFTP algorithm can be applied to continuous state spaces. The coalescence of a Markov chain initialized at the maximal state with a Markov chain initialized at the minimal state by time 0 implies the coalescence of all possible states of the state space, a property referred to as sandwiching. A great benefit of this sandwiching property is that only the chains initialized by the maximal and minimal states need to be monitored (Kendall and Thönnes 1999). In the context of point processes, there is no maximal state, so Kendall

(1998) and Kendall and Møller (2000) introduced the idea of dominated CFTP, a modification of the CFTP algorithm where set inclusion is used as the partial order structure, the empty configuration Ø is a stochastically minimal state, but no stochastically maximal state is assumed. Instead a dominating process is used to provide the desired sandwiching property. The dominating process is a spatial birth and death process (Preston 1977) with its stationary distribution given by a Poisson process with intensity measure $\Lambda(B) = K\ \mu_L(B)$, where B is a Borel subset of the space $\Omega = F \subseteq \mathbb{R}^2$ as in Sec. 2.12, and K is the upper bound of the point process' conditional intensity function. The dominated CFTP algorithm proceeds by extending the dominating process backwards in time through a series of point births and deaths, where a randomly chosen point in F is born in a configuration with $n(\mathbf{x})$ points with probability $K/(K + n(\mathbf{x}))$, or otherwise a randomly selected point is deleted (a "death"). At each time a point birth occurs, the point location is recorded; at each time a point death occurs, the point location is recorded and a random uniform number on [0,1] is generated. In this way the dominating process undergoes births and deaths backwards in time until the empty configuration is achieved. Next the "target" spatial birth and death process is initialized with the empty configuration. At each time the dominating process experienced a point birth going backwards in time, the target process experiences a death going forward in time using the recorded point coordinates for the backwards point death. When the dominating process experienced a point death while going backwards in time, the recorded point is added to the target process with configuration $\mathbf{x}$ only if the recorded random uniform number associated with the point exceeds the birth rate

$b(u;\mathbf{x}) = \lambda^*(u;\mathbf{x}) / K$. Through this recording of point locations and the random uniform numbers, the target process is "coupled" with the dominating process.

While various means have been proposed to speed efficiency of the algorithm, including the use of upper and lower sandwiching processes (Kendall and Møller 2000; Berthelsen and Møller 2002), for the purpose here a critical point in the implementation of dominated CFTP is that the dominating process must be initialized from a Poisson process with intensity measure $\Lambda(B) = K\ \mu_L(B)$. In this thesis, the pairwise interaction and triplets models proposed are all locally stable, as shown in Sec. 3.5. Because the value of *K* can be large for these processes, perfect sampling using dominated CFTP as well as *imperfect* sampling via Geyer and Møller's (1994) shift-birth-death Metropolis-Hastings algorithm were attempted as a means of proposing a sample $\mathbf{y}^*$ from $f(\mathbf{y},\theta \mid \mathbf{x})$.

The choice of the conditional density for **y** is guided by the need to provide good mixing of the joint density $f(\mathbf{y},\theta \mid \mathbf{x})$. One simple choice is to use the likelihood density evaluated at a fixed parameter vector $\tilde{\theta}$ rather than either θ^* or $\theta^{(t)}$ so that

$$r\left(\mathbf{y}^*, \mathbf{y}^{(t)}\theta^*, \theta^{(t)}\right) = \frac{f\left(\mathbf{y}^* \mid \theta^*, \mathbf{x}\right)}{f\left(\mathbf{y}^{(t)} \mid \theta^{(t)}, \mathbf{x}\right)} \frac{g\left(\mathbf{x} \mid \theta^*\right)}{g\left(\mathbf{x} \mid \theta^{(t)}\right)} \frac{f\left(\theta^*\right)}{f\left(\theta^{(t)}\right)} \frac{g\left(\mathbf{y}^{(t)} \mid \theta^{(t)}\right)}{g\left(\mathbf{y}^* \mid \theta^*\right)} \frac{q\left(\theta^{(t)} \mid \theta^*\right)}{q\left(\theta^* \mid \theta^{(t)}\right)}$$

$$= \frac{\dfrac{g\left(\mathbf{y}^* \mid \tilde{\theta}\right)}{Z\left(\tilde{\theta}\right)}}{\dfrac{g\left(\mathbf{y}^{(t)} \mid \tilde{\theta}\right)}{Z\left(\tilde{\theta}\right)}} \frac{g\left(\mathbf{x} \mid \theta^*\right)}{g\left(\mathbf{x} \mid \theta^{(t)}\right)} \frac{f\left(\theta^*\right)}{f\left(\theta^{(t)}\right)} \frac{g\left(\mathbf{y}^{(t)} \mid \theta^{(t)}\right)}{g\left(\mathbf{y}^* \mid \theta^*\right)} \frac{q\left(\theta^{(t)} \mid \theta^*\right)}{q\left(\theta^* \mid \theta^{(t)}\right)}$$

$$= \frac{g\left(\mathbf{y}^{*} \mid \tilde{\theta}\right)}{g\left(\mathbf{y}^{(t)} \mid \tilde{\theta}\right)} \frac{g\left(\mathbf{x} \mid \theta^{*}\right)}{g\left(\mathbf{x} \mid \theta^{(t)}\right)} \frac{f\left(\theta^{*}\right)}{f\left(\theta^{(t)}\right)} \frac{g\left(\mathbf{y}^{(t)} \mid \theta^{(t)}\right)}{g\left(\mathbf{y}^{*} \mid \theta^{*}\right)} \frac{q\left(\theta^{(t)} \mid \theta^{*}\right)}{q\left(\theta^{*} \mid \theta^{(t)}\right)} \qquad [137]$$

Møller et al. (2006) argued that so long as the Gibbs posterior distribution is concentrated around $\tilde{\theta}$, the algorithm should be expected to mix well. Determination of $\tilde{\theta}$, however is not necessarily an easy task. Møller et al. used a maximum pseudolikelihood estimate, but it is worth noting that Baddeley and Turner (2000) found that although the pseudo-likelihood estimator is expected to work well when the data are Poisson-like, the MPLE is inefficient when the data show strong spatial interaction. The pairwise and triplets models used here are designed to model localized spatial attraction, which makes the prospect of the MPLE not as appealing. Instead, when drawing inference from a field plot from the Coweeta dataset, the posterior mode obtained from the implementation of Bognar's algorithm (Sec. 3.71) with that dataset was used for $\tilde{\theta}$. Combining these choices for the conditional density of **y** along with the uniform priors specified in Table 1 and the independent normal proposals for the parameters θ, gives the implementation of the Møller et al.'s algorithm here:

1. Choose the initial configuration $\mathbf{y}^{(0)} \in \mathbb{N}$ and $\theta^{(0)}, \tilde{\theta} \in \Theta^{k}$ such that

   $f\left(\mathbf{y}^{(0)} \mid \tilde{\theta}\right) > 0$ and $f\left(\theta^{(0)} \mid \mathbf{x}\right) > 0$ and set $t = 0$.

2. Propose a new parameter vector θ^{*} from the multivariate normal distribution

   $q\left(\theta^{*} \mid \theta^{(t)}\right) = \prod_{i=1}^{k} N_i\left(\theta_i^{*} \mid \mu_i = \theta_i^{(t)}, \sigma_i^2\right)$.

3. Conditional upon θ^*, propose a new configuration $\mathbf{y}^*$ from the Gibbs distribution

   $f\left(\mathbf{y}^* \mid \theta^*\right)$

4. Accept the new parameter vector and configuration with probability

   $$\alpha\left(\mathbf{y}^{(t)}, \theta^{(t)}, \mathbf{y}^*, \theta^*\right) = \min\left\{1,\ I_{\Theta^k}\left(\theta^*\right) \frac{g\left(\mathbf{y}^* \mid \tilde{\theta}\right)}{g\left(\mathbf{y}^{(t)} \mid \tilde{\theta}\right)} \frac{g\left(\mathbf{x} \mid \theta^*\right)}{g\left(\mathbf{x} \mid \theta^{(t)}\right)} \frac{g\left(\mathbf{y}^{(t)} \mid \theta^{(t)}\right)}{g\left(\mathbf{y}^* \mid \theta^*\right)}\right\} \quad [138]$$

5. Increment t and repeat until $t = T$

4. Results

4.1 Spatial structure of observed data

During data preprocessing, all plots were found to have cases of multiple tree records with identical coordinate locations. The number of trees records with identical coordinates varied from just two records for one tree with two stems on Plot 218 to a total of 48 tree records resulting from 18 trees with two stems each and four trees with three stems each on Plot 527 (Table 2). Multiple stemmed trees were typically either red maple (*Acer rubrum* L.) or sourwood (*Oxydendrum arboreum* L.) on Plots 118, 218, 318, and 427, whereas on Plot 527 multiple stemmed trees were typically yellow birch (*Betula alleghaniensis* Britton), black birch (*Betula lenta* L.), or basswood (*Tilia americana* L.). As a consequence of multiple stemmed trees, the coordinate locations of 12, 1, 7, 14, and 26 tree records for Plots 118, 218, 318, 427, and 527 respectively were modified as described in Sec. 3.2.

Table 2. Distribution of total tree number by tree records per coordinate location.

*No. of Trees per Coord. Location**	*Plot 118*	*Plot 218*	*Plot 318*	*Plot 427*	*Plot 527*
1	330	262	335	359	220
2	20	2	14	16	36
3	3	0	0	9	12
Total Trees on Plot	353	264	349	384	268

*A value of 1 indicates trees with unique locations. A value of 2 or 3 indicates 2 or 3 tree records with one location recorded in the field; e.g., 12 trees listed for three trees per coordinate location implies four trees of three stems each.

Table 3. Coweeta gradient plot summary statistics and basal area by species, for all trees with 1998 DBH ≥ 10 cm.

	Plot 118	*Plot 218*	*Plot 318*	*Plot 427*	*Plot 527*
Elevation (m)*	786	802	866	1085	1387
Quad. Mean DBH (cm)	23.2	30.0	29.0	26.0	31.0
Trees / ha	551.56	412.50	545.31	600.00	418.75
Basal Area (m^2 ha^{-1})	23.23	29.14	36.00	31.79	31.66
Acer pensylvanicum L.	0.00	0.02	0.08	0.31	0.81
Acer rubrum L.	2.35	4.65	4.66	6.51	0.20
Acer saccharum Marsh.	0.00	0.30	0.00	0.00	0.71
Acer spicatum Lam.	0.00	0.00	0.00	0.00	0.47
Amelanchier arborea (Michx. f.) Fern	0.24	0.00	0.01	0.00	0.12
Betula alleghaniensis Britton	0.00	0.04	0.06	0.00	7.86
Betula lenta L.	0.00	1.97	0.34	0.10	5.85
Carya glabra (Mill.) Sweet	1.42	3.39	4.22	1.24	0.09
Carya ovata (Mill.) K. Koch	0.00	0.00	0.08	0.00	0.00
Castanea dentata (Marsh.) Borkh.	0.00	0.00	0.00	0.04	0.00
Castanea pumila Mill.	0.02	0.00	0.00	0.00	0.00
Cornus florida L.	0.13	0.50	0.35	0.36	0.00
Fagus grandifolia Ehrh.	0.05	0.00	0.00	0.00	0.30
Fraxinus americana L.	0.00	0.00	0.06	0.00	1.90
Ilex spp.	0.00	0.00	0.00	0.00	0.19
Liriodendron tulipifera L.	0.09	6.77	0.93	0.00	0.00
Magnolia acuminata L.	0.00	0.00	0.00	0.01	0.12
Magnolia fraseri Walt.	0.14	0.23	0.00	0.28	0.00
Nyssa sylvatica Marsh.	0.51	0.58	3.56	2.10	0.00
Oxydendrum arboreum (L.) DC.	2.10	0.83	3.21	3.59	0.00
Pinus rigida Mill.	4.25	0.00	0.00	0.00	0.00
Quercus alba L.	0.46	0.00	0.00	0.00	0.22
Quercus coccinea Muenchh.	2.33	1.80	0.10	1.52	0.00
Quercus marilandica Muenchh.	0.27	0.00	0.00	0.00	0.00
Quercus prinus L.	4.68	3.95	12.56	10.14	0.00
Quercus rubra L.	2.24	1.16	2.97	4.26	6.72
Quercus velutina Lam.	1.10	1.33	1.58	0.23	0.00
Quercus spp.	0.00	0.00	0.00	0.27	0.00
Robinia pseudoacacia L.	0.75	1.12	0.40	0.36	0.00
Sassafras albidum (Nutt.) Nees	0.11	0.00	0.05	0.00	0.00
Tilia americana L.	0.00	0.33	0.00	0.00	5.85
Tsuga canadensis (L.) Carr.	0.00	0.15	0.03	0.38	0.19
Unknown	0.00	0.01	0.75	0.07	0.04

*Source: Clark et al. (1998)

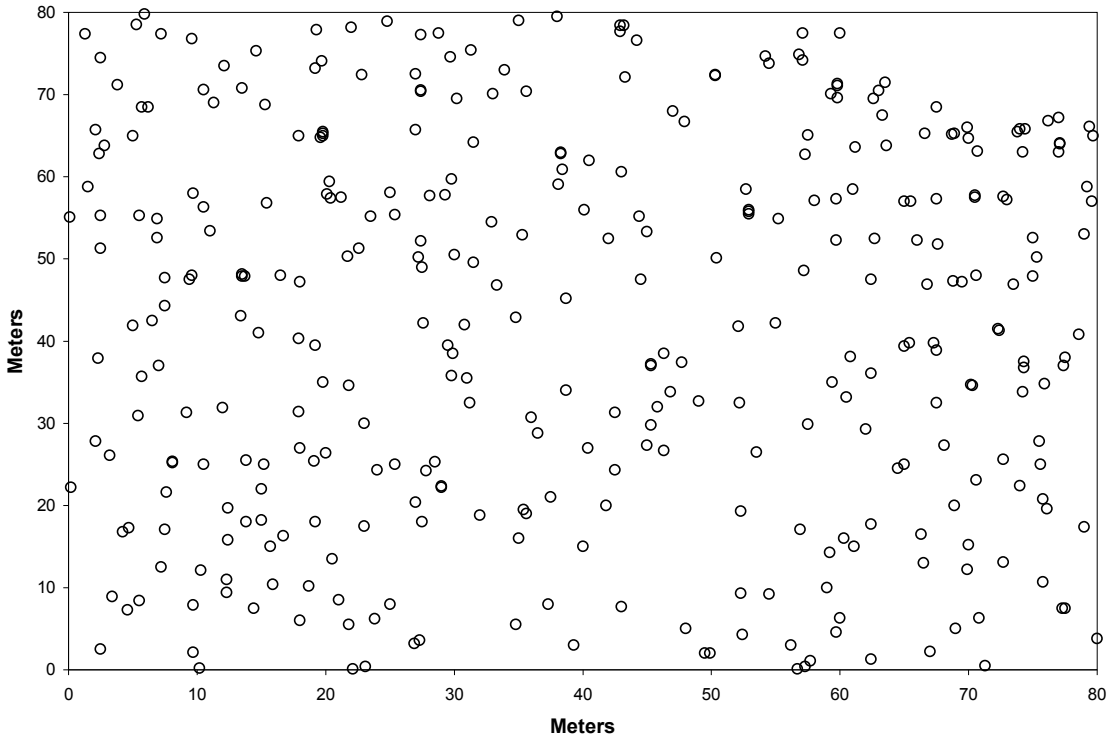

Figure 13. Stem map of Plot 118 (all trees with DBH ≥ 10cm).

In addition to restricting all trees to having a DBH at the 1998 remeasurement of at least 10 cm, the Coweeta dataset was also restricted to eliminate a few species (*Aralia spinosa* L. and *Hamamelis virginiana* L.) that are more commonly thought of as shrubs than trees, regardless of the observed stem diameter. Species composition of the final tree list for each plot (Table 3) varied from cove hardwoods (dominated by *Liriodendron tulipifera* L.) to oak-hickory (dominated by *Quercus* spp., *Carya* spp.). Plot tree density ranged from to 412.5 trees ha^{-1} on Plot 218 to 600 trees ha^{-1} on Plot 427 and plot basal areas ranged from 23.23 m^2 ha^{-1} on Plot 118 to 36.00 m^2 ha^{-1} on Plot 318 (Table 3).

After data preprocessing, including modifying coordinates of certain stems as illustrated in Figure 3, the spatial structure of each of the five Coweeta plots was assessed. Visual inspection of the stem maps for these plots (Figs. 13 – 17)

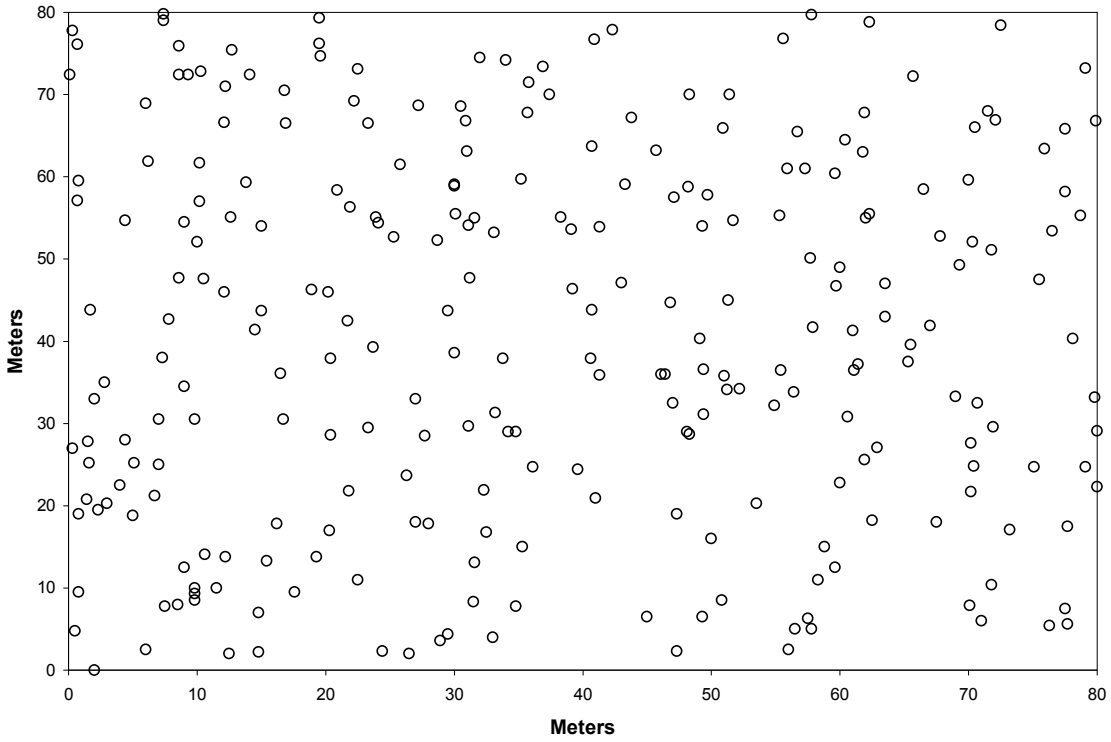

Figure 14. Stem map of Plot 218 (all trees with DBH ≥ 10cm).

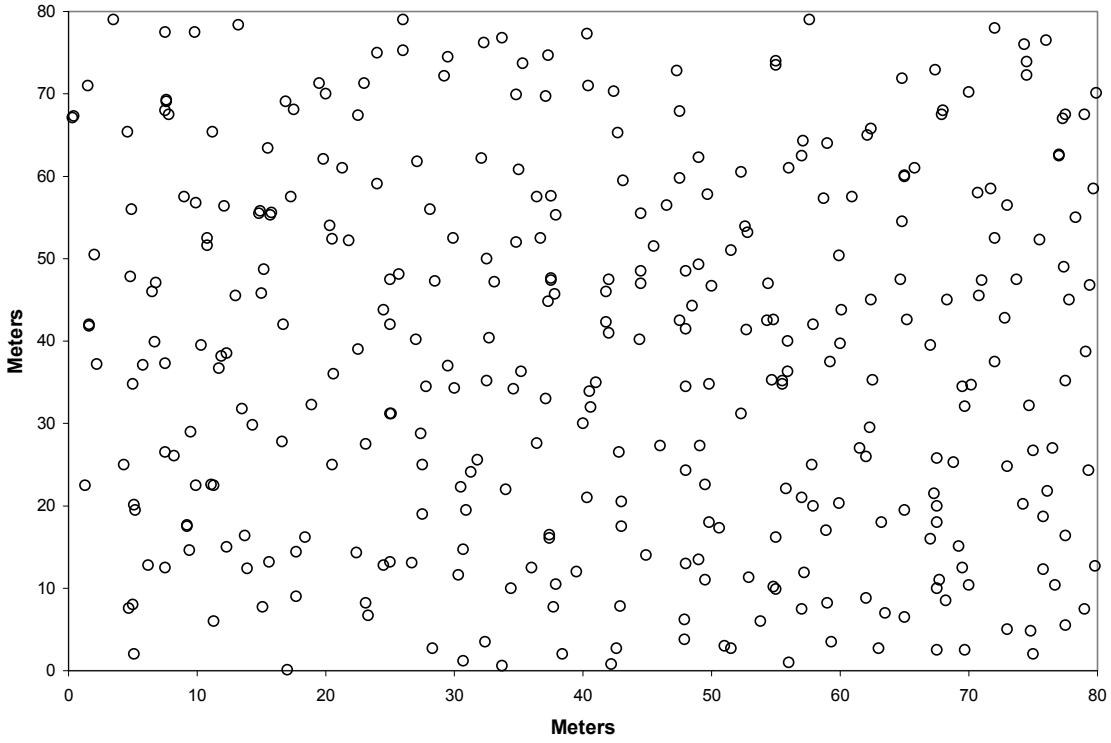

Figure 15. Stem map of Plot 318 (all trees with DBH ≥ 10cm).

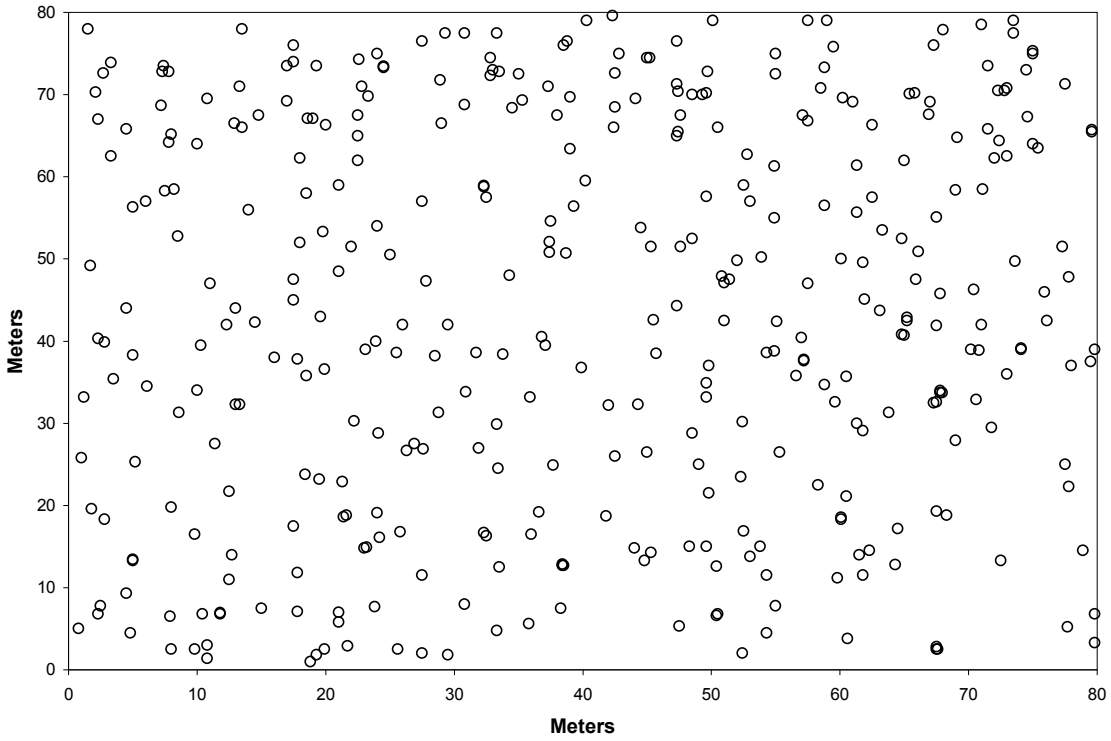

Figure 16. Stem map of Plot 427 (all trees with DBH ≥ 10cm).

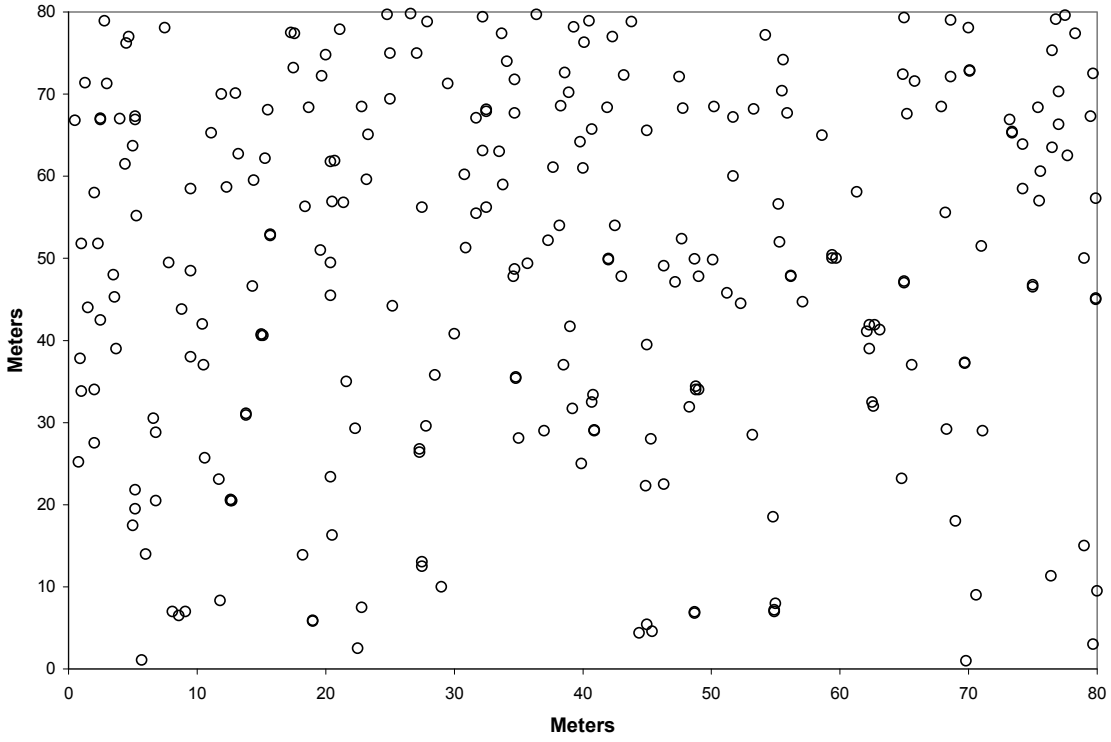

Figure 17. Stem map of Plot 527 (all trees with DBH ≥ 10cm).

indicated that clumps of trees were distributed throughout the plots, with clusters of up to four stems being present. This clustering is least apparent in Plot 218 and most apparent in Plot 527, which is dominated by yellow birch (*Betula alleghaniensis* Britton), northern red oak (*Quercus rubra* L.), black birch (*Betula lenta* L.), and basswood (*Tilia americana* L.).. Also apparent in Plot 527 is the presence of spatial inhomogeneity, where tree density is much lower in the lower right-hand corner of the plot than elsewhere.

Spatial structure of the Coweeta data was assessed by Ripley's *K*- and *L*-functions, where positive values of $L(r) - r$ indicate clustering, negative values indicate regularity in the pattern, and a plot of complete spatial randomness will equal zero for all distances *r*. The *L*-function was calculated by the R-language package Spatstat (Baddeley and Turner 2005) and plotted for distances up to 20 meters for each of the five Coweeta plots. For each plot, 1000 simulations of the homogeneous Poisson process were generated to form a 95% confidence envelope, which represents the 2.5th and 97.5th percentiles of the simulated *L*-functions (Fig. 18 – 22). The five Coweeta plots showed considerable differences in spatial structure as assessed by Ripley's *L*-function. While Plot 218 showed a spatially random structure, the other plots all had strong clustering within 3m of distance, with Plot 118, 427, and 527 showing the strongest clustering. In addition to this short range clustering, Plots 427 and 527 showed clustering at distances greater than 10 and 5 m respectively. In contrast to other plots, Plot 318 (Fig. 20) simultaneously shows clustering (from 0 – 2m) and over-dispersion of points (from 3 – 7m), providing an example of a point pattern that is neither purely repulsive nor purely attractive.

Analysis of the spatial structure of the Coweeta data using the pair correlation

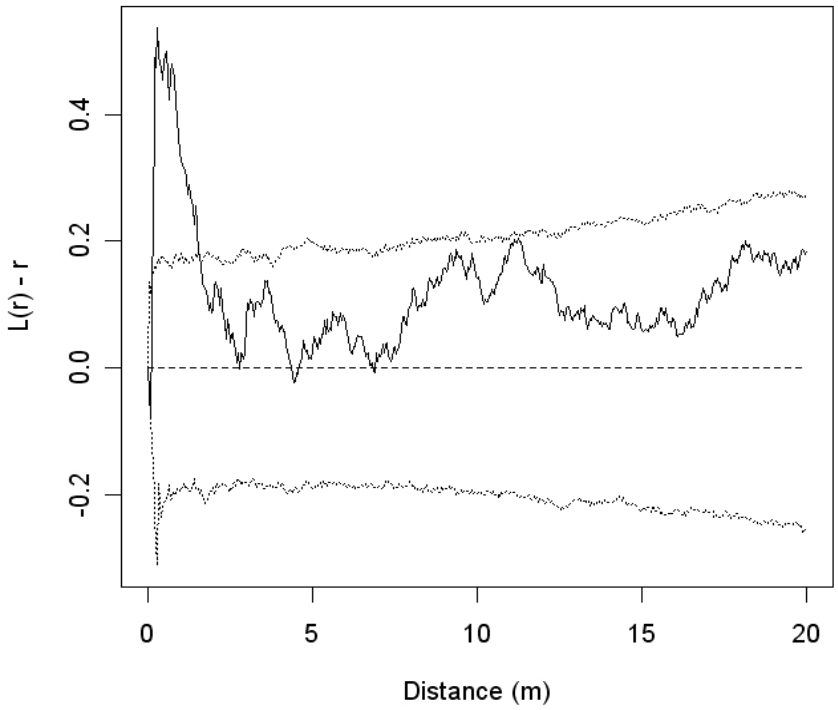

Figure 18. Observed $L(r) - r$ (solid line) and 95% Monte Carlo simulation envelope for CSR (dashed lines) for Plot 118.

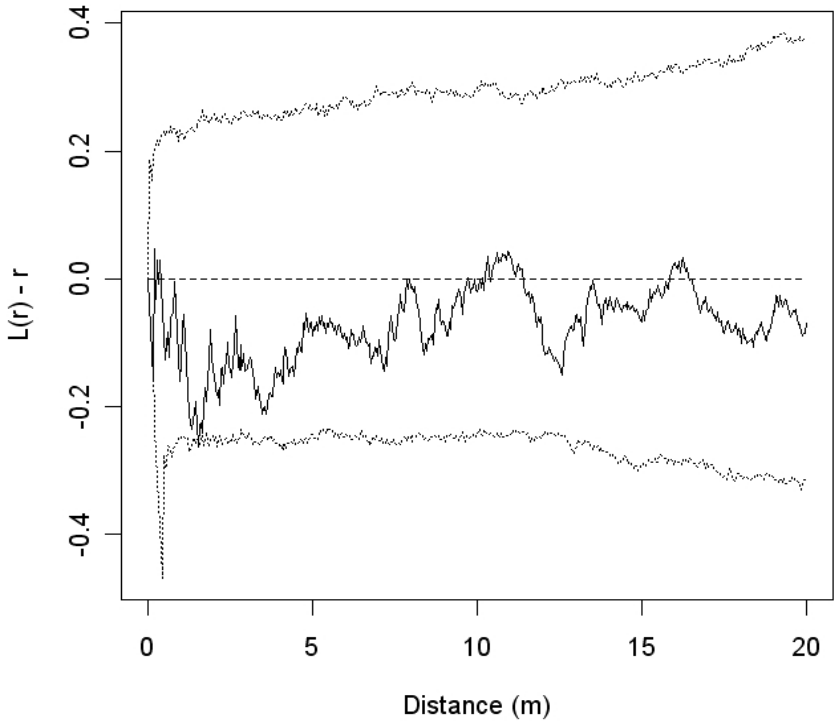

Figure 19. Observed $L(r) - r$ (solid line) and 95% Monte Carlo simulation envelope for CSR (dashed lines) for Plot 218.

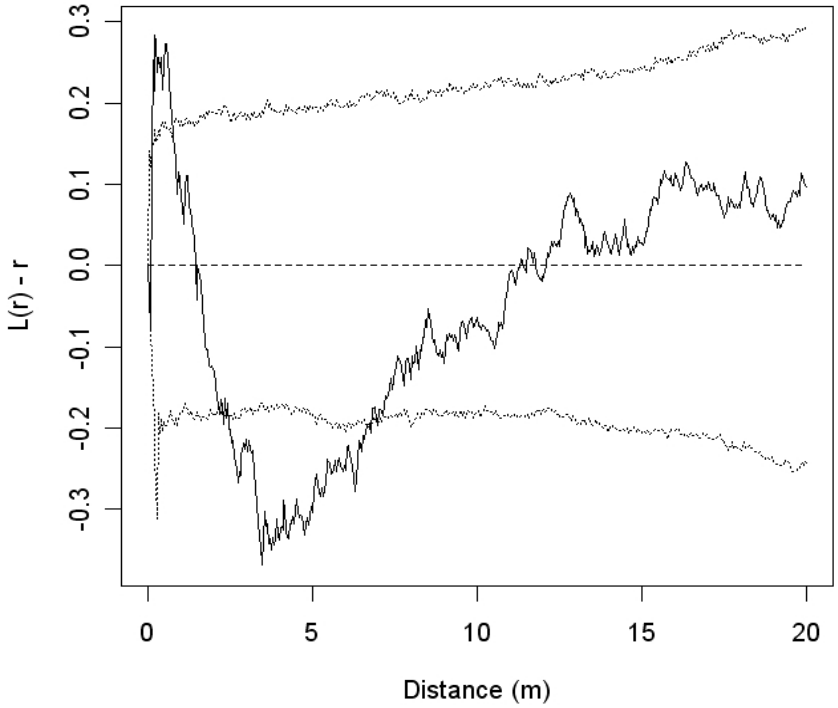

Figure 20. Observed $L(r) - r$ (solid line) and 95% Monte Carlo simulation envelope for CSR (dashed lines) for Plot 318.

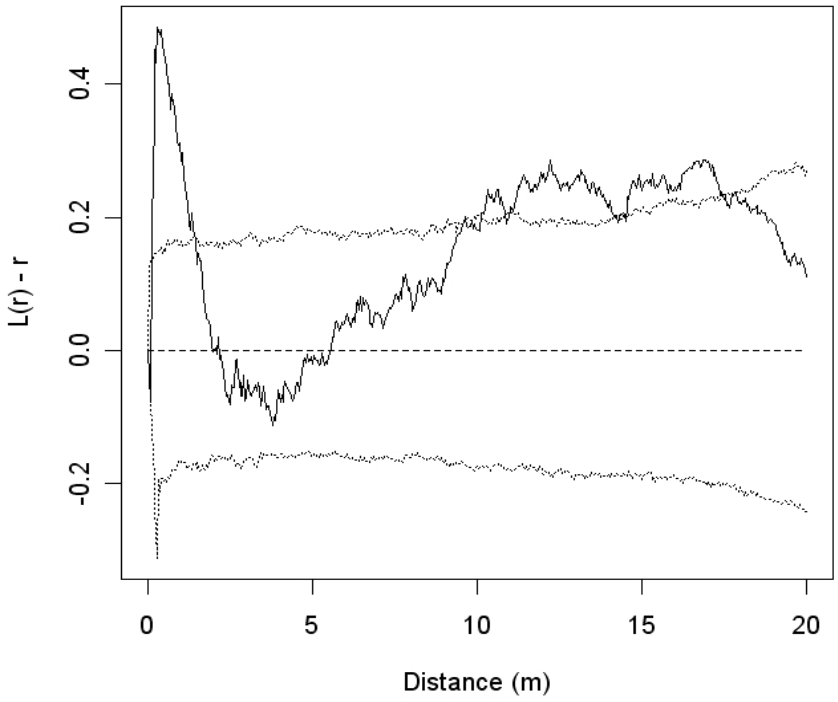

Figure 21. Observed $L(r) - r$ (solid line) and 95% Monte Carlo simulation envelope for CSR (dashed lines) for Plot 427.

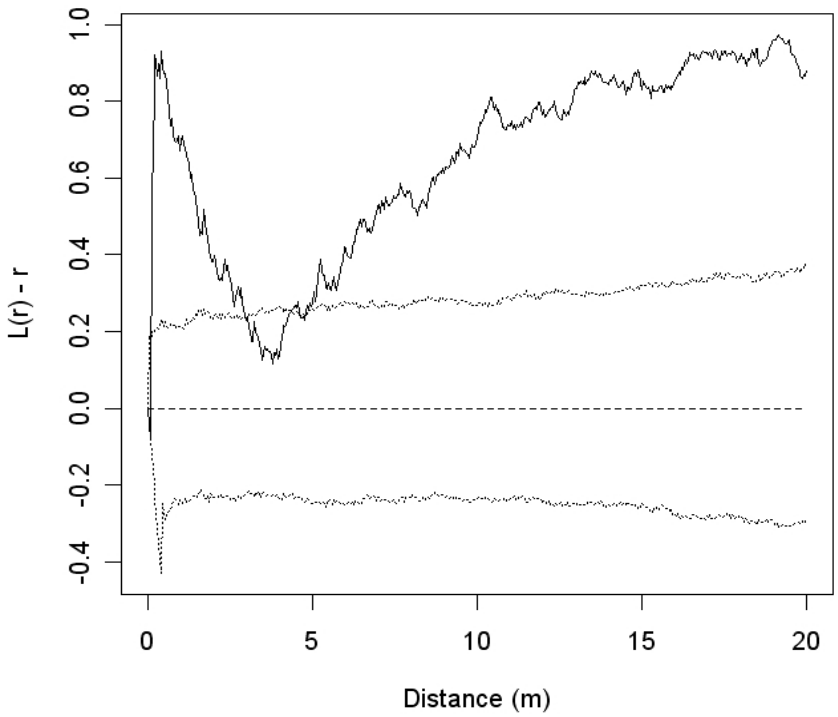

Figure 22. Observed $L(r) - r$ (solid line) and 95% Monte Carlo simulation envelope for CSR (dashed lines) for Plot 527.

function provided results (not all shown) very similar to those obtained from the *L*-function. Since values of the pair correlation function less than one indicate regularity and values greater than one at short distances indicate clustering (Stoyan and Penttinen 2000), the estimated pair correlation functions from the Coweeta data were compared against Monte Carlo envelopes of complete spatial randomness (CSR). As an example, the Monte Carlo test of the pair correlation function for Plot 318 indicates the presence of strong clustering at distances less than one meter and the presence of spatial regularity from one to four meters (Fig. 23).

Up to this point, the spatial analysis of the Coweeta data assessed the pattern

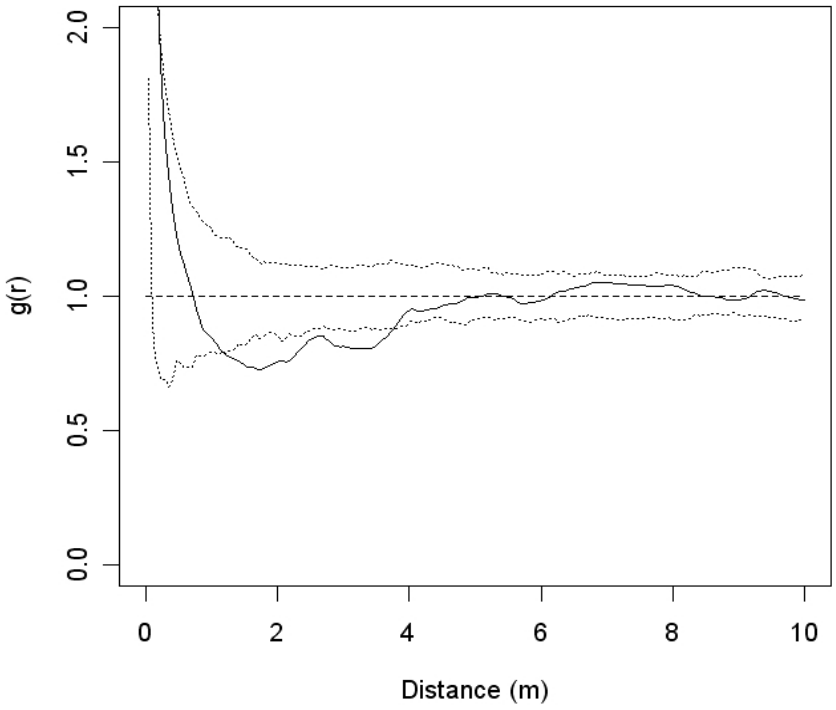

Figure 23. Observed pair correlation (solid line) and 90% Monte Carlo simulation envelope for CSR (dashed lines) for Plot 318.

among all trees with DBH ≥ 10 cm. Trees however are three dimensional objects which spatially stratify themselves by height as well as their location on the ground surface. The pattern of dominant and co-dominant trees may be more regular due to the size and position of their crown in the upper canopy layers. Intermediate and suppressed trees can exist much closer to dominant trees by locating their crowns *below* the crowns of the dominants and co-dominants. It was hypothesized that clustering at small distances (< 2m) is the result of stump sprouting in species that predominantly exist only in the understory or in otherwise suppressed stems of forest canopy trees. To examine this possibility, the Coweeta data were screened at increasing diameter limits of 10cm, 14cm, 18cm, 22cm, 26cm, and 30cm. In the absence of direct crown information such as crown

class, screening by tree diameter restricts the pattern to only the larger trees, where tree diameter is assumed to be positively correlated with tree height and crown size. For each diameter limit and Coweeta plot, Ripley's *L*-function was computed and plotted, again using the isotropic edge-effect correction (Figs. A1 – A30). For Plots 118 and 427, the *L*-function plots reveal a decrease in the magnitude of this short range clustering phenomenon. In Plot 118, for example, the significant clustering at distances less than 2m is weakened to the point that by the 30cm DBH limit, the pattern shows random interaction for all distances less than 15m. The *L*-function graphs for plots 318 and 427 also indicate a weakening of the short-range clustering phenomenon, whereas in Plot 527 significant clustering at distances less than 2m persists across the diameter limits. In the case of Plot 318, the clustering is present, but not significant for the 18, 22, and 26cm DBH limits. At the 30cm DBH limit, clustering becomes significant once more (Fig. A18). On the whole, short distance clustering is more evident when considering smaller diameter trees, but it still persists in some of the Coweeta datasets when screened to include only large diameter trees. Significant overdispersion in the spatial pattern of trees at Coweeta is evident at DBH limits ≥ 22cm for Plots 218, 427 and 527. In Plot 218, evidence of overdispersion becomes apparent at distances less than 5m once the DBH limit is at least 14cm. Plots 427 and 527 show increased repulsiveness at distances greater than 5m when the DBH limit exceeded 14cm. Significant overdispersion for plots 427 and 527 occurred at distances from 3 – 10m and 4 – 6m, respectively, when the DBH limit ≥ 22cm. In contrast, Plot 118 showed significant overdispersion only at the 18cm DBH limit and in Plot 318, overdispersion is present from 3 – 7m, but its significance weakens as the DBH limits increase.

4.2 Gibbs Point Process Simulation

4.21 Pairwise interaction models

For the simulation of Gibbs point processes, Geyer and Møller's (1994) Metropolis-Hastings algorithm was used with equal probability assigned to point births, deaths, and shifts. Unless otherwise noted, the simulation runs used 240,000 total iterations, with a burn-in of 40,000 iterations. Samples for Monte Carlo tests were drawn from the remaining 200,000 iterations by sub-sampling at a spacing of 200, giving 1,000 samples for each run. In most cases, the burn-in of Geyer and Møller's algorithm took only a few thousand iterations (as assessed by visual inspection of trace plots), but following Geyer and Møller (1994), 40,000 iterations were used as a conservative estimate. No attempt was made to assess an "optimal spacing" for sub-sampling, as described in Geyer (1999).

From the *L*-function analysis of the Coweeta gradient plots, Plot 318 provided the best example of short-distance clustering (< 2m) and longer range repulsion from 3 to 7m (Fig. 20). Since modeling this phenomenon is of key importance to the objectives pursued here, the proposed Gibbs point process models are simulated on 80x80 square "plots" to provide a comparison with the observed spatial structure in Plot 318. For each simulation, Geyer and Møller's algorithm is initialized by a sample from a homogeneous Poisson process with $n = 349$ points, the number of points observed on Plot 318. For the simulation of all Gibbs point process models, the intensity parameter was set to $\beta = \frac{n}{|F|} = \frac{349}{80 \cdot 80} \approx 0.0545$. As discussed in Sec. 3.71, the hard-core distance for all

models was set to 0.09m and the interaction distance for all models was set at 5m, unless otherwise noted.

Since simulation of Gibbs point processes using purely repulsive potentials is well established (Illian et al. 2008), parameters were chosen for simulation runs of the proposed models to represent either partial or complete spatial attraction. In the Fiksel model, the pair potential is attractive when $\theta_1 < 0$. Two example simulation runs are shown to illustrate how the model's behavior and the performance of the simulator can be affected strongly by the choice of values for θ_2. Both examples show an attractive potential with $\theta_1 = -1$, where Ex. 1 is chosen with $\theta_2 = 2.5$ and Ex. 2 is chosen with $\theta_2 = 1$ (Fig. 24). The execution of Geyer and Møller's algorithm for Ex. 1 produced trace plots that suggest convergence to the stationary distribution in less than 5,000 iterations corresponding to a point process with sampled configurations having about 375 points (Fig. B1 – B2). Sampled configurations show the presence of stump-sprout-like clusters of two or three trees, but also the presence of "chains" of 5 – 10 trees, such as one found at (22m, 75m) in Fig. 25. The 95% Monte Carlo simulation envelope of the *L*-function from this model confirms the attractiveness of the spatial pattern, which weakens to a Poisson-like structure after 10m (Fig. 26).

While the example above resulted in reasonable performance of the point pattern simulator and general agreement between the *L*-function plot and underlying theoretical model, the Fiksel model using a more attractive potential with $\theta_2 = 1$ was unable to converge to a stationary distribution. After 20,000 iterations, the number of points in the configurations generated from the Markov chain had exceeded 1,000 (Fig. B3) and the

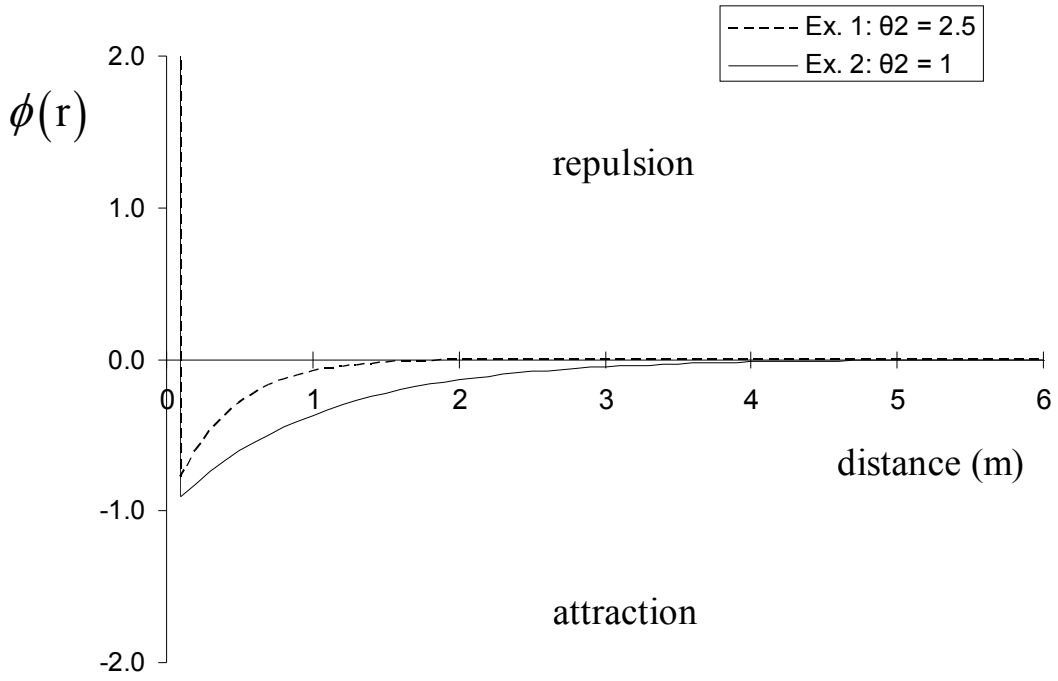

Figure 24. Pair potential plots of simulated Fiksel model. Ex. 1 corresponds to the parameter values $(\beta = 0.0545,\ \theta_1 = -1,\ \theta_2 = 2.5,\ R = 5)$; parameter values for Ex. 2 is $(\beta = 0.0545,\ \theta_1 = -1,\ \theta_2 = 1,\ R = 5)$

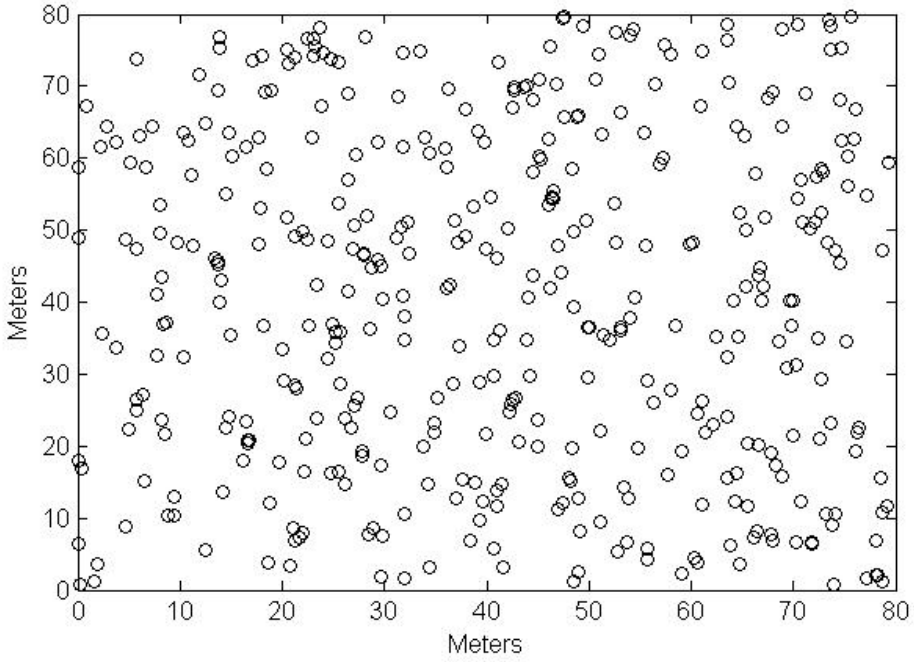

Figure 25. Sample configuration of Ex. 1 of the Fiksel model with parameter values $(\beta = 0.0545,\ \theta_1 = -1,\ \theta_2 = 2.5,\ R = 5)$.

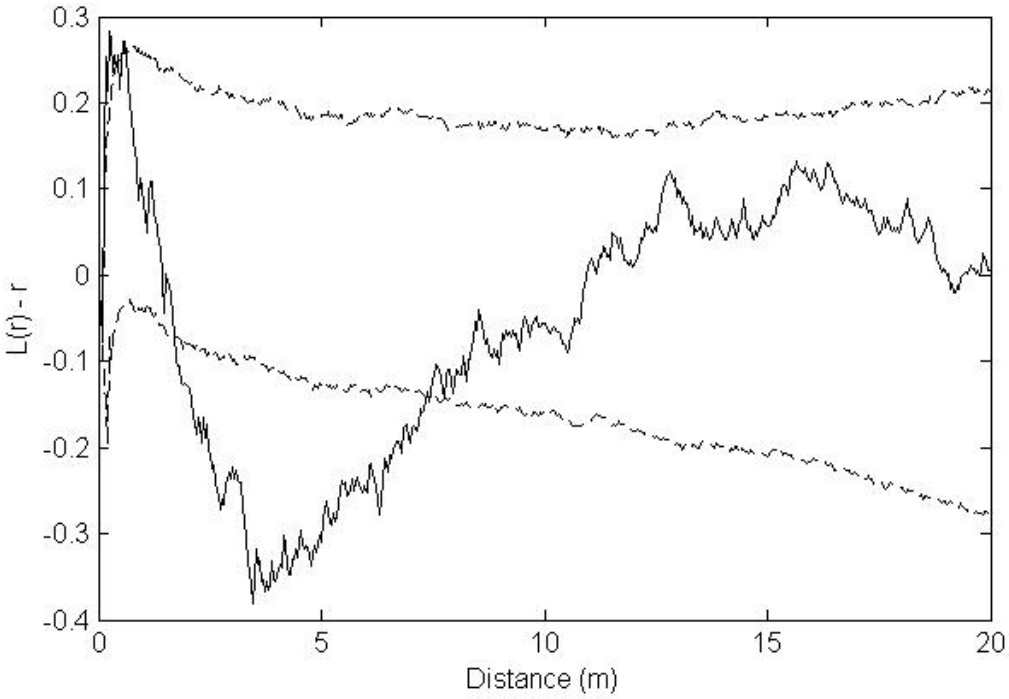

Figure 26. Monte Carlo envelope of *L*-function of Ex. 1 of the Fiksel model with parameter values $(\beta = 0.0545,\ \theta_1 = -1,\ \theta_2 = 2.5,\ R = 5)$ against the observed *L*-function of Coweeta Plot 318.

corresponding canonical statistic for θ_1 steadily decreased (Fig.B4). A sample obtained at the 20,000th iteration showed extreme clustering (Fig. 27), which is confirmed also by the *L*-function for this sample (Fig. 28). Beyond 20,000 iterations the Metropolis-Hastings algorithm was terminated as the number of interacting pairs of points generated from configurations in excess of 1,000 points was so large that the MATLAB code running the algorithm exhausted system memory on a PC equipped with 2GB of RAM.

Results such as Fig. 27 are unreasonable for use in modeling forest tree patterns and positions because the point process which generates it allows the number

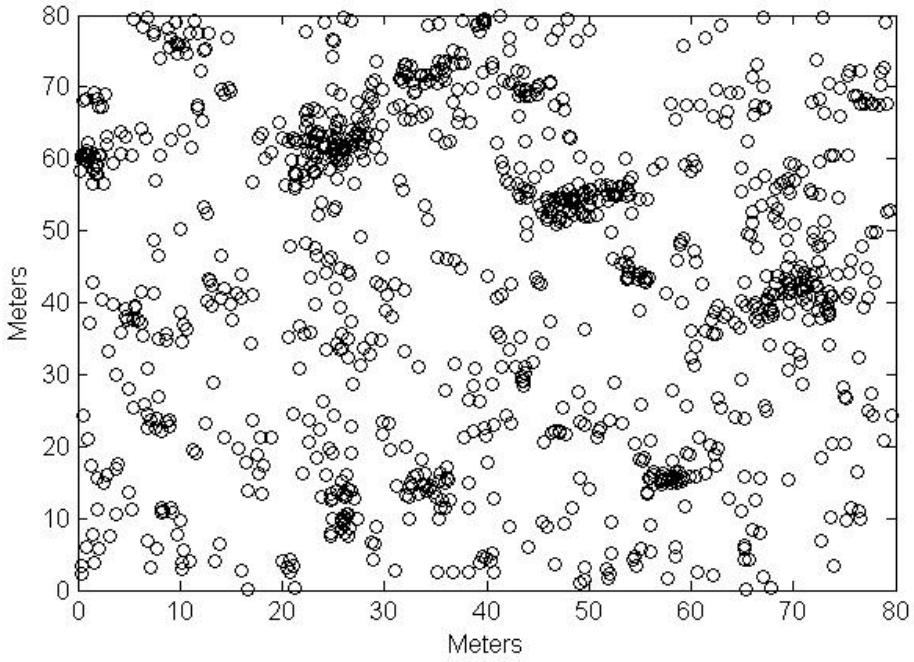

Figure 27. Sample configuration (n = 1,043 points) after 20,000 iterations of Ex. 2 of the Fiksel model with parameter values $(\beta = 0.0545,\ \theta_1 = -1,\ \theta_2 = 1,\ R = 5)$.

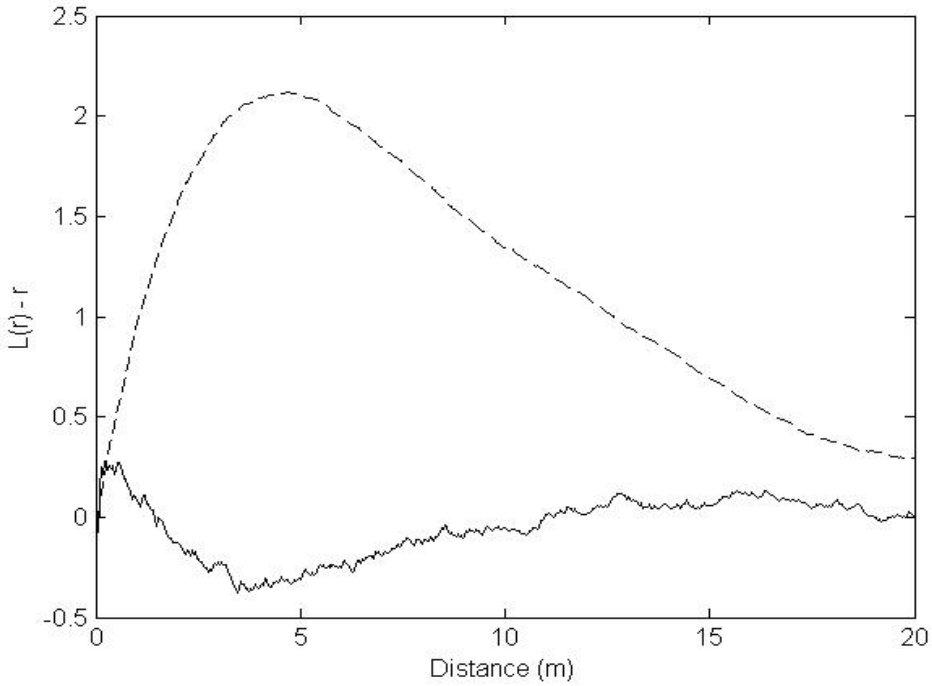

Figure 28. L-function of a sample configuration after 20,000 iterations of Ex. 2 of the Fiksel model with parameter values $(\beta = 0.0545,\ \theta_1 = -1,\ \theta_2 = 1,\ R = 5)$ against the observed L-function of Coweeta Plot 318.

of points of the process to be excessively large. Tree density for forests like those in Coweeta may be known or reasonably approximated through practical sampling and estimation procedures (Packard and Radtke 2007). If sample estimates of tree density were available, then the Gibbs process could be specified by conditioning on a specific number of points for the desired plot size. Simulating a Gibbs point process conditional on the number of points is straightforward in the context of Geyer and Møller's Metropolis-Hastings algorithm. By choosing the probability of a point shift to be one, Eq. 120 implies that the transition kernel of Geyer and Møller's algorithm is equivalent to the transition kernel of the Markov chain originally published by Metropolis et al. (1953). With this in mind, the attractive Fiksel model with $\theta_2 = 1$ was re-simulated with n fixed at 349 points, corresponding to the point density of Plot 318. Again, samples from the chain showed extreme clustering; after 150,000 iterations, where nearly all of the 349 points were located in one tight cluster located approximately at the coordinates (45m, 70m) (Fig. 29). The trace plot of the pseudo-canonical statistic (Fig. B5) showed a steady decrease across the 150,000 iteration run of the Metropolis algorithm. The behavior of simulations from the Fiksel model with ($\beta = 0.0545$, $\theta_1 = -1$, $\theta_2 = 1$, $R = 5$) provides an example of what Geyer (1999) termed the "one-clump process". By studying the behavior of the conditional Strauss process at the extremes of the pair potential's parameter values, Geyer (1999, p. 98) found that the conditional Strauss process, when specified to be strongly attractive, resulted in a process where all realizations resulted in sample configurations where all points were located in a ball with radius equal to the model's interaction distance.

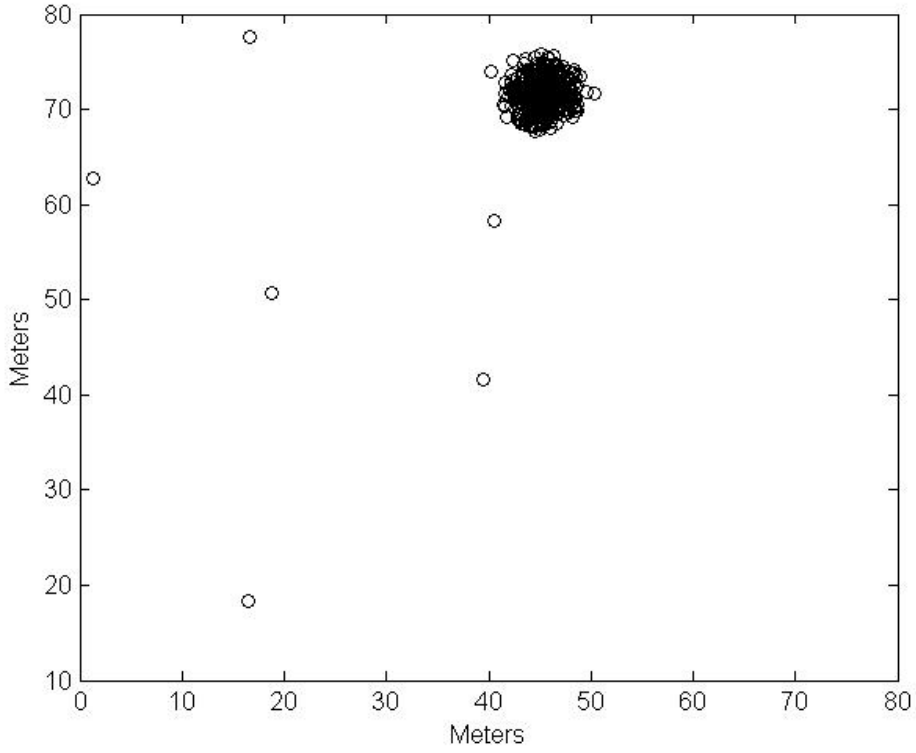

Figure 29. Sample configuration after 150,000 iterations of Ex. 2 of the Fiksel model with parameter values $(\beta = 0.0545,\ \theta_1 = -1,\ \theta_2 = 1,\ R = 5)$ and fixed n = 349.

Because the pair potential of the Fiksel model is essentially an exponential curve, the pair potential can either be repulsive, or attractive, but not both at the same time (c.f. Fig. 4). Since it was desired to model both localized attraction from stump sprouts and repulsion from intertree competition, the hard-core modified exponential (HCME) and hard-core serpentine (HCSP) models were proposed with pair potentials that can be attractive and repulsive simultaneously using smooth parametric functions. Both of these models were simulated using Geyer and Møller's Metropolis-Hastings algorithm using parameter values that gave partially attractive as well as totally attractive pair potential

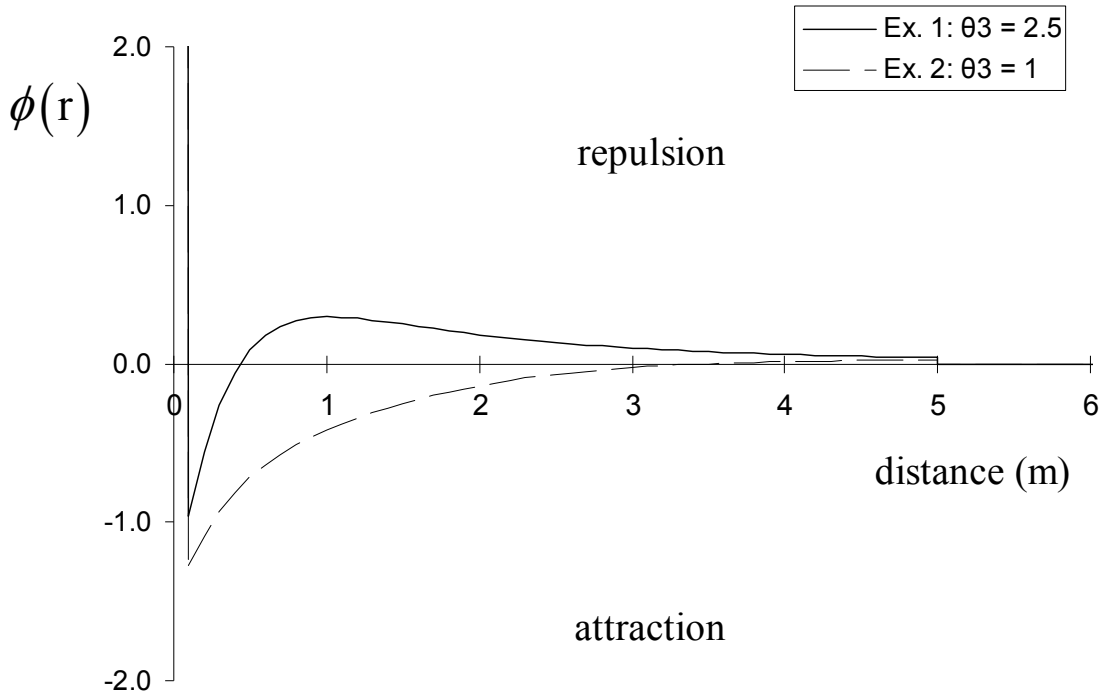

Figure 30. Pair potential plots of simulated Hard-Core Modified Exponential model. Ex. 1 corresponds to $(\beta = 0.0545,\ \theta_1 = 1,\ \theta_2 = -2.5,\ \theta_3 = 2.5,\ R = 5)$; Ex. 2 corresponds to $(\beta = 0.0545,\ \theta_1 = 1,\ \theta_2 = -2.5,\ \theta_3 = 1,\ R = 5)$

plots. For the HCME model, two examples are chosen corresponding to the parameter values ($\beta = 0.0545$, $\theta_1 = 1$, $\theta_2 = -2.5$, $\theta_3 = 2.5$, $R = 5$) and ($\beta = 0.0545$, $\theta_1 = 1$, $\theta_2 = -2.5$, $\theta_3 = 1$, $R = 5$). The first example, with $\theta_3 = 2.5$, corresponds to a potential that is attractive to a distance of 0.5m, becoming increasingly repulsive to 1m, with decreasing repulsion up to the interaction distance of 5m. The second example, with $\theta_3 = 1$, corresponds to a potential that is attractive to about 3.5m, being repulsive from 3m to 5m. (Fig. 30). Just as in the case of the Fiksel model, the two examples show widely varying

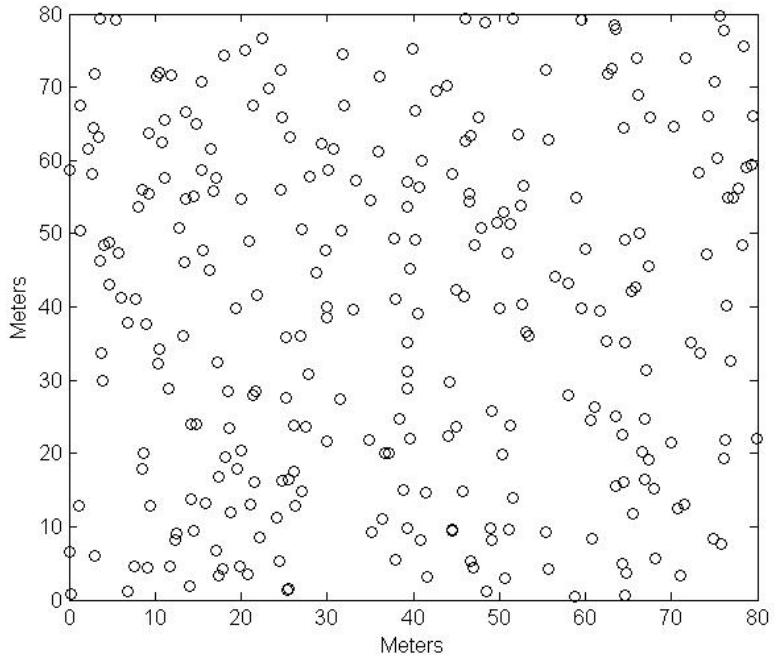

Figure 31. Sample configuration of Ex. 1 of the HCME model with parameter values $(\beta = 0.0545,\ \theta_1 = 1,\ \theta_2 = -2.5,\ \theta_3 = 2.5,\ R = 5)$.

behavior. Trace plots from the first example with $\theta_3 = 2.5$ suggest that the Markov chain has reached a stationary distribution with about 260 points (Figs. B6-B8). Samples from the HCME model with these parameter values do not show excessive attractiveness or any tendency to degenerate into the "one-clump" process (Fig. 31). Upon visual inspection, simulated point patterns appear to be repulsive, with a few clusters of two or three points dispersed throughout the plot (Fig. 31). A 95% Monte Carlo envelope of the *L*-function confirms this, and shows attractiveness at distances less than two meters and repulsion from approximately two to ten meters that closely mimics the *L*-function from Coweeta Plot 318 (Fig. 32). When the HCME model is simulated with $\theta_3 = 1$, the model degenerates quickly into excessively attractive patterns (Fig. 33). The number of points

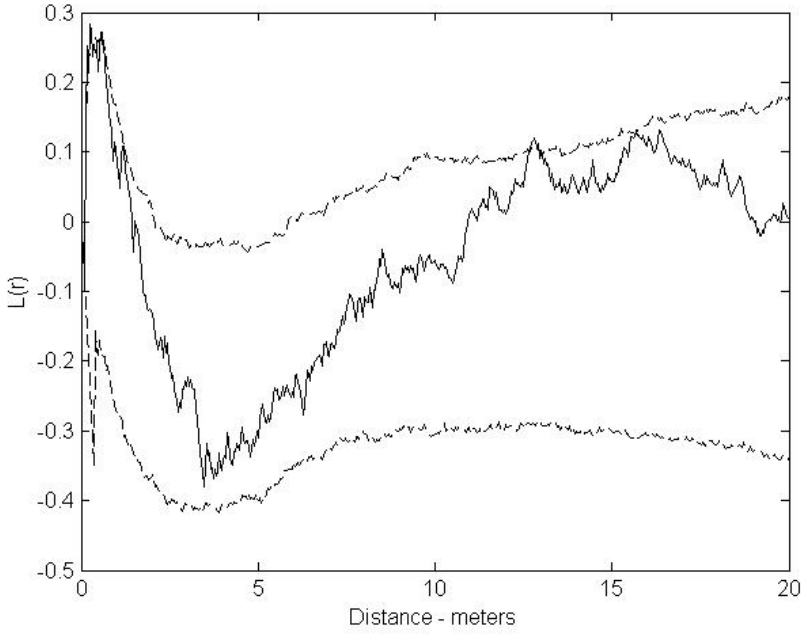

Figure 32. Monte Carlo envelope of *L*-function of Ex. 1 of the HCME model with parameter values $(\beta = 0.0545,\ \theta_1 = 1,\ \theta_2 = -2.5,\ \theta_3 = 2.5,\ R = 5)$ against observed *L*-function of Coweeta Plot 318.

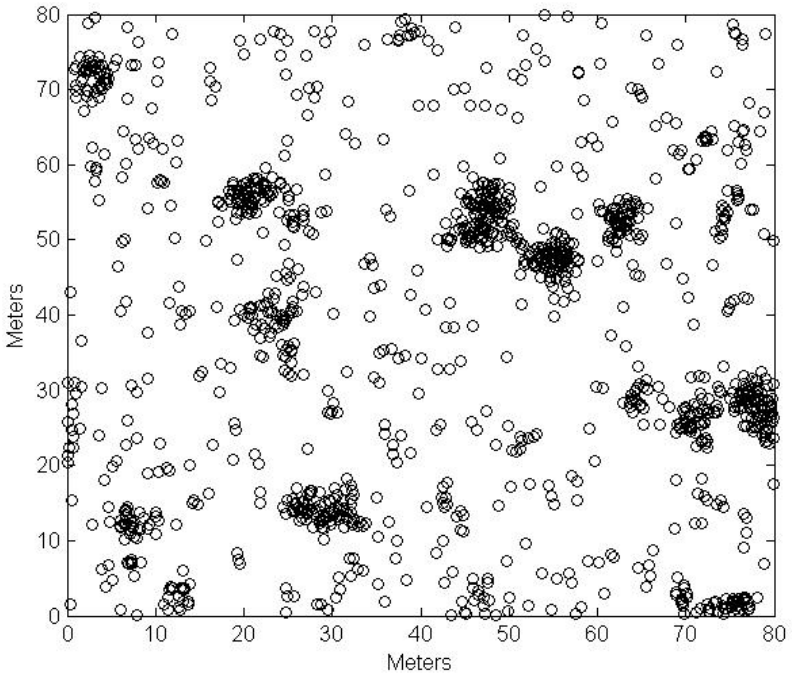

Figure 33. Sample configuration of Ex. 2 of the HCME model with parameter values $(\beta = 0.0545,\ \theta_1 = 1,\ \theta_2 = -2.5,\ \theta_3 = 1,\ R = 5)$ after 40,000 iterations.

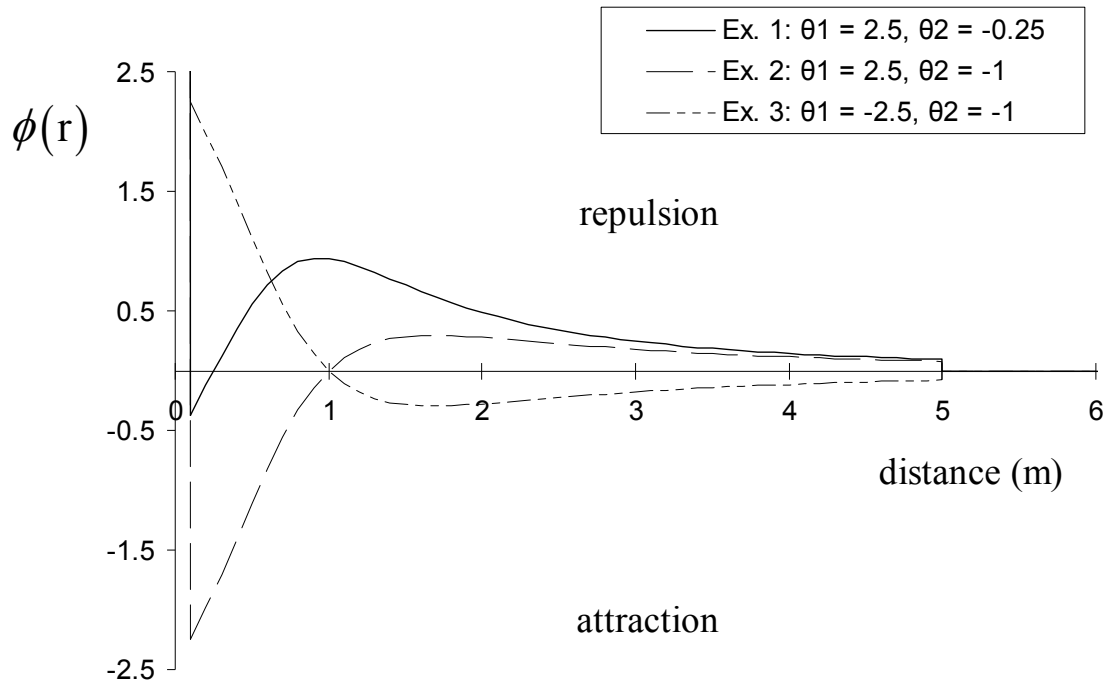

Figure 34. Pair potential plots of simulated HCSP model. Ex. 1 corresponds to the parameter values ($\beta = 0.0545$, $\theta_1 = 2.5$, $\theta_2 = -0.25$, $R = 5$); parameter values for Ex. 2 is ($\beta = 0.0545$, $\theta_1 = 2.5$, $\theta_2 = -1$, $R = 5$); parameter values for Ex. 3 is ($\beta = 0.0545$, $\theta_1 = -2.5$, $\theta_2 = -1$, $R = 5$).

exceeded 1,200 after 40,000 iterations, beyond which the algorithm failed for lack of sufficient computer memory (Fig. B9).

The apparent lack of stability of the Fiksel and HCME models is replicated as well in the hard-core serpentine (HCSP) model. As an illustration, three examples are chosen that present mixtures of attraction and repulsion. The first with parameter values ($\beta = 0.0545$, $\theta_1 = 2.5$, $\theta_2 = -0.25$, $R = 5$), gives a potential with attraction at distances less than 0.25m, and repulsion otherwise through the potential's range of interaction. The second example, with θ_2 decreased to -1.0, shows spatial attraction to distances up to 1m, with repulsion from 1m to 5m (Fig. 34). As an example of the flexibility of the HCSP

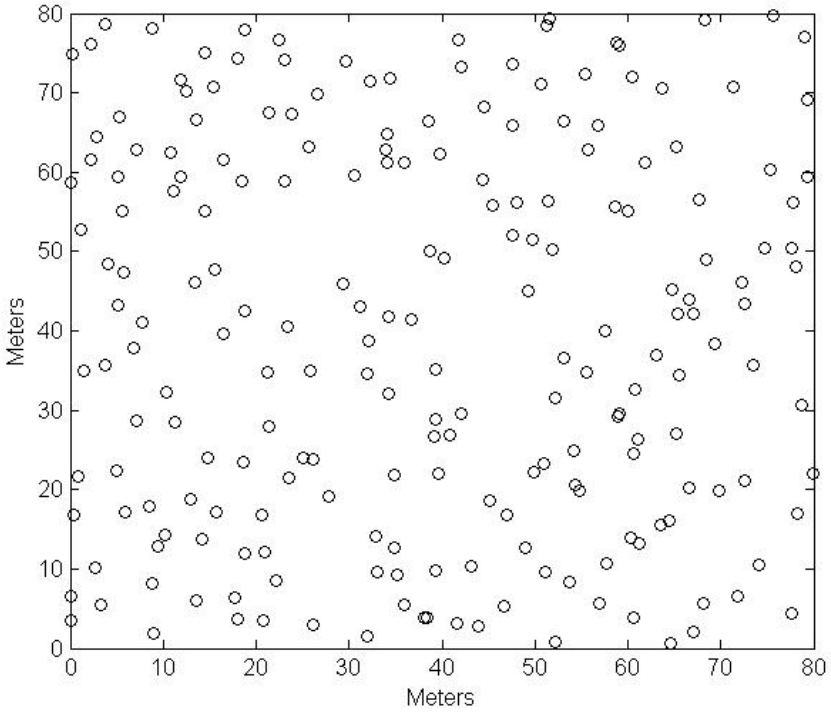

Figure 35. Sample configuration of Ex. 1 of the HCSP model with parameter values $(\beta = 0.0545,\ \theta_1 = 2.5,\ \theta_2 = -0.25,\ R = 5)$.

potential, the third example demonstrates short range repulsion (up to 1m) combined with spatial attraction from 1m to 5m. While this spatial structure is not consistent with Coweeta Plot 318, it is included here since some studies (Degenhardt and Pofahl 2000) have attempted to model forest pattern using Gibbs processes with potentials that show short range repulsion combined with longer range attraction.

For the first example with $\theta_1 = 2.5$, and $\theta_2 = -0.25$, the Markov chain simulating the HCSP model appears to have reached the stationary distribution rapidly (Figs. B12 – B13). The trace plot of the number of points (Fig. B12) shows no indication of any instability in the model, with samples having around 200 points. Upon visual inspection, sample configurations (Fig. 35) appear to show overdispersion with a few stray clusters

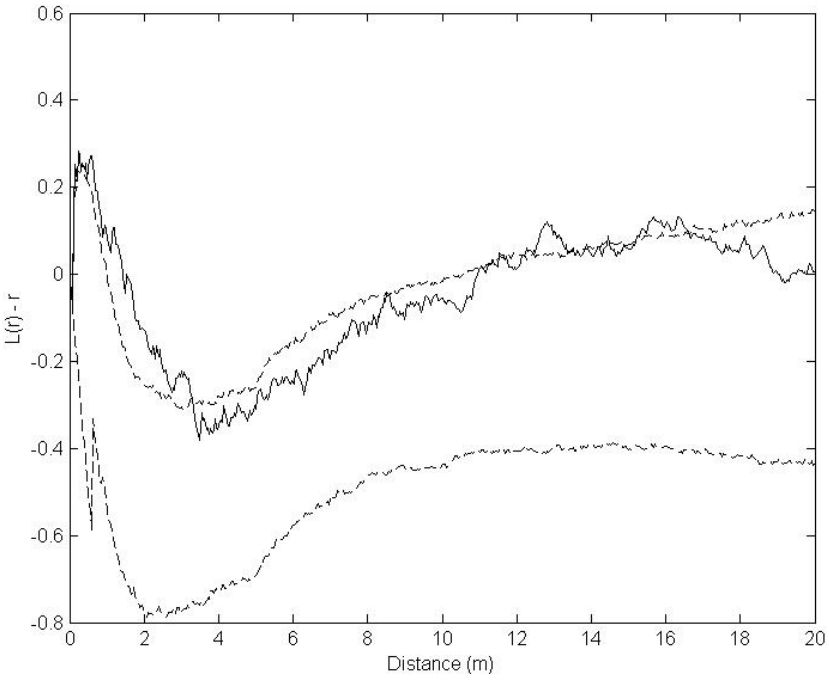

Figure 36. Monte Carlo envelope of *L*-function of Ex. 1 of the HCSP model with parameter values $(\beta = 0.0545,\ \theta_1 = 2.5,\ \theta_2 = -0.25,\ R = 5)$ against observed *L*-function of Coweeta Plot 318.

of two points.The Monte Carlo envelope of the *L*-function confirms a slight amount of clustering at distances less than one meter. Otherwise the HCSP model shows a repulsive spatial structure that overall shows less clustering than Coweeta Plot 318 (Fig. 36).

The second example, with $\theta_1 = 2.5$ and $\theta_2 = -1$, did not converge to a stationary distribution. Although the simulation run proceeded through the 240,000 iterations specified, trace plots indicate that neither the number of points nor the pseudo-canonical statistic are from samples of the stationary distribution of the chain (Fig. B14 – B15). Instead of degenerating into a "one-clump" process, a sample at the 240,000th iteration

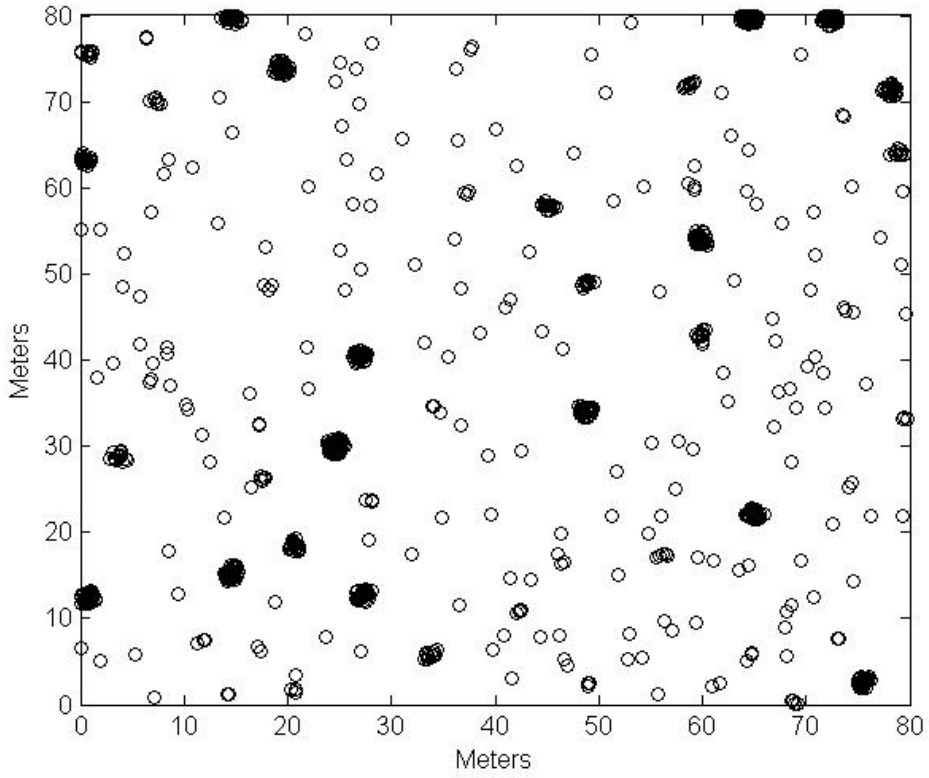

Figure 37. Sample configuration of Ex. 2 of the HCSP model with parameter values $(\beta = 0.0545,\ \theta_1 = 2.5,\ \theta_2 = -1,\ R = 5)$ after 240,000 iterations; n = 782 points.

suggests that the HCSP model is degenerating into a "multiple clump" process, where all points will be drawn into one of multiple clumps that appear randomly distributed across the plot (Fig. 37).

Attempts to stabilize the HCSP model by conditioning on the number of points also proved unsuccessful. With n = 349 fixed, the trace plot of the pseudo-canonical statistic (Fig. B16) showed a steady increase across the 240,000 iteration run. A sample

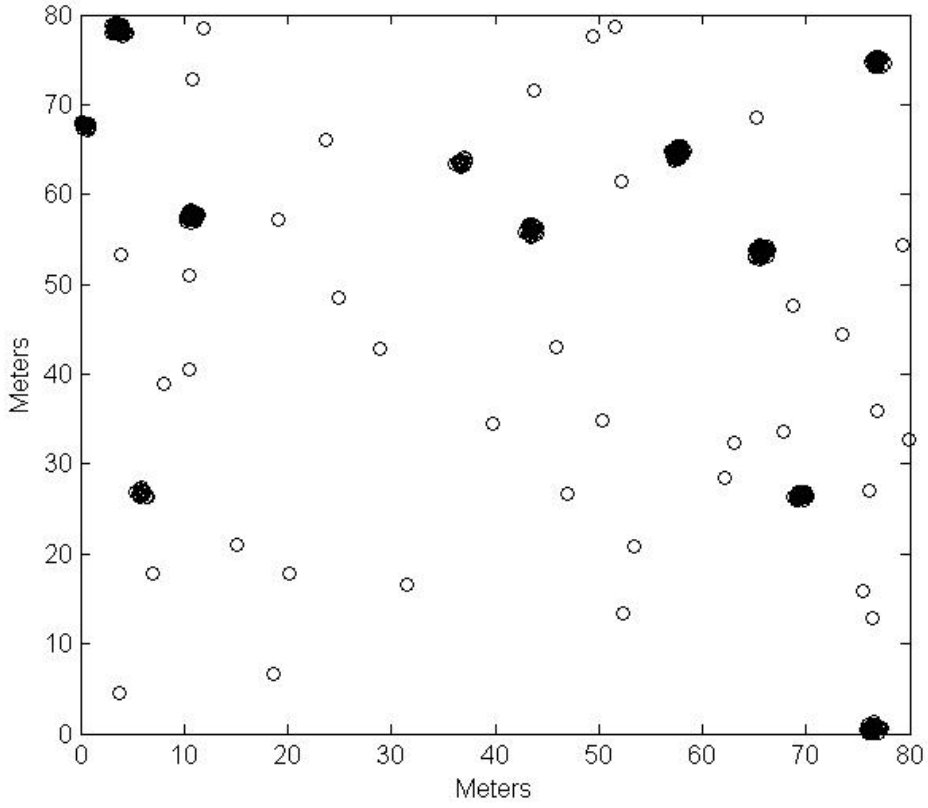

Figure 38. Sample configuration of Ex. 2 of the HCSP model with parameter values $(\beta = 0.0545,\ \theta_1 = 2.5,\ \theta_2 = -1,\ R = 5)$ after 240,000 iterations with fixed $n = 349$.

taken at the end of the run (Fig. 38) showed that even with n fixed, the HCSP samples continue to degenerate into samples from a "multiple clump" process, with the clumps located randomly throughout the plot.

Simulation of the third example of the HCSP process with $\theta_1 = -2.5$, $\theta_2 = -1$, corresponding to repulsion at distances up to 1m and attraction from 1 to 5m proved to be unstable as well. In less than 20,000 iterations, Geyer and Møller's algorithm had failed

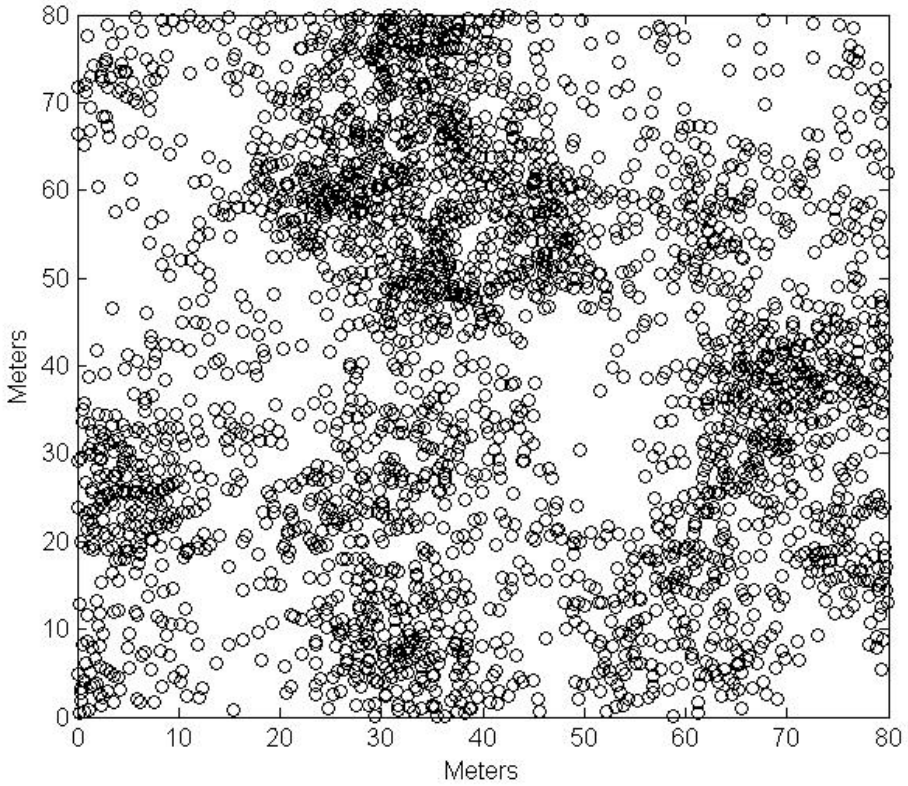

Figure 39. Sample configuration of Ex. 3 of the HCSP model with parameter values $(\beta = 0.0545,\ \theta_1 = -2.5,\ \theta_2 = -1,\ R = 5)$ after 16,000 iterations; n = 2,906 points.

for lack of sufficient computer memory. By 16,000 iterations the Markov chain was generating point patterns with almost 3,000 points (Fig. B17). Unlike the previous example, where sample configurations returned multiple tight clusters of points, the potential caused massive attraction that was quite diffuse, with ill-defined "clusters" spread out over 200 square meters of plot area or more (Fig. 39). Again, conditioning on the number of points, as was required in the partially attractive potentials of Degenhardt and Pofahl (2000), failed to prevent the instability of the HCSP model. After 50,000

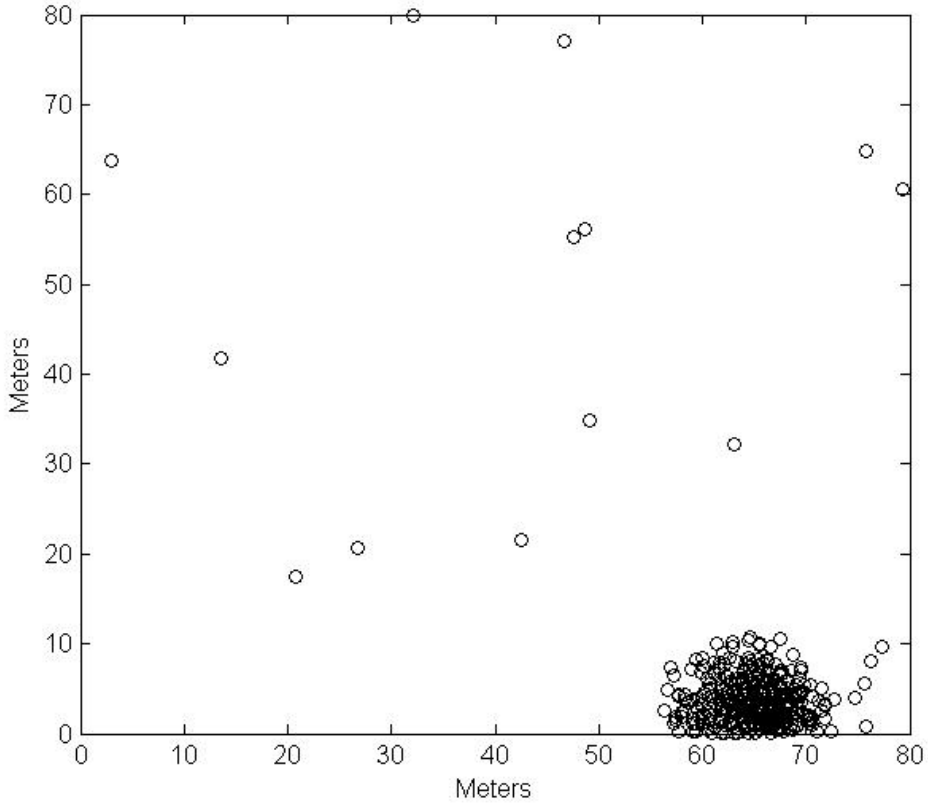

Figure 40. Sample configuration of Ex. 3 of the HCSP model with parameter values $(\beta = 0.0545,\ \theta_1 = -2.5,\ \theta_2 = -1,\ R = 5)$ after 50,000 iterations with fixed $n = 349$.

iterations, sampled configurations (Fig. 40) showed the behavior of a "one-clump" process, similar in appearance to that described by Geyer (1999). While the large cluster of points in Fig. 40 is more spread out than the tight single cluster of Fig. 29, the repulsiveness of the HCSP potential at distances less than 1m, failed to prevent the model from degenerating into a one clump process just as the purely attractive Fiksel model did.

A summary of the simulation results of all pairwise models tested highlights two key features of pairwise models with partial or purely attractive pair potentials. First

Table 4. Summary of simulation results for pairwise interaction processes.

Model	*Parameters* (θ_1, θ_2, θ_3)	*Point number*	*Potential at distances < 1m*	*Potential at distances > 1m*	*Outcome*
Fiksel	−1, 2.5	Random	Weak attraction	Random	Pairs, triples in repulsive pattern
Fiksel	−1, 1	Random	Weak attraction	Weak attraction	Failed, massive multiple clusters
Fiksel	−1, 1	Fixed	Weak attraction	Weak attraction	One clump
HCME	1, −2.5, 2.5	Random	Weak attraction	Weak repulsion	Pairs in repulsive pattern
HCME	1, −2.5, 1	Random	Weak attraction	Weak attraction and random	Failed, massive multiple clusters
HCSP	2.5, −0.25	Random	Weak repulsion	Weak repulsion	Pairs in repulsive pattern
HCSP	2.5, −1	Random	Strong attraction	Weak repulsion	Failed, multiple clumps
HCSP	2.5, −1	Fixed	Strong attraction	Weak repulsion	Multiple clumps
HCSP	−2.5, −1	Random	Strong repulsion	Weak attraction	Failed, massive diffuse clusters
HCSP	−2.5, −1	Fixed	Strong repulsion	Weak attraction	One clump

potentials that when plotted may appear to be weakly attractive at some distances, can in fact be extremely attractive (Table 4). This was true for both purely attractive potentials such as the Fiksel model and partially attractive and partially repulsive models like the HCME and HCSP, regardless whether the attraction reflected short range (< 1m) or long range behavior. Secondly, if a pair potential was sufficiently attractive, it would

degenerate into a one clump process, despite conditioning on the number of points of the process.

4.22 Triplets models

As a suggestion to address the inadequacy of existing Gibbs point process models to simulate attractive patterns, Geyer (1999) proposed two models that, in different ways, attempted to address this problem. One, called the "saturation process" attempted to limit the number of interacting points by introducing a parameter that puts an upper bound on the maximum contribution to a configuration that any particular point may have. The other introduced repulsive third order interactions to counteract attractive pairwise interactions in a "triplets process". Geyer's triplets process, along with two new formulations of triplets processes introduced here, are constructed to allow for either partially or completely attractive 2^{nd}-order interactions as means of simulating clustering of trees at very short distances, while using a 3^{rd}-order repulsive potential to prevent patterns from degenerating to one or more clumps of points.

Geyer's model is perhaps the simplest construction of a triplets process possible. Both the attractive 2^{nd}-order and repulsive 3^{rd}-order potentials are given by parameter values that keep the potentials constant across the range of interaction (see Sec. 3.32, Eqs. 97, 98). To provide an illustration of the behavior of Geyer's model, two levels of attraction, corresponding to $\theta_1 = -1$ and $\theta_1 = -2.5$ were selected along with two levels of repulsion, corresponding to $\theta_2 = 1$ and $\theta_2 = 2.5$ (Figs 40, 41). From this four examples of

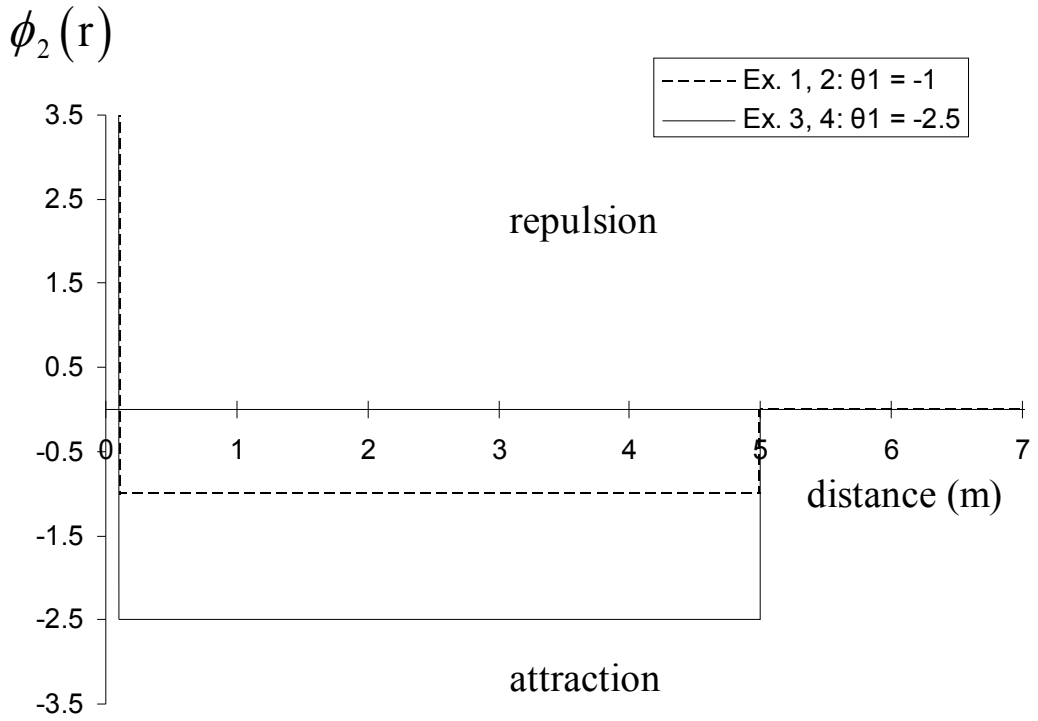

Figure 41. Pair potential plots of simulated Geyer triplets model. Ex. 1 and 2 correspond to $\theta_1 = -1$; Ex. 3 and 4 correspond to $\theta_1 = -2.5$.

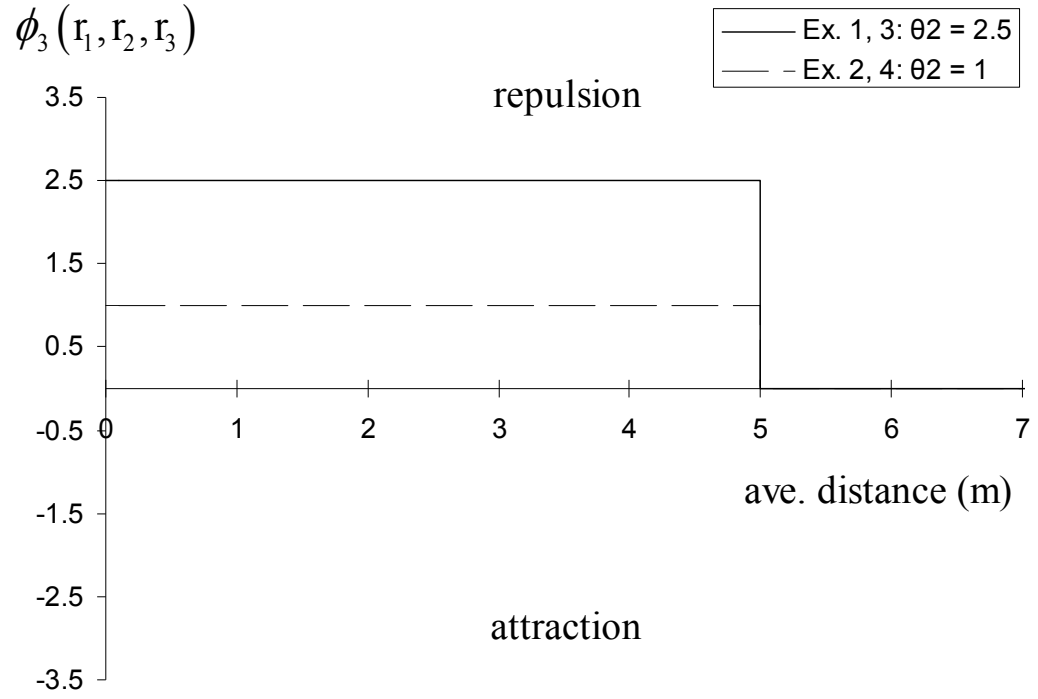

Figure 42. Third order potential plots of simulated Geyer triplets model. Ex. 1 and 3 correspond to $\theta_2 = 2.5$; Ex. 2 and 4 correspond to $\theta_2 = 1.0$.

Geyer's model are created to illustrate the effects of different degrees of 2nd-order attraction with different degrees of 3rd-order repulsion. As with the pairwise models, the parameter β was set to 0.0545, the hard-core distance was set to 0.09m, and the interaction distance was set to 5m.

The first example of Geyer's triplets model where $\theta_1 = -1.0$ and $\theta_2 = 2.5$, represents a combination of weak 2nd-order attraction and strong 3rd-order repulsion. Given the result from the previous example, one might expect that these parameter values will give point patterns that show strong point repulsion, which does in fact occur. Trace plots from the simulation (Figs. C1 – C3) show much faster convergence (10 – 20,000 iterations) and convergence to a point process with a density comparable to that of Plot 318. Despite the presence of a few close pairs of trees, the overall pattern of samples from this example show strong regularity in pattern (Fig. 43). Monte Carlo envelopes of the *L*-function indicate that Geyer's triplets process with parameter values $\theta_1 = -1.0$ and $\theta_2 = 2.5$ give configurations with strong repulsion throughout the range of interaction (Fig. 44).

The second example of Geyer's triplets process, where $\theta_1 = -1.0$ and $\theta_2 = 1.0$, represents a combination of weak 2nd order attraction and weak 3rd order repulsion. The Metropolis-Hastings sampler converges quickly to a point process with just under 500 points (Fig. C4 – C6). Sampled configurations show repulsive interactions and Monte Carlo envelopes of the *L*-function show repulsion throughout the range of interaction, though not quite as large as in the first example with $\theta_1 = -1.0$ and $\theta_2 = 2.5$ (Figs. 44, 45).

The results for the third example of Geyer's triplets process, where $\theta_1 = -2.5$ and

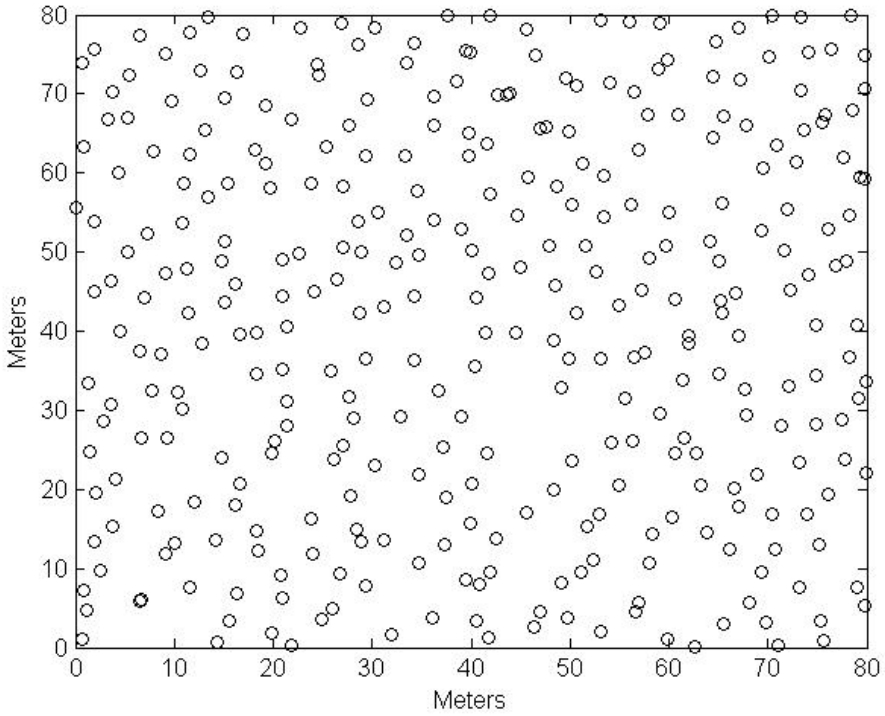

Figure 43. Sample configuration of Ex. 1 of the Geyer triplets model with parameter values $(\beta = 0.0545,\ \theta_1 = -1,\ \theta_2 = 2.5,\ R = 5)$.

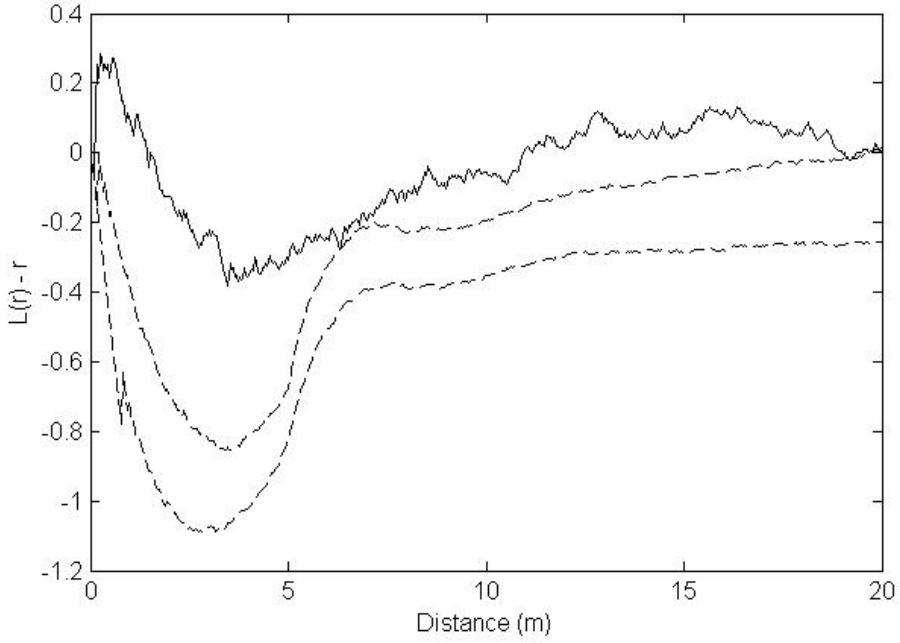

Figure 44. Monte Carlo envelope of *L*-function of Ex. 1 of the Geyer triplets model with parameter values $(\beta = 0.0545,\ \theta_1 = -1,\ \theta_2 = 2.5,\ R = 5)$ against observed *L*-function of Coweeta Plot 318.

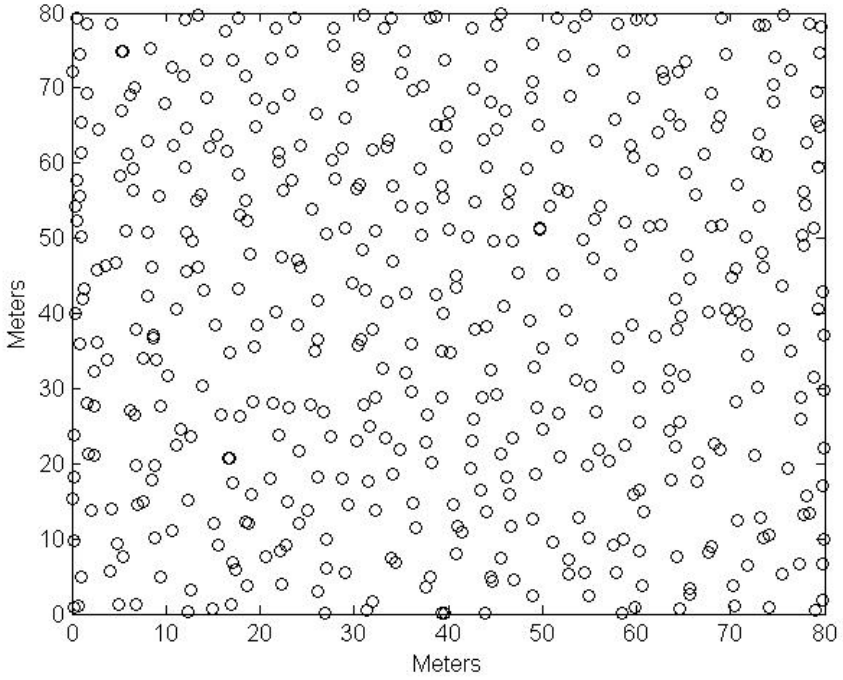

Figure 45. Sample configuration of Ex. 2 of the Geyer triplets model with parameter values $(\beta = 0.0545,\ \theta_1 = -1,\ \theta_2 = 1,\ R = 5)$

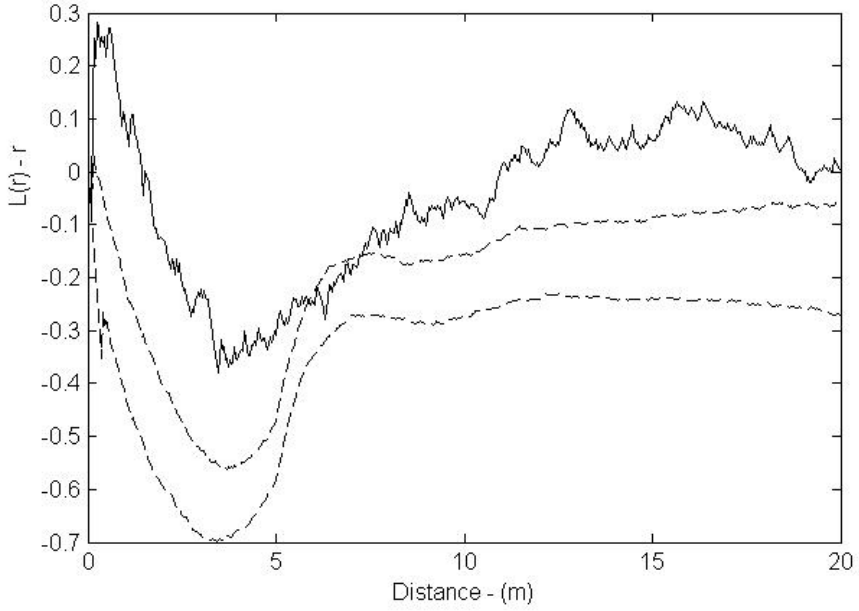

Figure 46. Monte Carlo envelope of *L*-function of Ex. 2 of the Geyer triplets model with parameter values $(\beta = 0.0545,\ \theta_1 = -1,\ \theta_2 = 1,\ R = 5)$ against observed *L*-function of Coweeta Plot 318.

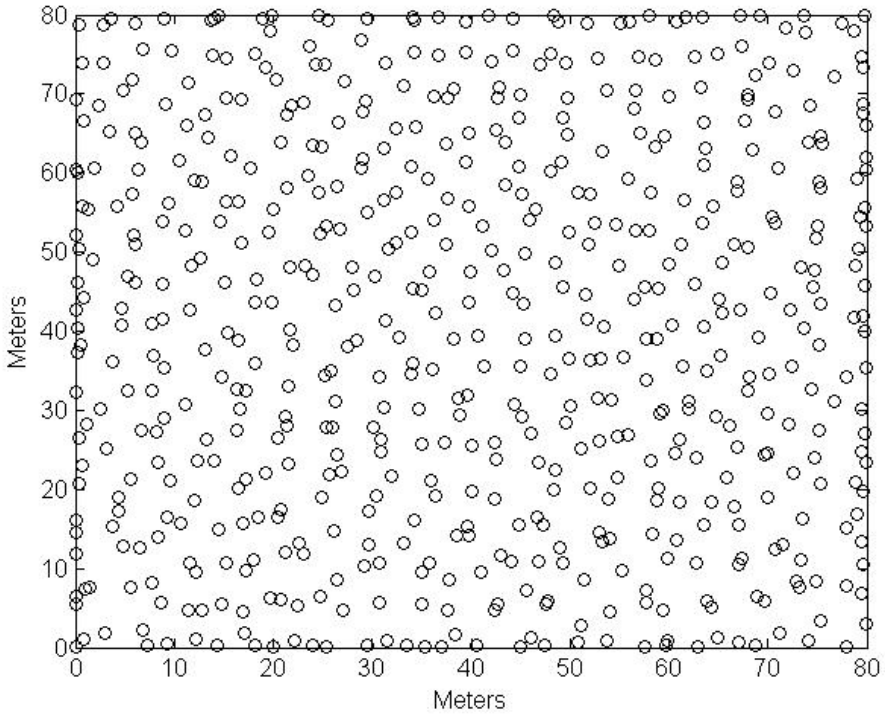

Figure 47. Sample configuration of Ex. 3 of the Geyer triplets model with parameter values $(\beta = 0.0545,\ \theta_1 = -2.5,\ \theta_2 = 2.5,\ R = 5)$

$\theta_2 = 2.5$ illustrate a combination of strong 2nd order attraction and strong 3rd order repulsion, are similar to the results from the previous two examples. Simulation with Geyer and Møller's algorithm resulted in apparent convergence after 30 – 40,000 iterations to a point process with about 580 points (Figs. C7 – C9). Sampled configurations (Fig. 47) reflect the higher point density (the number of points on Plot 318 is 349) in a fairly repulsive pattern. The Monte Carlo test of the *L*-function indicates that while Geyer's triplets model contains highly attractive second-order interactions, the patterns generated from the model using the parameter values θ_1 = -2.5 and θ_2 = 2.5 show

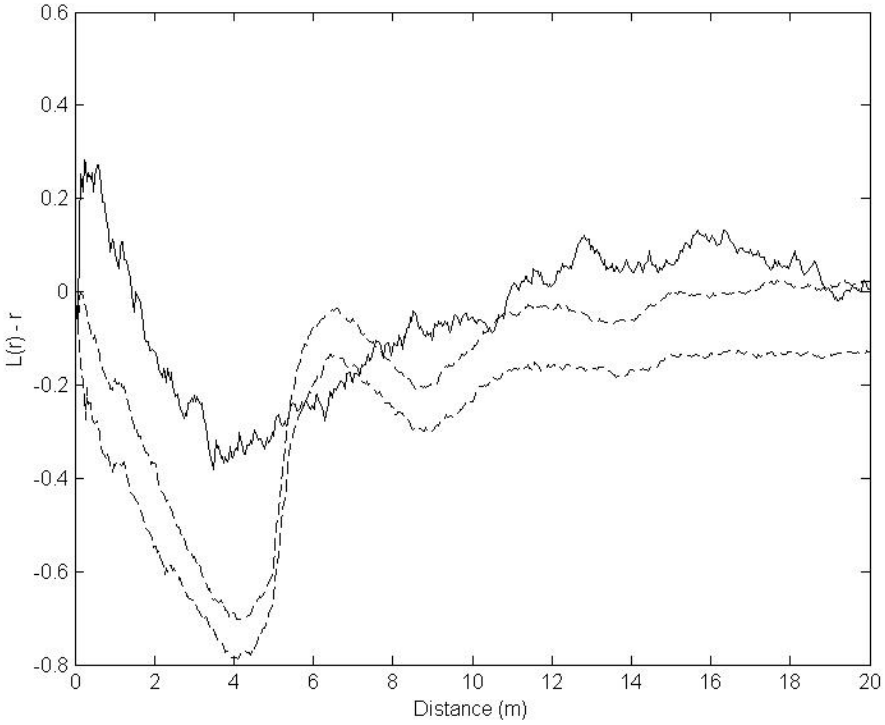

Figure 48. Monte Carlo envelope of *L*-function of Ex. 3 of the Geyer triplets model with parameter values $(\beta = 0.0545,\ \theta_1 = -2.5,\ \theta_2 = 2.5,\ R = 5)$ against observed *L*-function of Coweeta Plot 318.

in fact strong point repulsion throughout the 5m range of interaction of the model (Fig. 48).

As the first three examples illustrate, Geyer's triplets model illustrates purely repulsive point behavior despite the specification of an attractive pair potential. In these three examples, the absolute value θ_2 was greater than or equal to the absolute value of θ_1 = -1.0. Now consider an example when this is not true, that is when θ_1 = -2.5 and θ_2 = 1.0. Simulation of the model takes up to 100,000 iterations to burn-in (Fig. C10 –

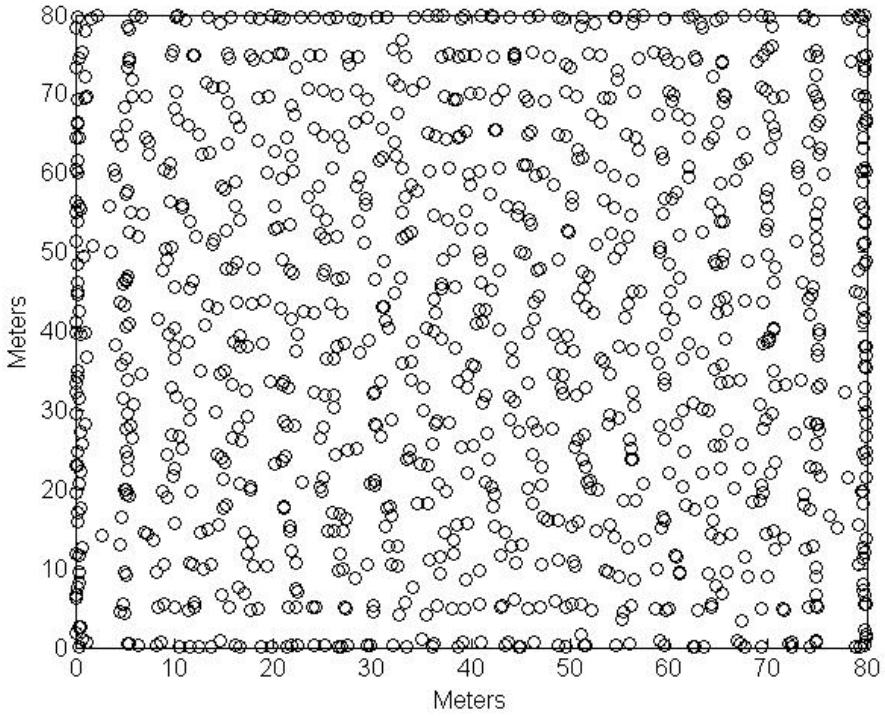

Figure 49. Sample configuration of Ex. 4 of the Geyer triplets model with parameter values $(\beta = 0.0545,\ \theta_1 = -2.5,\ \theta_2 = 1,\ R = 5)$.

C12) and reaches a stationary distribution corresponding to a point process with nearly 1,100 points. Sampled configurations show linear clusters or chains of points that are regularly spaced throughout the "plot" window (Fig. 49). The strong attraction of points has lead to a buildup of clusters along the edge of the plot, indicating that despite the use of isotropic edge correction factors in the 2nd- and 3rd-order potentials, edge effects remain. Because of the longer burn-in needed for this set of parameters, the desired Monte Carlo envelope using 1,000 samples at a spacing of 200 samples necessitated a

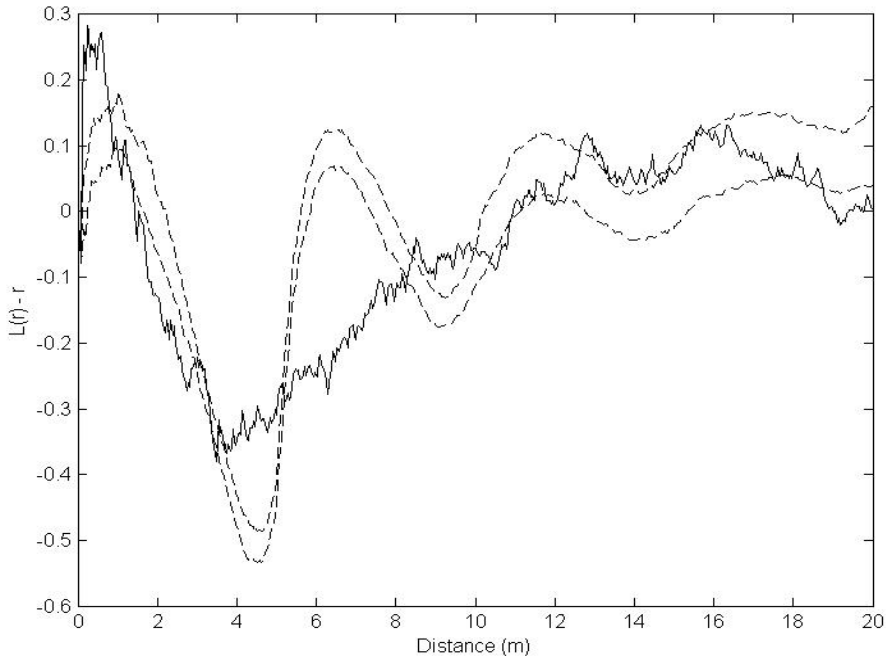

Figure 50. Monte Carlo envelope of *L*-function of Ex. 4 of the Geyer triplets model with parameter values $(\beta = 0.0545,\ \theta_1 = -2.5,\ \theta_2 = 1,\ R = 5)$ against observed *L*-function of Coweeta Plot 318.

300,000 iteration run of Geyer and Møller's , with a 100,000 iteration burn-in used. The resulting 95% envelope of the *L*-function did show attraction up to 2m, followed by gradually increasing repulsion at the interaction distance of 5m. The pattern of attraction and repulsion is then replicated at multiples of the interaction distance, in oscillations of decreasing amplitude (Fig. 50).

The simulation of the exponential triplets (EXPTP) model (see Sec. 3.32, Eqs. 95, 96) is illustrated using four examples with the same parameter values (θ_1 = -1.0 and -2.5; θ_2 = 1.0 and 2.5) as those used to demonstrate the Geyer triplets model. Unlike Geyer's model, the EXPTP model uses the exponential function to model weakening 2nd-order

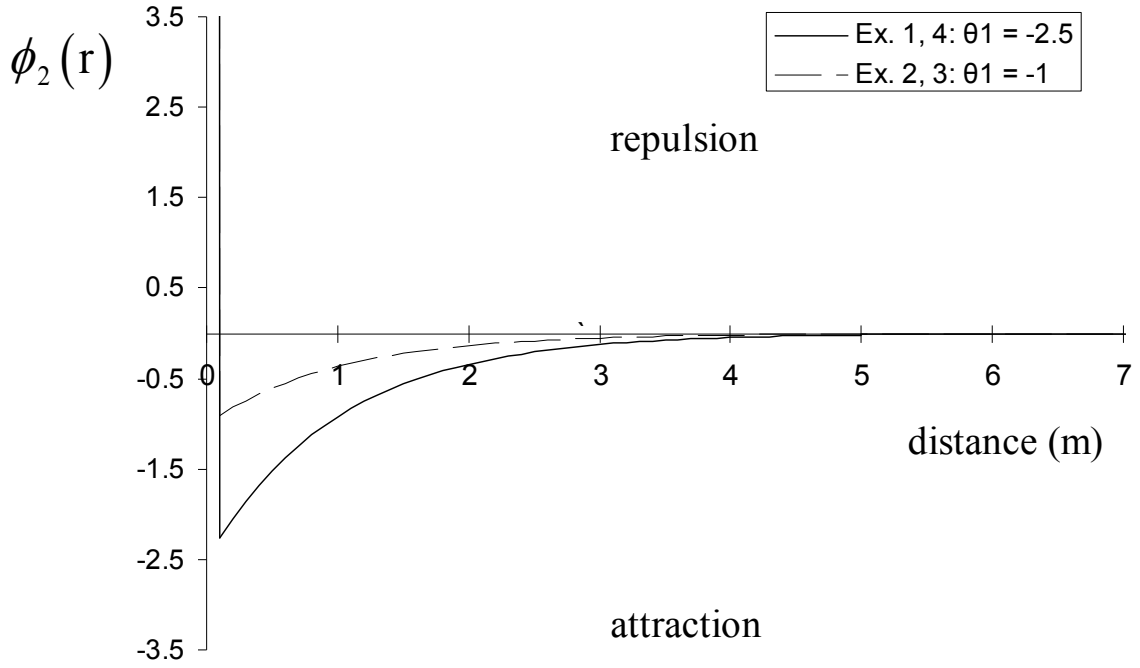

Figure 51. Pair potential plots of simulated Exponential triplets model. Ex. 1 and 4 correspond to θ_1 = -2.5; Ex. 2 and 3 correspond to θ_1 = -1.0.

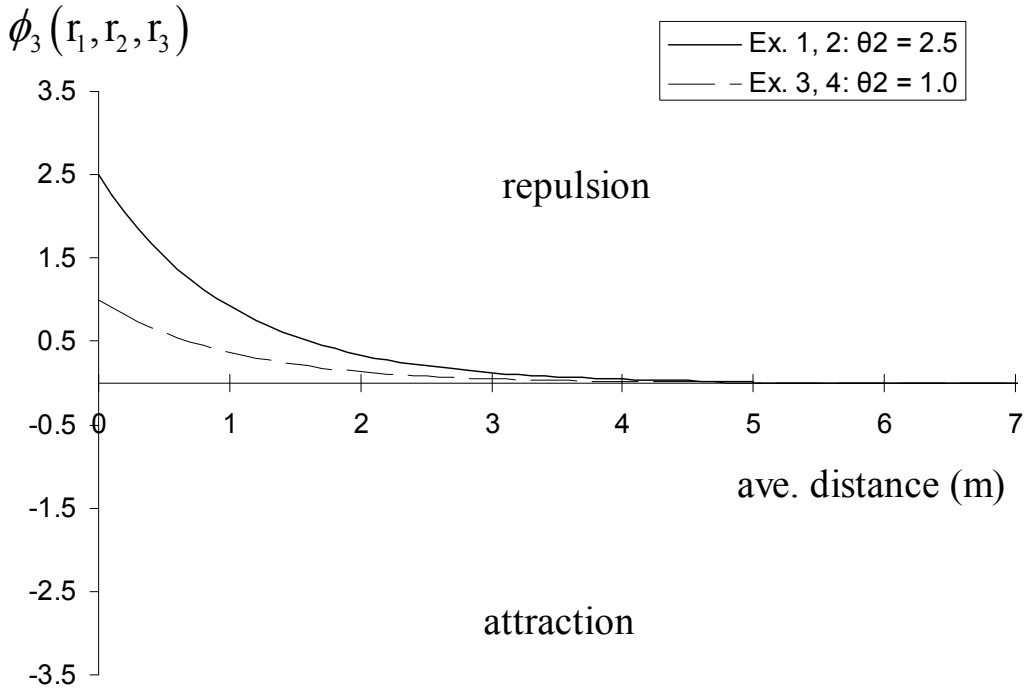

Figure 52. Third order potential plots of simulated Exponential triplets model. Ex. 1 and 2 correspond to θ_2 = 2.5; Ex. 3 and 4 correspond to θ_2 = 1.0.

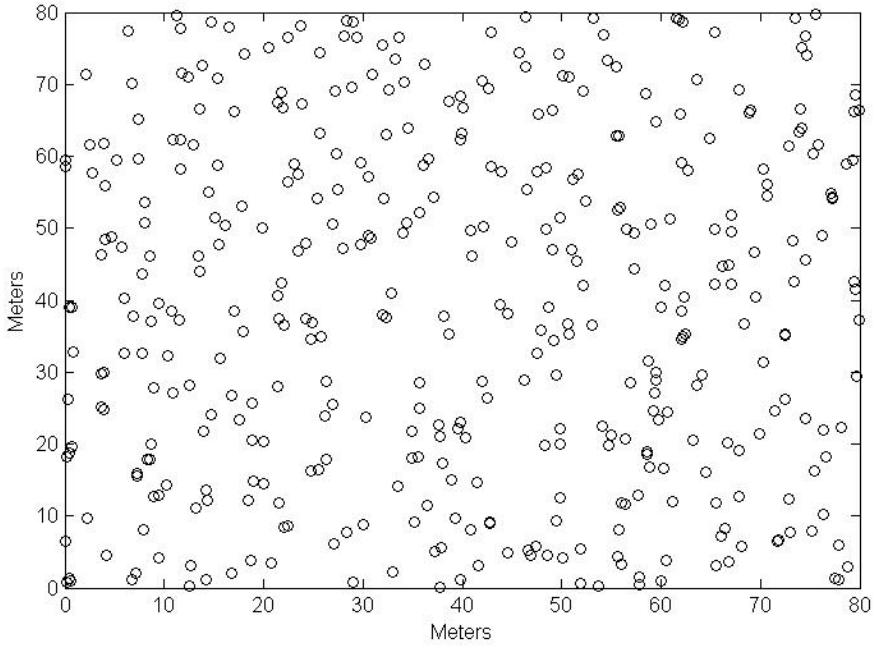

Figure 53. Sample configuration of Ex. 1 of the EXPTP model with parameter values $(\beta = 0.0545,\ \theta_1 = -2.5,\ \theta_2 = 2.5,\ R = 5)$.

spatial attraction and 3rd-order repulsion with distance (Figs. 50, 51). The 2nd-order potential of the EXPTP model is essentially the Fiksel potential with the "irregular" parameter of the Fiksel model (θ_2) fixed to one; the 3rd-order potential describes decreasing repulsion as a function of the average pairwise distance of the three pairs formed from any interacting triple of points.

For the simulation of the first example of the EXPTP process with (θ_1 = -2.5; θ_2 = 2.5) Geyer and Møller's algorithm burned in after 20,000 iterations to a process with approximately 375 points (Fig. C13 – C15). Sample configurations showed a mixture of clusters of two or three points interspersed in a somewhat regular pattern (Fig. 53). The Mone Carlo envelope of the *L*-function showed features similar to the observed

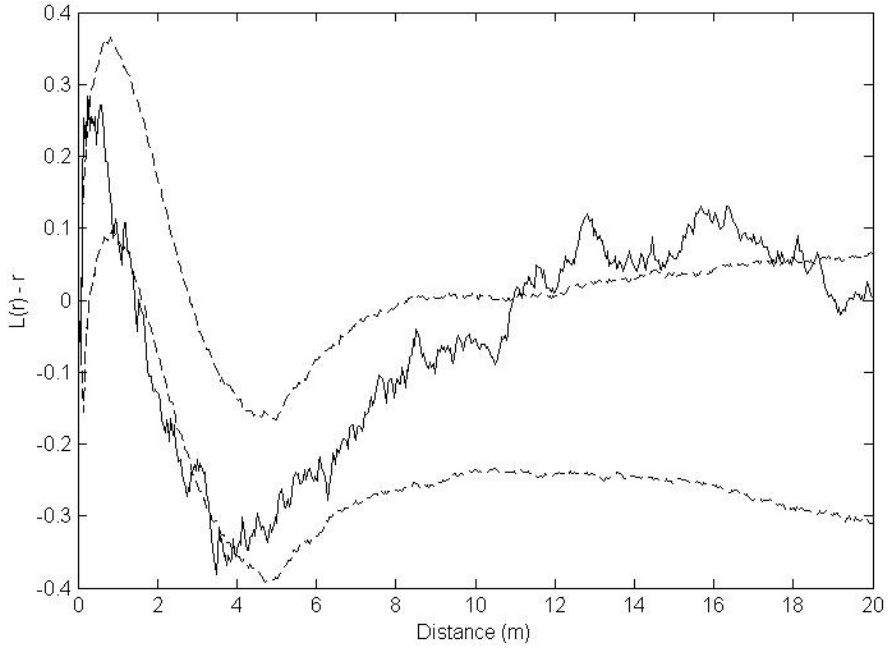

Figure 54. Monte Carlo envelope of *L*-function of Ex. 1 of the EXPTP model with parameter values $(\beta = 0.0545,\ \theta_1 = -2.5,\ \theta_2 = 2.5,\ R = 5)$ against observed *L*-function of Coweeta Plot 318.

L-function of Plot 318 (Fig. 54). Both the EXPTP model and the Coweeta data showed similar levels of clustering at distances less than 2m, although the maximum level of point repulsion given by the EXPTP model occurs at somewhat larger distances (5m) than does the Plot 318. In contrast to the Plot 318 data, the EXPTP model shows repulsion at distances greater than 10m, whereas the observed data was devoid of any non-random spatial structure at these distances.

Weakening the strength of the attractive pair potential in the second example of the EXPTP process with (θ_1 = -1; θ_2 = 2.5), resulted in fast convergence to a point

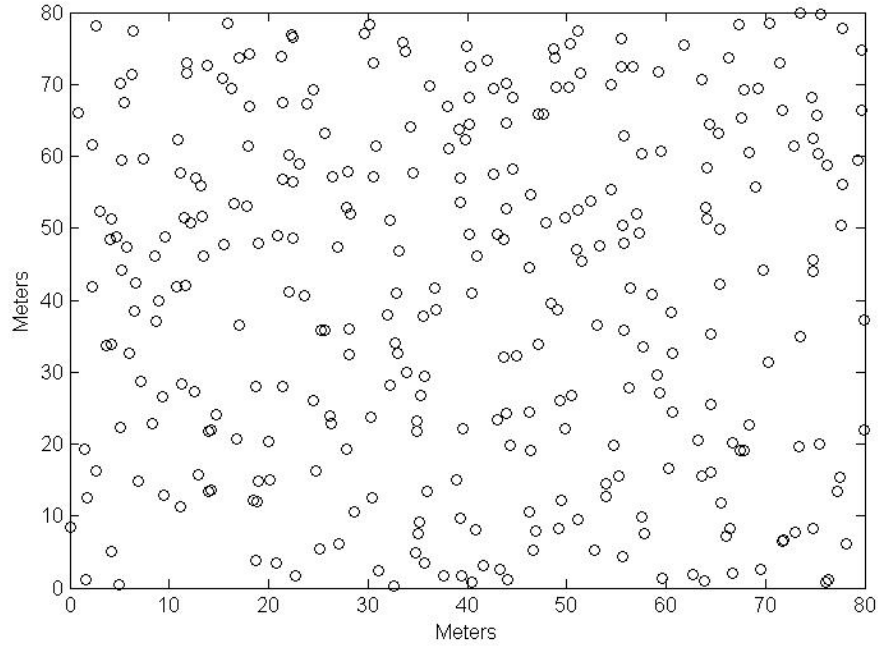

Figure 55. Sample configuration of Ex. 2 of the EXPTP model with parameter values $(\beta = 0.0545,\ \theta_1 = -1,\ \theta_2 = 2.5,\ R = 5)$.

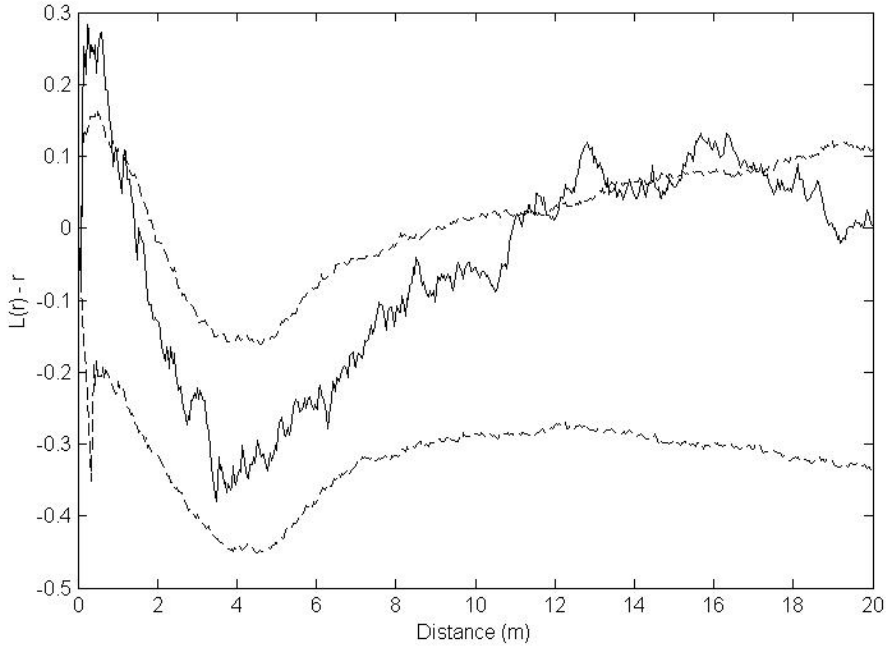

Figure 56. Monte Carlo envelope of *L*-function of Ex. 2 of the EXPTP model with parameter values $(\beta = 0.0545,\ \theta_1 = -1,\ \theta_2 = 2.5,\ R = 5)$ against observed *L*-function of Coweeta Plot 318.

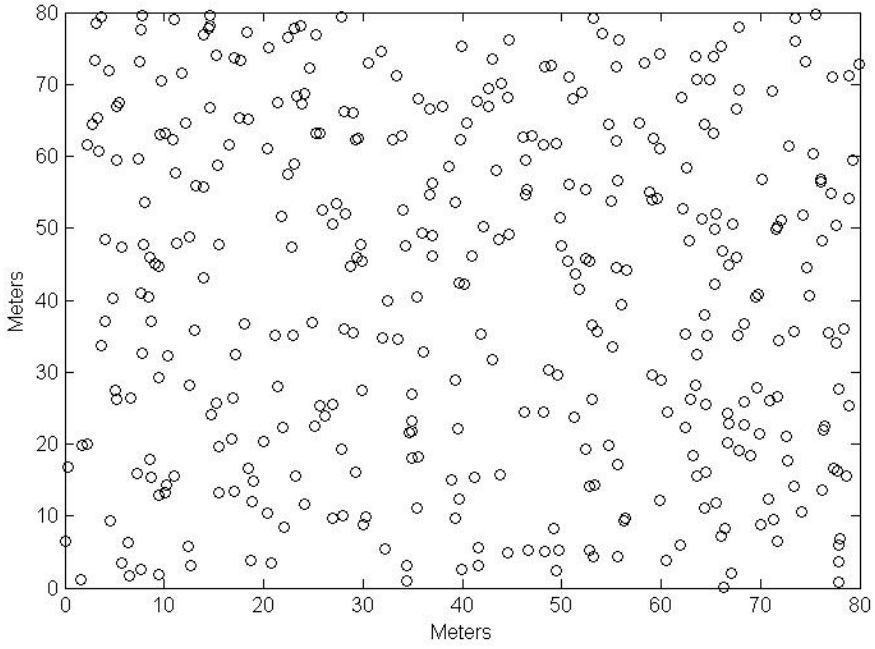

Figure 57. Sample configuration of Ex. 3 of the EXPTP model with parameter values $(\beta = 0.0545,\ \theta_1 = -1,\ \theta_2 = 1,\ R = 5)$.

process with under 300 points (Fig. C16 – 18). Sampled configurations appear very similar to the previous example, with scattered clusters of two or three trees interspersed in otherwise overdispersed point configurations (Fig. 55). By reducing the strength of the pair potential with θ_1 = -1, the Monte Carlo envelope of the EXPTP process no longer overlaps with the *L*-function of the Plot 318 data at distances less than 1m (Fig. 56), implying that the observed data is more strongly clustered than what the EXPTP model would indicate with (θ_1 = -1; θ_2 = 2.5),

In the third example of the EXPTP process, with (θ_1 = -1; θ_2 = 1), both the strength of the attractiveness of the 2nd-order potential and the strength of the repulsion of

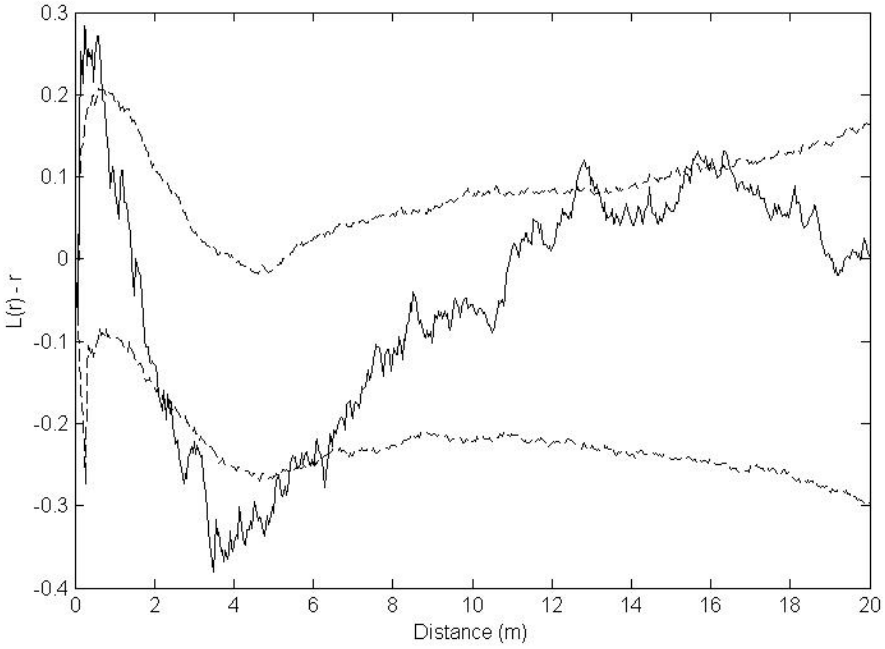

Figure 58. Monte Carlo envelope of *L*-function of Ex. 3 of the EXPTP model with parameter values $(\beta = 0.0545,\ \theta_1 = -1,\ \theta_2 = 1,\ R = 5)$ against observed *L*-function of Coweeta Plot 318.

the 3rd-order potential has been reduced. As might be expected, this leads to sampled configurations that are closer to CSR, though it does not eliminate clustering at distances less than 2m or repulsion around 5m (Fig. 57, 58). As evidenced by the *L*-function Monte Carlo envelope, the observed data on Plot 318 show stronger clustering at distances less than 1m and stronger repulsion from 3-5m than what the EXPTP process provides with parameters (θ_1 = -1; θ_2 = 1).

As in the case with Geyer's triplets process, when a strongly attractive pair potential is joined with a weakly attractive 3rd order potential, Geyer and Møller's

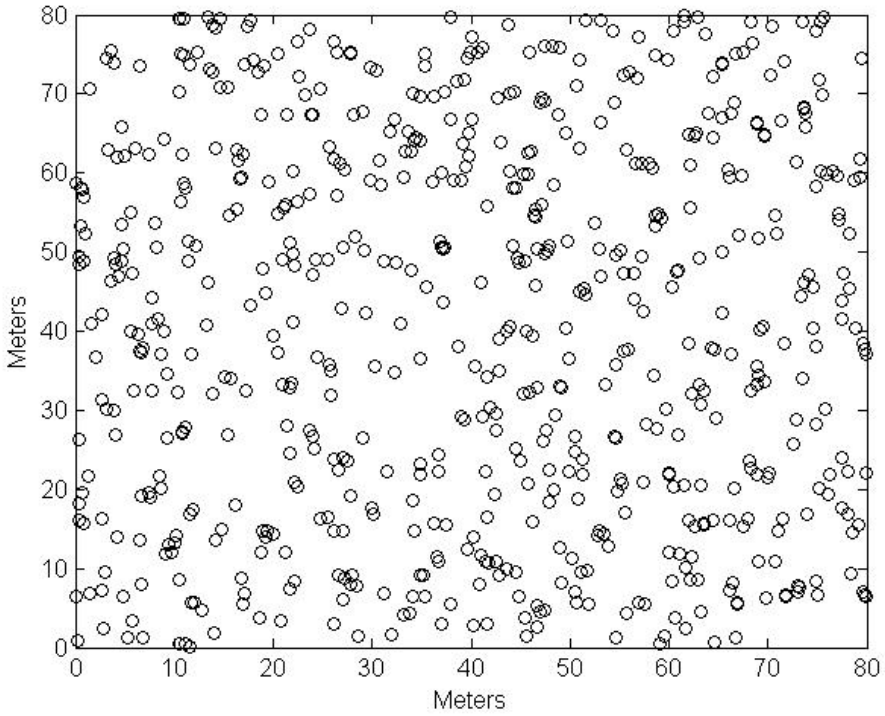

Figure 59. Sample configuration of Ex. 4 of the EXPTP model with parameter values $(\beta = 0.0545,\ \theta_1 = -2.5,\ \theta_2 = 1,\ R = 5)$.

algorithm converges more slowly and converges to a process with a much higher point density than would be expected for a given value of β and interaction distance. The EXPTP process converged to a stationary distribution with sampled configurations having nearly 600 points (Fig. C22 – C24). These samples also show larger numbers of clusters and more points in clusters (Fig. 59). The Monte Carlo envelope of the *L*-function confirms the higher degree of clustering, with clustering stronger than what is observed on Plot 318 at distances less than 3m. Additionally the overdispersion in sample point patterns has been reduced in the 3 – 7m range (Fig. 60).

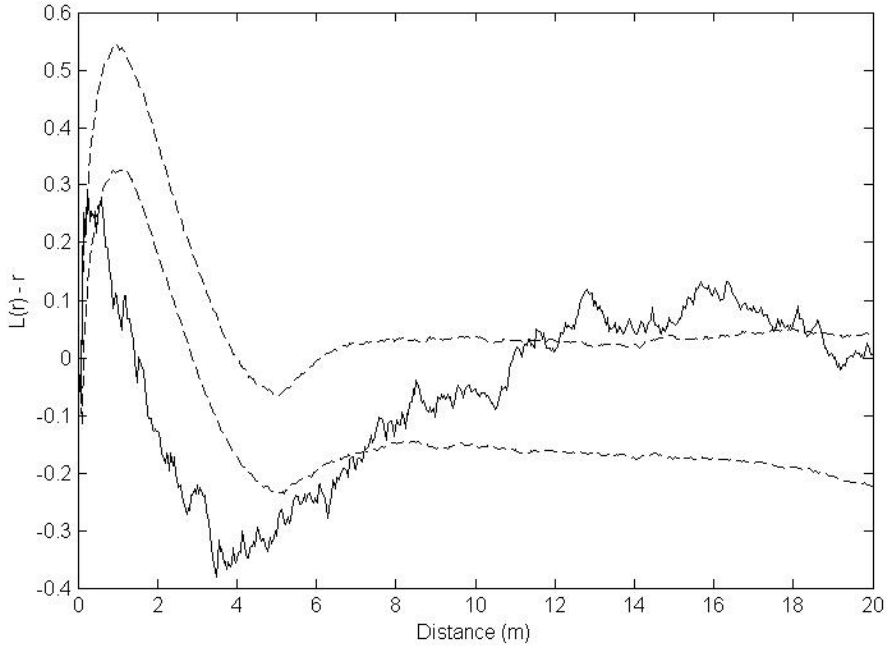

Figure 60. Monte Carlo envelope of *L*-function of Ex. 4 of the EXPTP model with parameter values $(\beta = 0.0545,\ \theta_1 = -2.5,\ \theta_2 = 1,\ R = 5)$ against observed *L*-function of Coweeta Plot 318.

The serpentine triplets process (SPTP) was created by joining the pair potential from the hard-core serpentine process with the same repulsive 3rd-order potential used in the EXPTP process (see Sec. 3.32, Eqs. 99, 100). Unlike either Geyer's triplets process or the EXPTP process, the pair potential of the SPTP process can be simultaneously attractive and repulsive, modeling either short range attraction combined with longer range repulsion (when $\theta_1 > 0$ and $-2.5 \le \theta_2 \le -0.25$) or modeling short range repulsion combined with longer range attraction (when $\theta_1 < 0$ and $-2.5 \le \theta_2 \le -0.25$). When θ_2 falls outside the interval [-2.5, -0.25], the pair potential will be either purely attractive when the parameters θ_1 and θ_2 are of different sign, or will be purely repulsive when they are

both positive or both negative. To illustrate the behavior of the SPTP process, five examples are selected with the first four examples comparing different levels of short range attraction and longer range repulsion in the pair potential with different levels of repulsion in the 3rd-order potential (Figs. 60, 61). The fifth example provides for a pair potential with repulsion up to 1m, attraction from 1 to 5m, and a 3rd order potential with strong repulsion.

In the first example, with $\theta_1 = 2.5$, $\theta_2 = -1$, and $\theta_3 = 1$, strong attraction of points up to 1m is joined with repulsion beyond 1m in the pair potential and a weakly repulsive 3rd-order potential. The trace plots from Geyer and Møller's algorithm suggest quick convergence to a process with under 250 points (Figs. C25 – C27), but mixing of the chain is poor as evidenced by the irregular jumps in the trace plots of the pseudo-canonical statistics in Figs. C26 and C27. Upon visual inspection, sample configurations show tight clusters of two and three points embedded in an overdispersed pattern (Fig. 63). The Monte Carlo envelope of the *L*-function indicates that Ex. 1 of the SPTP process ($\theta_1 = 2.5$, $\theta_2 = -1$, $\theta_3 = 1$) shows strong clustering up to 2m, exceeding the degree of clustering in Plot 318. Overdispersion is present at distances larger than 2m, and appears to be close in magnitude to that of Plot 318 (Fig. 64).

Increasing the strength of repulsion in the 3rd order potential, while maintaining the same features of the partially attractive serpentine pair potential provides the basis for the second example, where $\theta_1 = 2.5$, $\theta_2 = -1$, and $\theta_3 = 2.5$. The trace plots from Geyer and Møller's algorithm show quick convergence to a process with slightly less points (about 225) and better mixing (Figs. C28 – C30). As might be expected, the SPTP process now shows more repulsive configurations, with fewer clusters of 2-3 points (Fig. 65). The

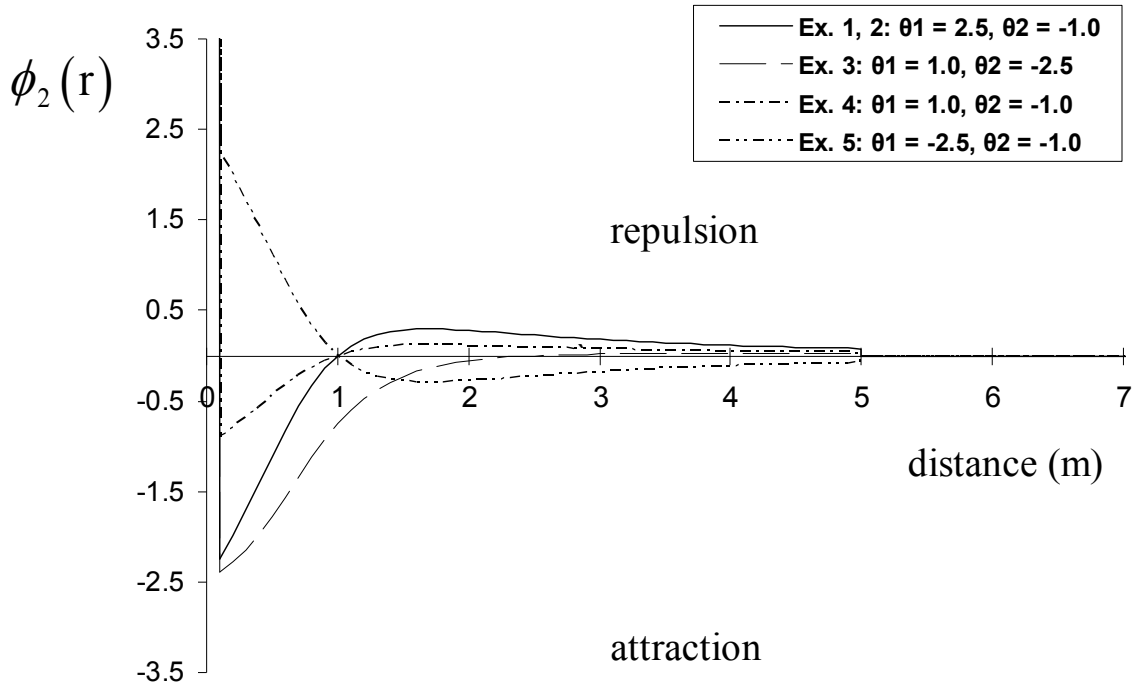

Figure 61. Pair potential plots of simulated serpentine triplets model. Ex. 1 and Ex. 2 correspond to $\theta_1 = 2.5$ and $\theta_2 = -1$; Ex. 3 corresponds $\theta_1 = 1$ and $\theta_2 = -2.5$; Ex. 4 corresponds to $\theta_1 = 1$ and $\theta_2 = -1$; Ex. 5 corresponds to $\theta_1 = -2.5$ and $\theta_2 = -1$.

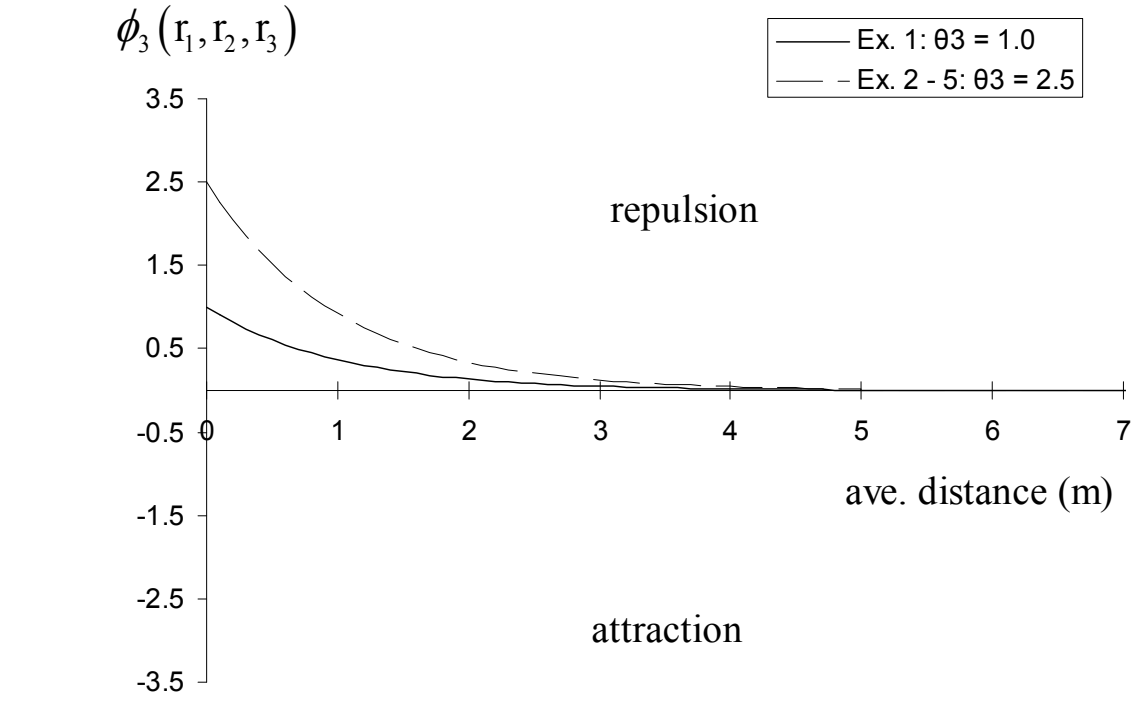

Figure 62. Third order potential plots of simulated serpentine triplets model. Ex. 1 corresponds to $\theta_3 = 1$; Ex. 2, 3, 4, 5 correspond $\theta_3 = 2.5$

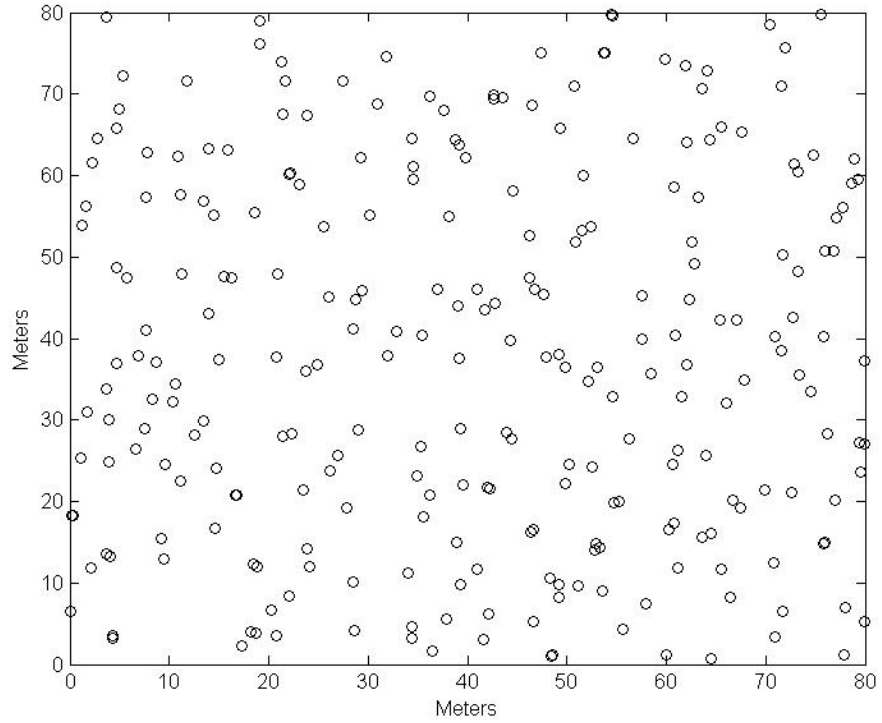

Figure 63. Sample configuration of Ex. 1 of the SPTP model with parameter values $(\beta = 0.0545,\ \theta_1 = 2.5,\ \theta_2 = -1,\ \theta_3 = 1,\ R = 5)$.

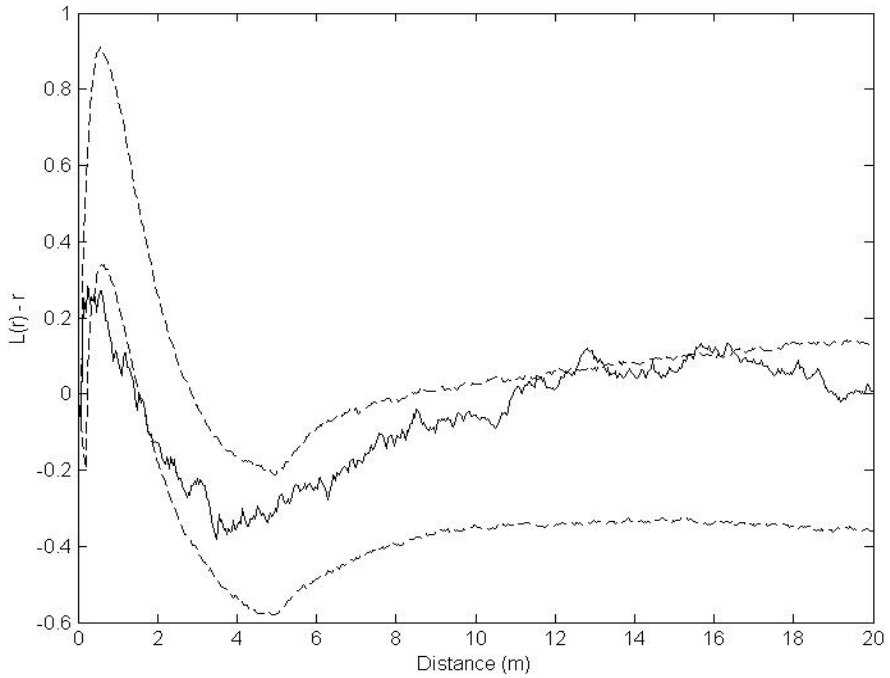

Figure 64. Monte Carlo envelope of *L*-function of Ex. 1 of the SPTP model with parameter values $(\beta = 0.0545,\ \theta_1 = 2.5,\ \theta_2 = -1,\ \theta_3 = 1,\ R = 5)$ against observed *L*-function of Coweeta Plot 318.

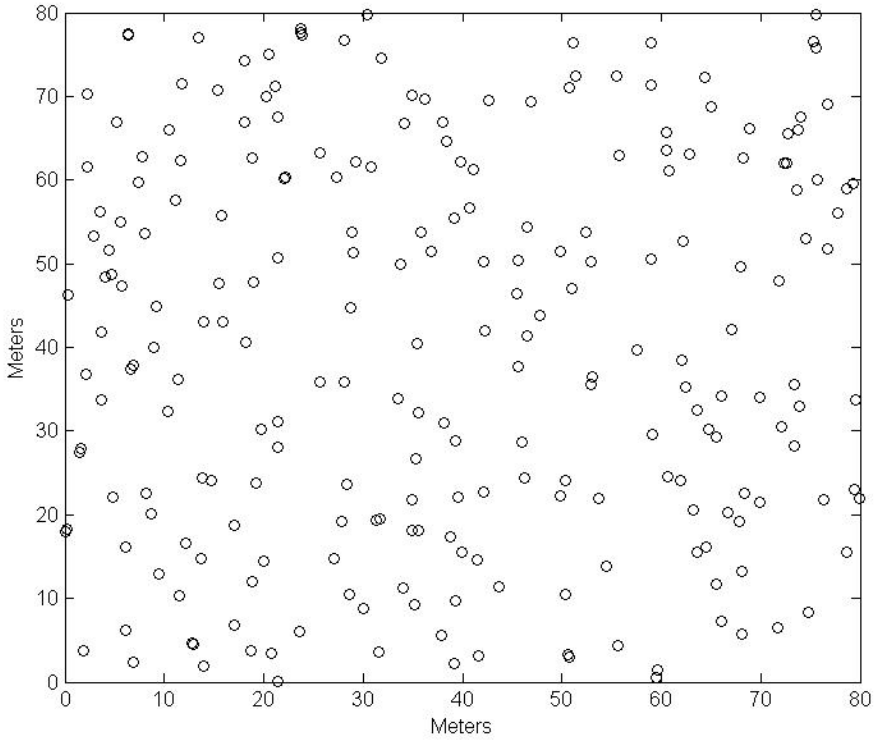

Figure 65. Sample configuration of Ex. 2 of the SPTP model with parameter values $(\beta = 0.0545,\ \theta_1 = 2.5,\ \theta_2 = -1,\ \theta_3 = 2.5,\ R = 5)$.

Increase in the repulsive strength of the 3rd-order potential has resulted in a decrease in the clustering at distances up to 2m, where now the Plot 318 data is consistent with the Monte Carlo envelope for the SPTP process at these distances (Fig. 66). The overdispersion at larger distances has also been increased to the point where the Plot 318 data is only marginally consistent with the process at distances greater than 5m (Fig. 66).

The third example, with $\theta_1 = 1$, $\theta_2 = -2.5$, and $\theta_3 = 2.5$, serves to illustrate the effect of a combining a pair potential showing relatively strong attraction with a strongly

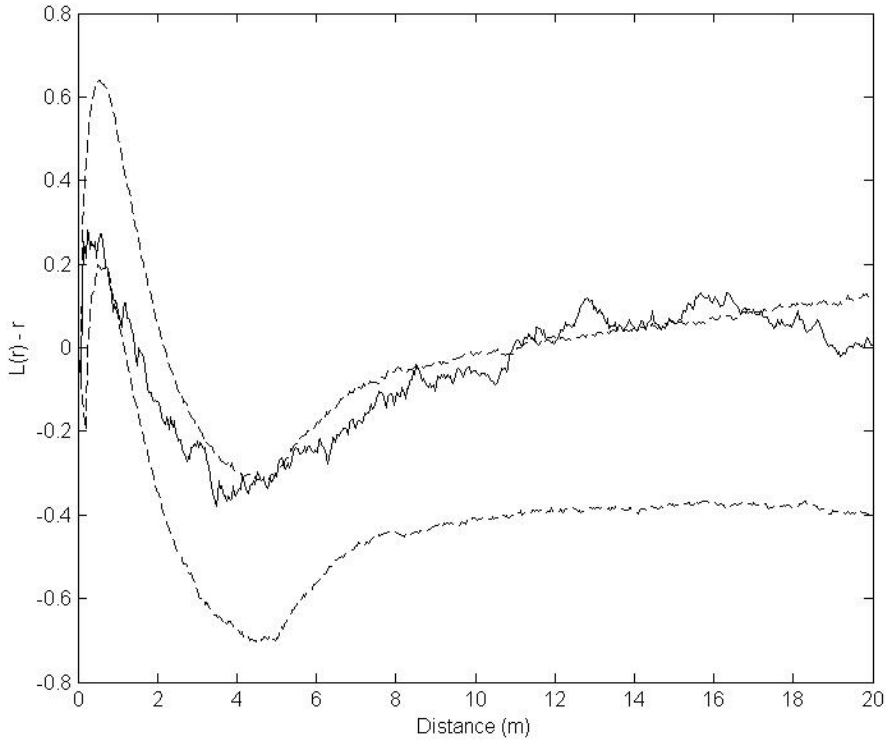

Figure 66. Monte Carlo envelope of *L*-function of Ex. 2 of the SPTP model with parameter values $\left(\beta = 0.0545,\ \theta_1 = 2.5,\ \theta_2 = -1,\ \theta_3 = 2.5,\ R = 5\right)$ against observed *L*-function of Coweeta Plot 318.

repulsive 3rd-order potential. Geyer and Møller's algorithm burned in quickly to a process that was stable at 300 points and appeared to mix reasonably well (Figs. C31 – C33). Sample configurations showed many clusters of two and three points, but no tendency for single points to be necessarily drawn into clusters or for clusters to contain more than three points (Fig. 67). Monte Carlo simulation of the *L*-function for the SPTP process with these parameter values indicates that the SPTP process has more clustering than the Plot 318 data, but the repulsion in the 3-7m range is quite comparable (Fig. 68). At

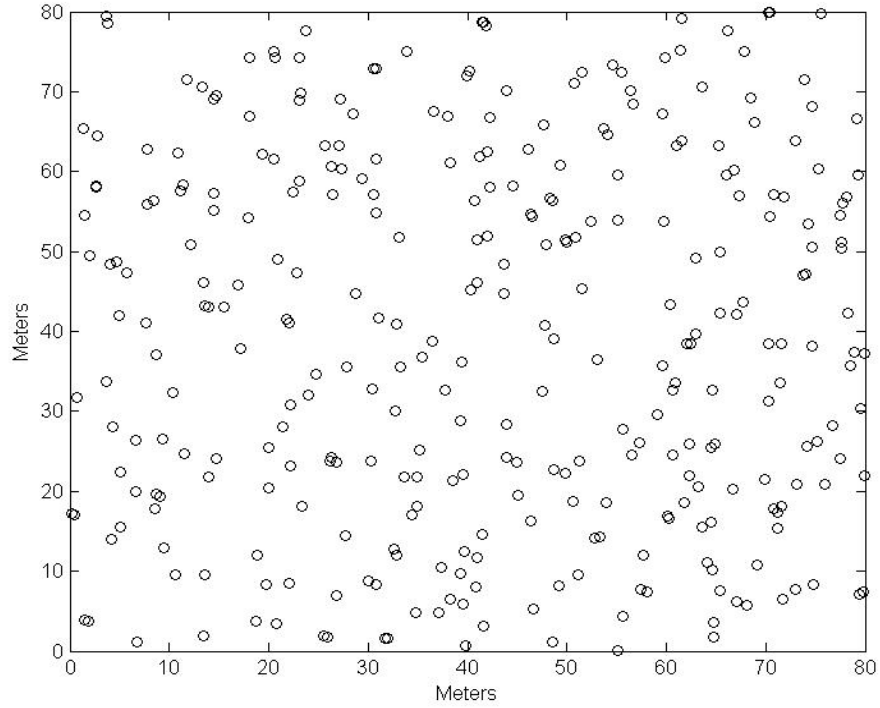

Figure 67. Sample configuration of Ex. 3 of the SPTP model with parameter values $(\beta = 0.0545,\ \theta_1 = 1,\ \theta_2 = -2.5,\ \theta_3 = 2.5,\ R = 5)$.

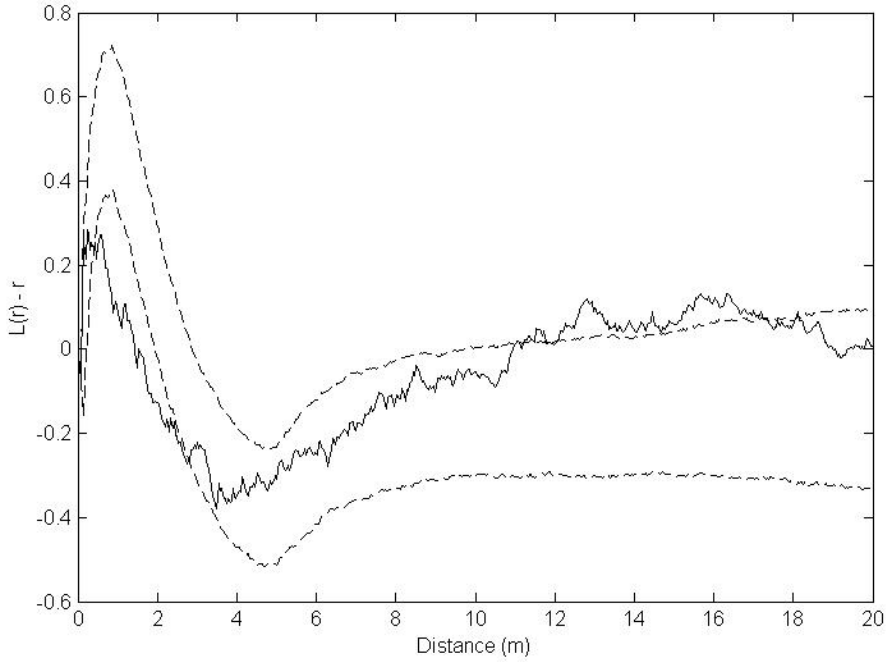

Figure 68. Monte Carlo envelope of *L*-function of Ex. 3 of the SPTP model with parameter values $(\beta = 0.0545,\ \theta_1 = 1,\ \theta_2 = -2.5,\ \theta_3 = 2.5,\ R = 5)$ against observed *L*-function of Coweeta Plot 318.

distances from 10 – 17m, the Plot 318 data shows random spatial structure, but the SPTP process still models repulsion, despite the fact that both the 2nd- and 3rd-order potentials have an interaction distance of 5m.

The fourth example of the SPTP process illustrates a combination of weak spatial attraction up to 1m combined with strong repulsion in the 3rd-order potential, where θ_1 = 1, θ_2 = -1, and θ_3 = 2.5. Convergence was quite rapid to a process with just less than 250 points and mixing is adequate. (Figs. C34 – C36). Sampled configurations show overdispersion in a pattern of points and a few close pairs of points (Fig. 69). The combination of weak 2nd-order attraction and strong 3rd-order repulsion does not eliminate all clustering from the process, however. The Monte Carlo simulation envelope of the *L*-function shows that the SPTP process still has clustering up to 2m that is not much less than the clustering found in the Plot 318 data (Fig. 70).

The fifth example of the SPTP process illustrates the unusual combination of repulsion at short distances and attraction at larger distances. The parameters chosen for this example (θ_1 = -2.5, θ_2 = -1, and θ_3 = 2.5) give the same pair potential that was used for the HCSP model which, when simulated, degenerated into large diffuse clumps of points (with *n* random) or one large clump (with *n* fixed) as seen in Figs. 38 and 39. This pair potential is then joined with a strongly repulsive 3rd-order potential to form the SPTP process. Simulation of the SPTP process is markedly different than what was observed for the HCSP process with the same pair potential. Trace plots show quick convergence to a stable process with about 350 points (Figs C37 – C39). Sample configurations appear to show no clustering at all (Fig. 71), a feature confirmed by the *L*-function envelope of

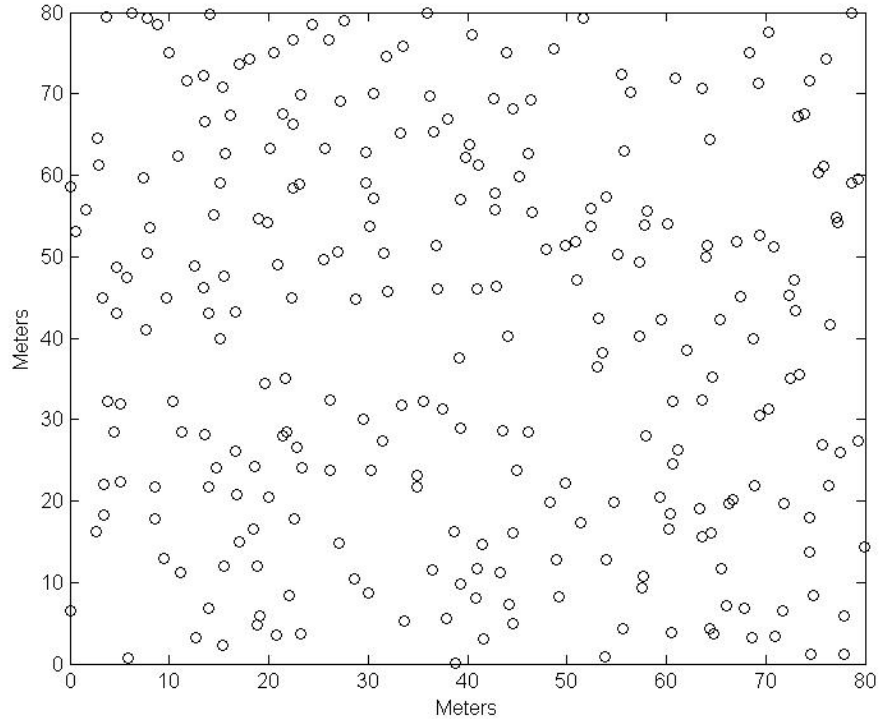

Figure 69. Sample configuration of Ex. 4 of the SPTP model with parameter values $(\beta = 0.0545,\ \theta_1 = 1,\ \theta_2 = -1,\ \theta_3 = 2.5,\ R = 5)$.

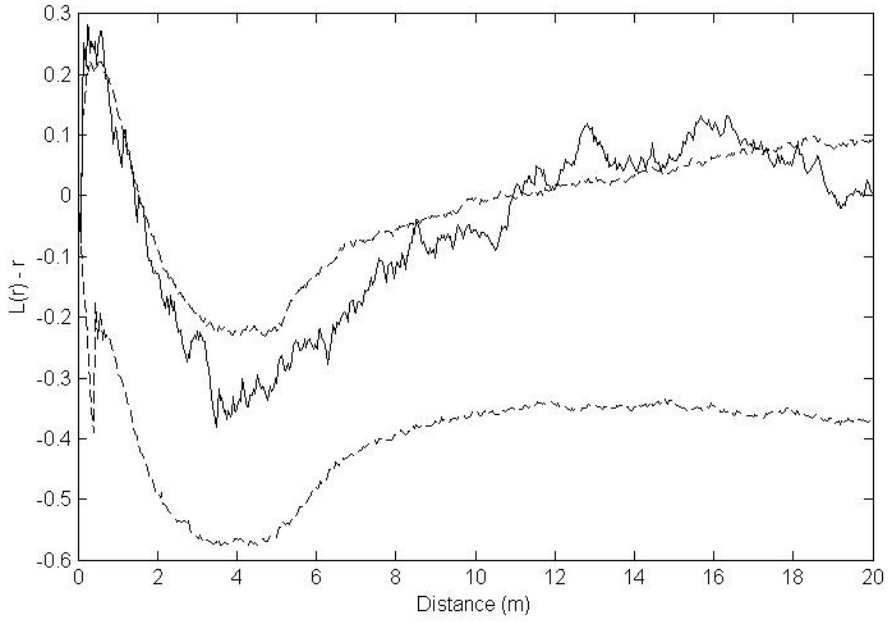

Figure 70. Monte Carlo envelope of *L*-function of Ex. 4 of the SPTP model with parameter values $(\beta = 0.0545,\ \theta_1 = 1,\ \theta_2 = -1,\ \theta_3 = 2.5,\ R = 5)$ against observed *L*-function of Coweeta Plot 318.

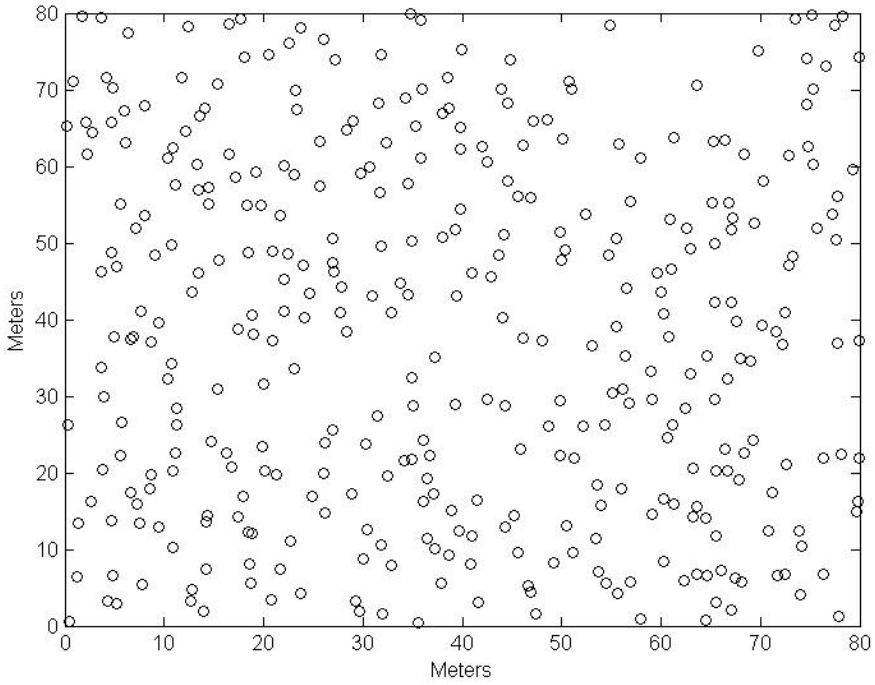

Figure 71. Sample configuration of Ex. 5 of the SPTP model with parameter values $(\beta = 0.0545,\ \theta_1 = -2.5,\ \theta_2 = -1,\ \theta_3 = 2.5,\ R = 5)$.

the SPTP process showing only overdispersion of points; the long range attraction in the pair potential appears to be completely overwhelmed by the repulsion of the 3rd-order potential (Fig 71).

Although the previous examples of the GTP, EXPTP, and SPTP processes have all combined strong pairwise attraction relative to weak 3rd-order repulsion, none have contained so strong attraction that these triplets processes degenerated into a "one-clump" process as some of the attractive pairwise processes did. To explore the behavior of the triplets models under more extreme pairwise attractiveness, the three triplets processes were simulated on a smaller 30 x 30m square, with β = 0.0545 and R = 5m as in previous

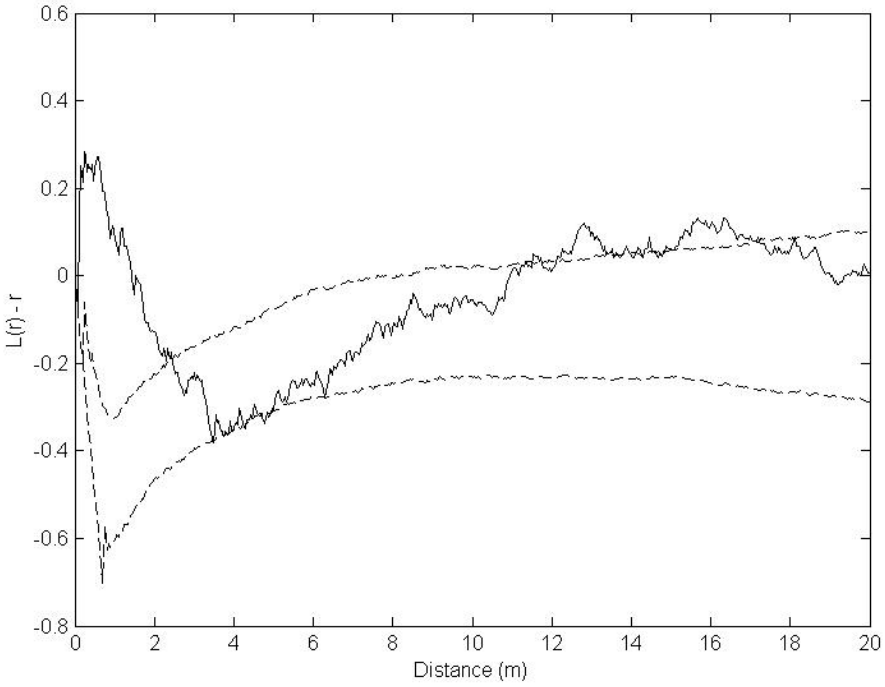

Figure 72. Monte Carlo envelope of *L*-function of Ex. 5 of the SPTP model with parameter values $(\beta = 0.0545,\ \theta_1 = -2.5,\ \theta_2 = -1,\ \theta_3 = 2.5,\ R = 5)$ against observed *L*-function of Coweeta Plot 318.

test cases. To provide a test of "extreme" attractiveness, the parameters for the pair potential of each triplets model was chosen so that at the distance 0m, each model's pair potential would be 5 times as large, in absolute value, as the 3rd-order potential at that distance. For Geyer's triplets model, this is achieved by choosing θ_1 = -5 and θ_2 = 1. Similarly the EXPTP process was simulated with θ_1 = -5 and θ_2 = 1, and the SPTP process was simulated with θ_1 = 1, θ_2 = -5, and θ_3 = 1, giving a purely repulsive potential that approaches 0 by 3m.

Trace plots from Geyer and Møller's algorithm (not presented here) indicated slow convergence for each of the triplets processes. The GTP process required nearly

150,000 iterations to converge to a process with almost 400 points on the 30 x 30m square. The EXPTP process required 200,000 iterations to converge to a process with 325 points, and the SPTP process also needed 200,000 iterations to converge to a process with about 380 points. As a basis for comparison, the center 30 x 30m piece of Plot 318 contained 60 trees, so it is clear that triplets processes with very strong attraction in the pair potential and weak 3rd-order repulsion lead to simulated point patterns with exceedingly high point densities. Sample configurations from the triplets processes show a noticeable difference with the GTP process. Configurations show linear "chains" of points not more than 5m in length and contain up to 8 points (Fig. 73). The EXPTP and SPTP processes, which limit the strong attraction to short distances, show clusters of 10 – 12 points arranged in a regular fashion (Fig. 74 - 75).

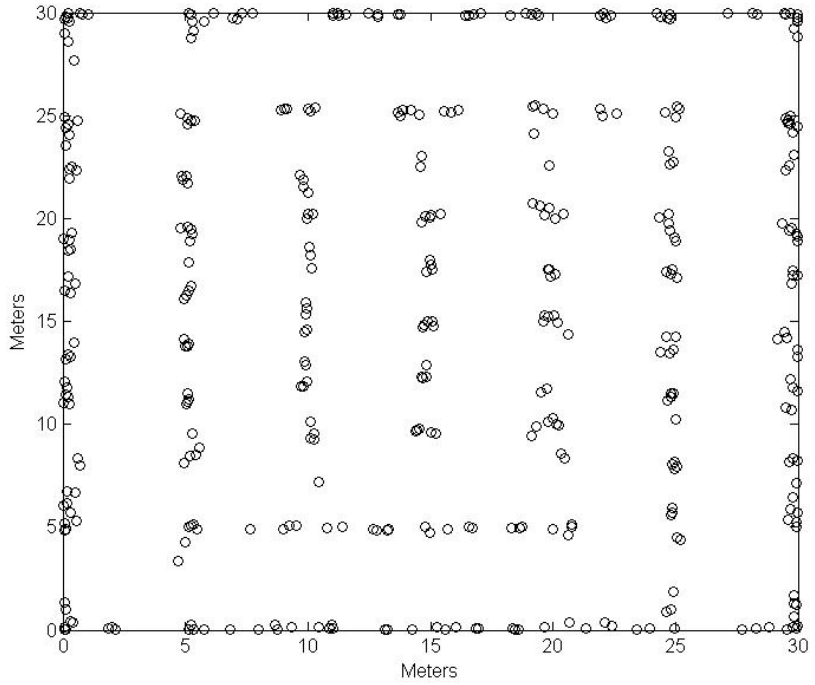

Figure 73. Sample configuration of simulated GTP model with parameter values $(\beta = 0.0545,\ \theta_1 = -5,\ \theta_2 = 1,\ R = 5)$.

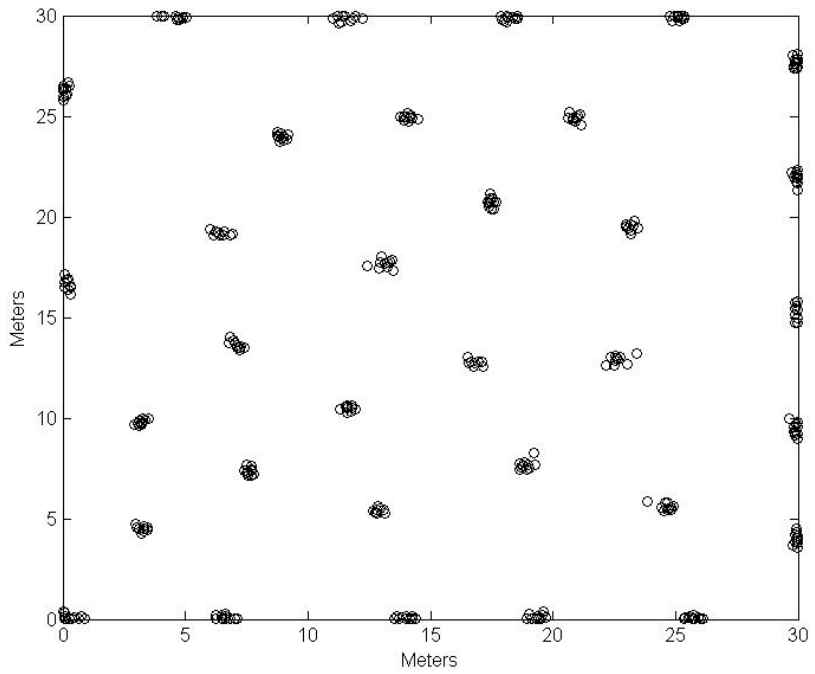

Figure 74. Sample configuration of simulated EXPTP model with parameter values $(\beta = 0.0545,\ \theta_1 = -5,\ \theta_2 = 1,\ R = 5)$.

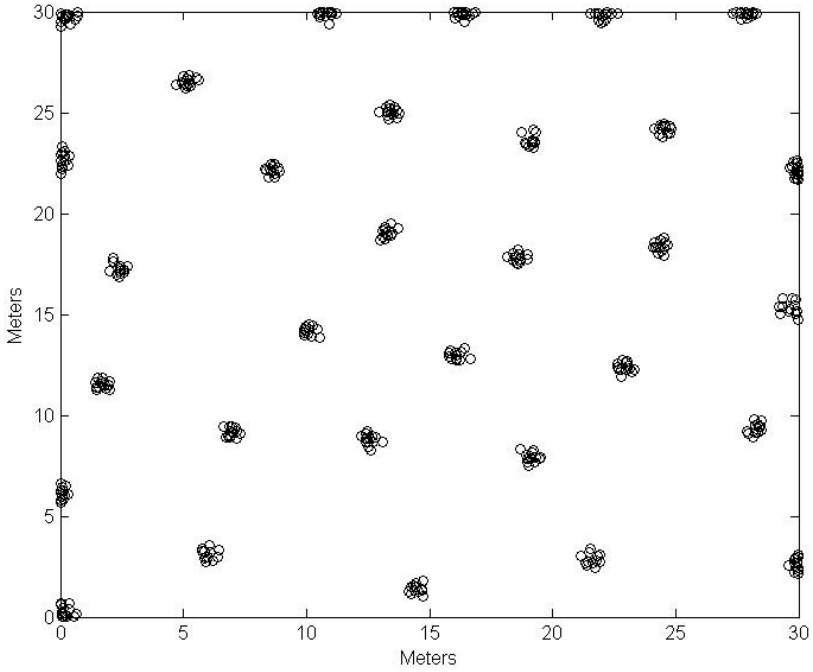

Figure 75. Sample configuration of simulated SPTP model with parameter values $\left(\beta = 0.0545,\ \theta_1 = 1,\ \theta_2 = -5,\ \theta_3 = 1,\ R = 5\right)$.

4.3 Gibbs Point Process Inference Using Importance Sampling

As demonstrated in Section 4.2, the number of iterations needed to burn in a simulation of a triplets process using Geyer and Møller's algorithm can vary widely depending upon the strength of the attraction proposed in the pair potential. For example the serpentine triplets process with very strong 2^{nd}-order attraction and weak 3^{rd}-order repulsion (Fig. 75) needed nearly 200,000 iterations to converge. When the SPTP process has strong repulsion, trace plots indicating a rapid burn-in of less than a few thousand iterations (Figs. C29 – C36). In the context of Bayesian inference for point processes, Geyer and Møller's algorithm is used to generate importance samples to estimate the ratio of normalizing constants at each iteration of Bognar's algorithm. Since importance samples cannot be drawn until Geyer and Møller's algorithm has burned in, the length of this burn-in becomes important for determining the total time needed for Bognar's algorithm to complete one iteration. In practice the burn-in must be small enough so that the iteration time of Bognar's algorithm is manageable. For example even a 1 minute iteration time for a 20,000 iteration run of Bognar's algorithm would require nearly two weeks (333.3 hours) of continuous run time. On the other hand, if the burn-in is chosen to be too small, it is possible that for many iterations importance samples will be drawn under proposed parameter vectors where Geyer and Møller's algorithm has not sufficiently burned in. As a point of reference, Bognar (2004, 2005, 2008) used burn-ins of 1,000, 5,000 and 20,000 iterations for unmarked homogeneous, unmarked inhomogeneous, and marked Gibbs pairwise interaction processes with purely repulsive interactions respectively.

To explore the question of the length of burn-in of Geyer and Møller's algorithm, a simulation study was constructed. Since Geyer and Møller's Metropolis-Hastings algorithm is nested within Bognar's Metropolis-Hastings sampler of the Gibbs posterior distributions, this burn-in is referred to as the "sub-chain" burn-in. The serpentine triplets process was used to test burn-in requirements, assigning parameters ($\theta_1 = 2.5$, $\theta_2 = -1$, $\theta_3 = 2.5$) to represent attractive 2nd-order interactions over distances up to 1m, with repulsive 2nd-order interactions up to a 2m interaction distance. To reduce the computational burden of Bognar's algorithm the SPTP process is simulated on a 10 x 10m "plot" with approximately 50 points, which was achieved by choosing the value of 2.5 for the parameter β. Using Geyer and Møller's algorithm, the SPTP process with parameter values ($\beta = 2.5$, $\theta_1 = 2.5$, $\theta_2 = -1$, $\theta_3 = 2.5$, $R = 2$) converged to a process with roughly 50 points as was expected, and the 100,000th iteration was selected as the test data, referred to hereafter as the SimTest50 dataset (Fig. 76). The SimTest50 data contained 51 points in a pattern containing multiple pairs and close triplets of points, but relatively few singletons. The burn-in for the SPTP process with these parameter values appears to be less than 5,000 iterations based on the trace plot for the number of points (Fig. 77), but trace plots of the pseudo-canonical statistics suggest that the burn-in may be up to 25,000 iterations. To test the effect of the sub-chain burn-in, Bognar's algorithm was run seven times with sub-chain burn-ins of 1,000, 2,500, 5,000, 10,000, 20,000, 50,000, and 75,000 iterations. After each burn-in, 100 importance samples were drawn from the next 10,000 iterations of Geyer and Møller's algorithm at a spacing of 100 iterations for each test run. To avoid a burn-in of Bognar's algorithm, each test run was

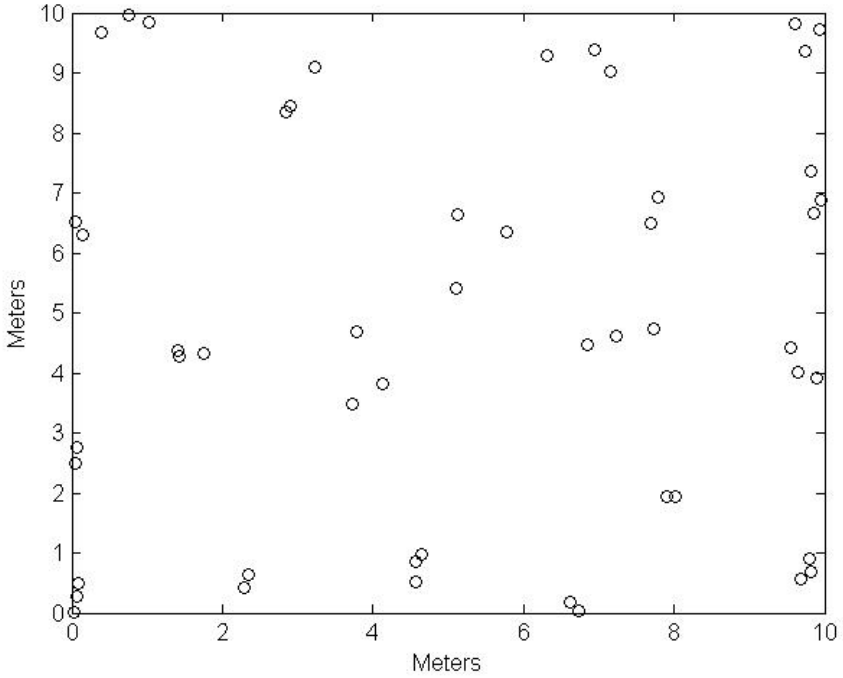

Figure 76. Configuration of the SimTest50 data with parameter values $(\beta = 2.5,\ \theta_1 = 2.5,\ \theta_2 = -1,\ \theta_3 = 2.5,\ R = 2)$.

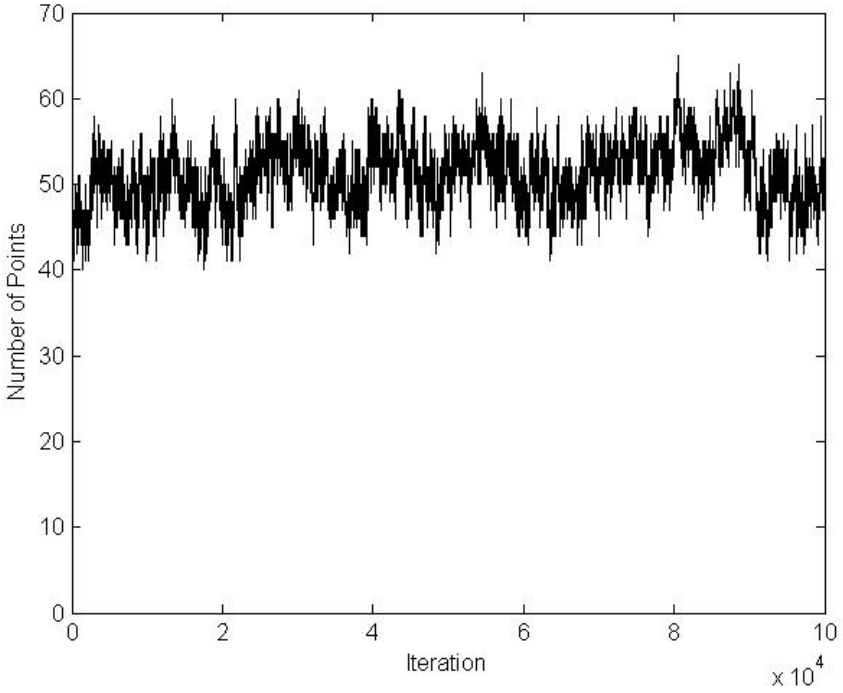

Figure 77. Trace plot of the number of points for the simulation of the SPTP model with $(\beta = 2.5,\ \theta_1 = 2.5,\ \theta_2 = -1,\ \theta_3 = 2.5,\ R = 2)$.

started on the actual parameter values that generated the SimTest50 dataset. For all runs the tuning parameters of the independent normal parameter distributions for proposing the SPTP parameters were set as follows: $\sigma_\beta = 0.01$, $\sigma_{\theta_1} = 0.05$, $\sigma_{\theta_2} = 0.05$,and $\sigma_{\theta_3} = 0.05$. Each run of Bognar's algorithm was completed for 20,000 iterations, with the exception of the 5,000 burn-in test (18,532 iterations) and the 10,000 burn-in test (19,385 iterations) which were terminated early due to the loss of communication with the server running the algorithms. Trace plots for the SPTP process parameters were obtained for each of the seven runs (Figs. 77 – 80) after completion of the C++ code that implemented Bognar's algorithm. Strong trends in the trace plots with respect to the sub-chain burn-in period are not readily apparent. Trace plots for the parameter θ_1, for example, showed substantial overlap, where all runs overestimate the target value (2.5) of this parameter. Very small sub-chain burn-in periods (1,000, 2,500, or 5,000 iterations) lead to a greater underestimation of θ_2 (target value is -1.0) and a greater overestimation of θ_3 (target value is 2.5) than larger burn-in periods, but none of the runs estimated these parameters correctly. Moreover, all of the trace plots indicate slow mixing of the Markov chains through the support of the Gibbs posterior distributions. In particular the trace plots for the parameter β do not appear to freely mix around any particular value for any of the runs (Fig. 78). As further evidence of slow mixing, the acceptance rates from the seven runs of Bognar's algorithm lie in the range of 68% to 75% (Table 5). As a basis for comparison, Robert and Casella (2004, pp 316-318) discuss the notion of desirable acceptance rates for Metropolis-Hastings algorithms being in the range of 25 to 50%.

After completion of the seven runs of Bognar's algorithm, the multivariate normal

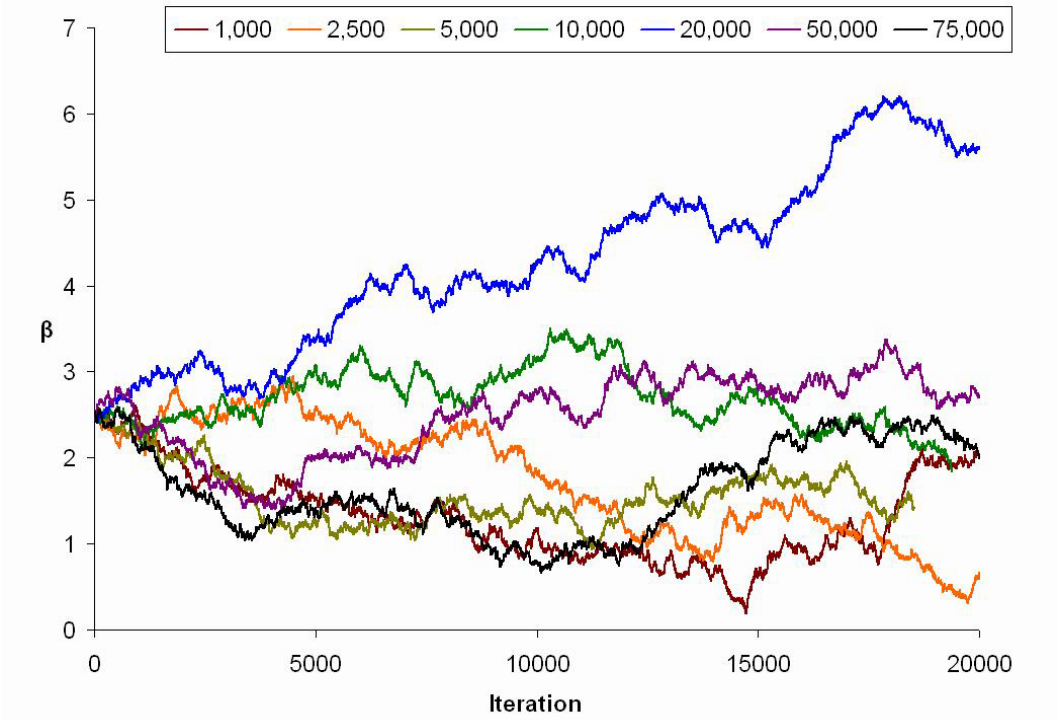

Figure 78. Trace plot of the parameter β of the SPTP model for simulated data on 10 x 10 plot. Sub-chain burn-in ranges from 1,000 iterations (red) to 75,000 iterations (black).

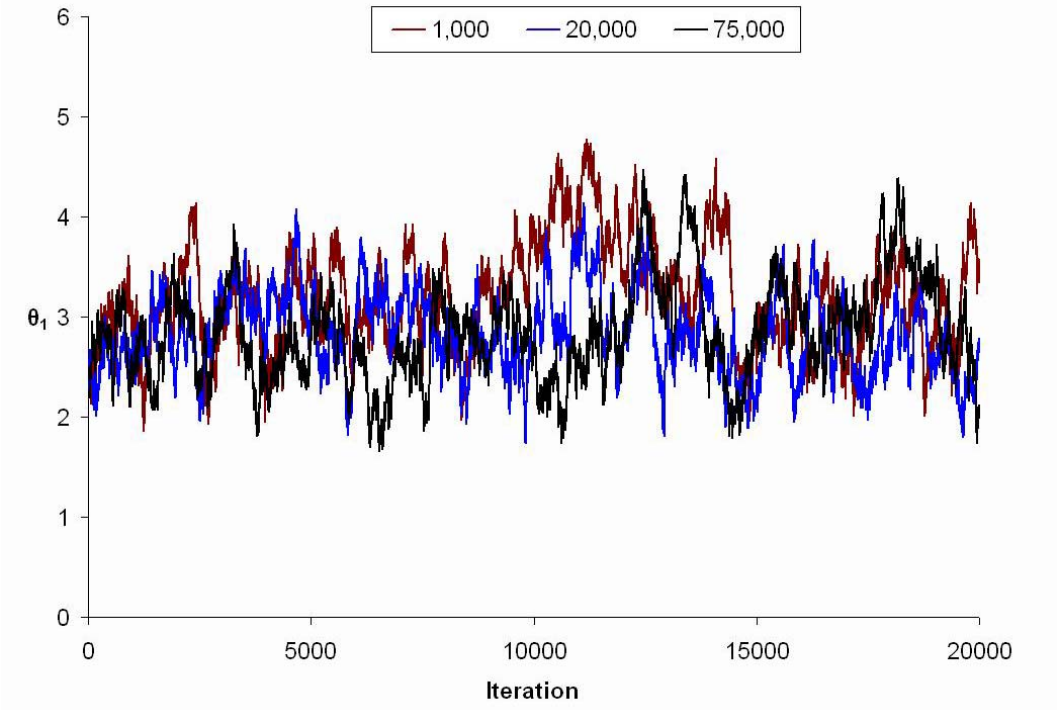

Figure 79. Trace plot of the parameter θ_1 of the SPTP model for simulated data on a 10 x 10 plot. Sub-chain burn-in ranges from 1,000 iterations (red) to 75,000 iterations (black).

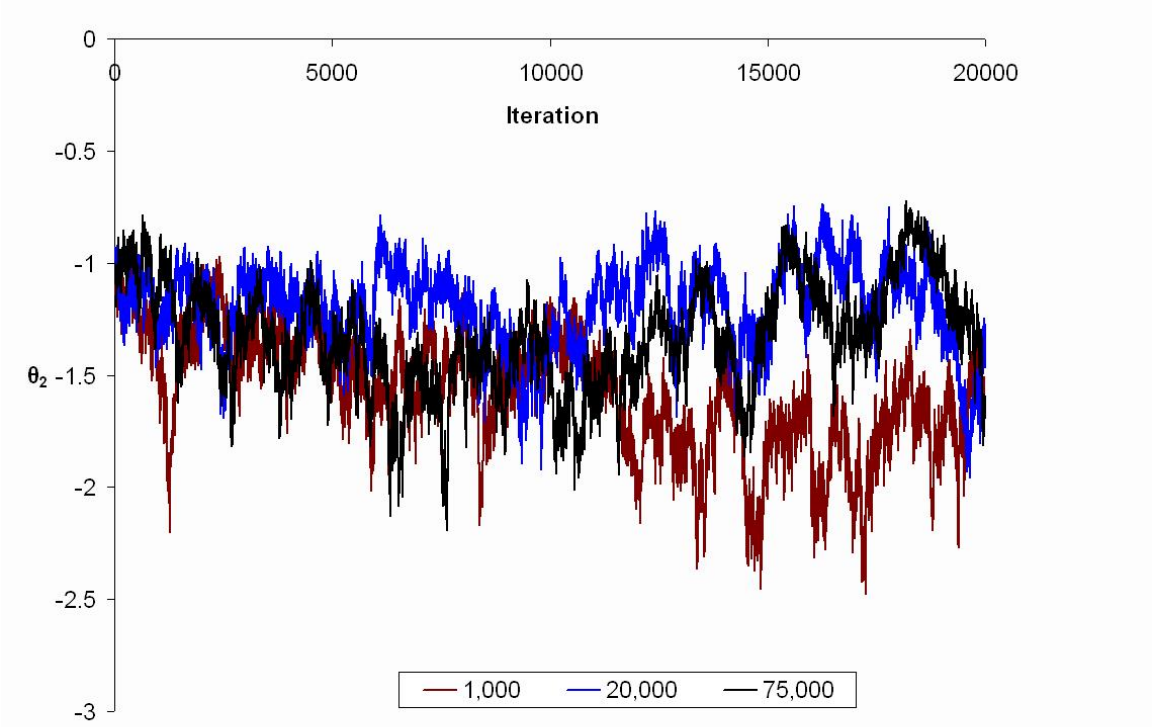

Figure 80. Trace plot of the parameter θ_2 of the SPTP model for simulated data on a 10 x 10 plot. Sub-chain burn-in ranges from 1,000 iterations (red) to 75,000 iterations (black).

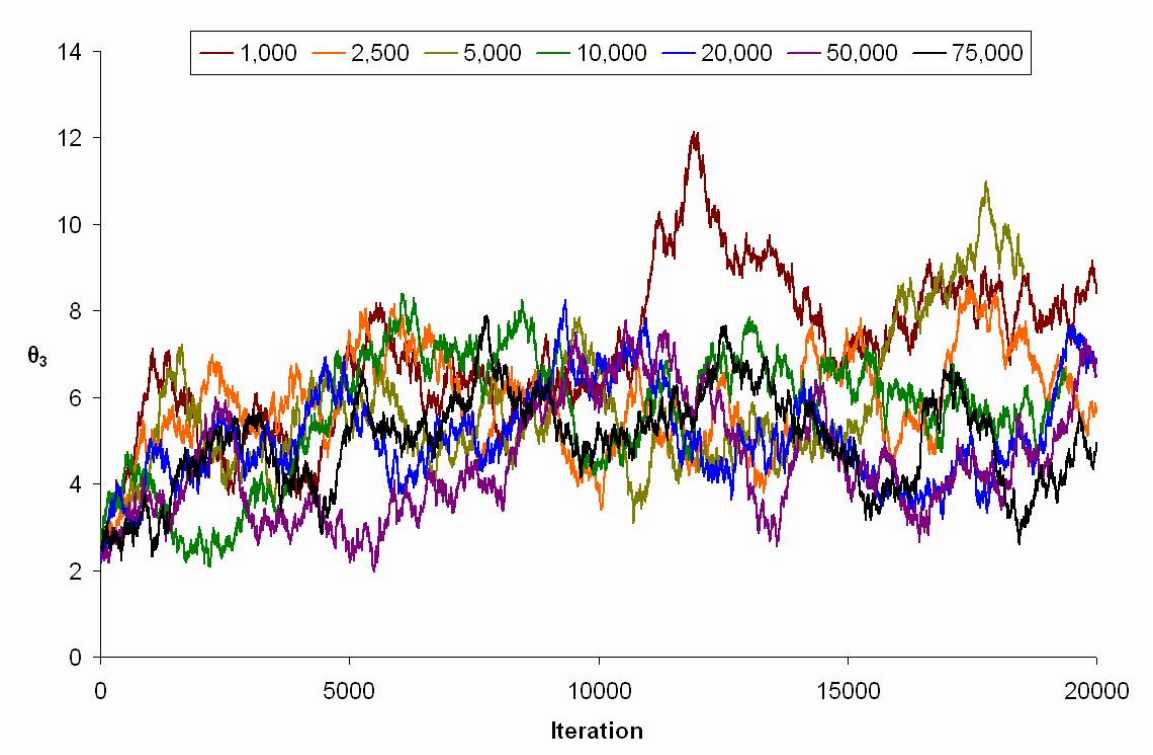

Figure 81. Trace plot of the parameter θ_3 of the SPTP model for simulated data on a 10 x 10 plot. Sub-chain burn-in ranges from 1,000 iterations (red) to 75,000 iterations (black).

Table 5. Posterior modes and acceptance rates from Bognar's algorithm with SPTP model for simulated data on 10 x 10 plot. Sub-chain burn-in ranges from 1,000 iterations to 75,000 iterations.

Sub-chain burn-in	*SPTP parameter*				*Acceptance*
(iterations)	β	θ_1	θ_2	θ_3	*Rate (%)*
1,000	1.26	3.00	-1.55	6.63	74.6
2,500	2.48	3.11	-1.34	5.98	72.9
5,000	1.71	2.78	-1.36	5.04	71.7
10,000	2.78	3.01	-1.40	6.92	69.9
20,000	2.85	3.14	-1.09	4.60	70.4
50,000	2.80	2.97	-1.08	3.99	67.9
75,000	0.96	2.68	-1.53	5.66	68.8

kernel density estimator (Scott 1992, p. 152) was used to smooth the histograms of the parameter values from each run, with the bandwidth given by the normal reference rule

$$h_i = \left(\frac{4}{d+2} \right)^{1/(d+4)} \sigma_i \; n^{-1/(d+4)}$$

where in this application d =4 is the dimension of the multivariate smoother, σ_i is the standard deviation of the i^{th} variable being smoothed, and n is the number of samples. Since Bognar's algorithm was started on the true parameter values, the burn-in for Bognar's algorithm was assumed to be zero iterations, implying that n = 20,000 samples would be available for the multivariate normal smoother for each run. After smoothing, the mode of the joint posterior distribution was obtained for each run (Table 5), in

addition to the marginal modes and marginal densities for each parameter. For all runs, the joint posterior modes reflect overestimation of θ_1 (target value = 2.5), underestimation of θ_2 (target value = -1.0), and overestimation of θ_3 (target value = 2.5). The overestimation of θ_3 is particularly large in magnitude, with the 10,000 and 1,000 iteration sub-chain burn-in runs showing the largest overestimation at 6.92 and 6.63, respectively.

Despite the large numerical differences between the estimated posterior modes and the true parameter values that generated the SimTest50 data, simulation of the SPTP process with the posterior modes for all sub-chain burn-in runs using Geyer and Møller's algorithm resulted in Monte Carlo envelopes of *L*-functions that were consistent with the *L*-function calculated from the SimTest50 data. Using the posterior mode from the 5,000 sub-chain burn-in run as an example, the Monte Carlo envelope of the *L*-function from the SPTP process shows strong attraction at 0.5m combined with strong repulsion at 2m that closely mimics the SimTest50 data (Fig. 82). The Monte Carlo envelope from the 50,000 burn-in run is only slightly more repulsive across all distances up to 3m (Fig. 83), owing to the greater degree of attractiveness in the pair potential of the 5,000 sub-chain burn-in run ($\theta_2 = 2.78$, $\theta_3 = -1.36$). as compared to the 50,000 sub-chain burn-in run ($\theta_2 = 2.97$, $\theta_3 = -1.08$). Nevertheless, the Monte Carlo envelope encloses the *L*-function from the SimTest50 data at all distances up to 3m. Of all the burn-in runs, only the 2,500 iteration run generated a Monte Carlo envelope that did not enclose the SimTest50 *L*-function at all distances up to 3m. In this case, the Monte Carlo envelope was too attractive at distances from 0.3 to 0.5m, but otherwise enclosed the *L*-function of the SimTest50 data.

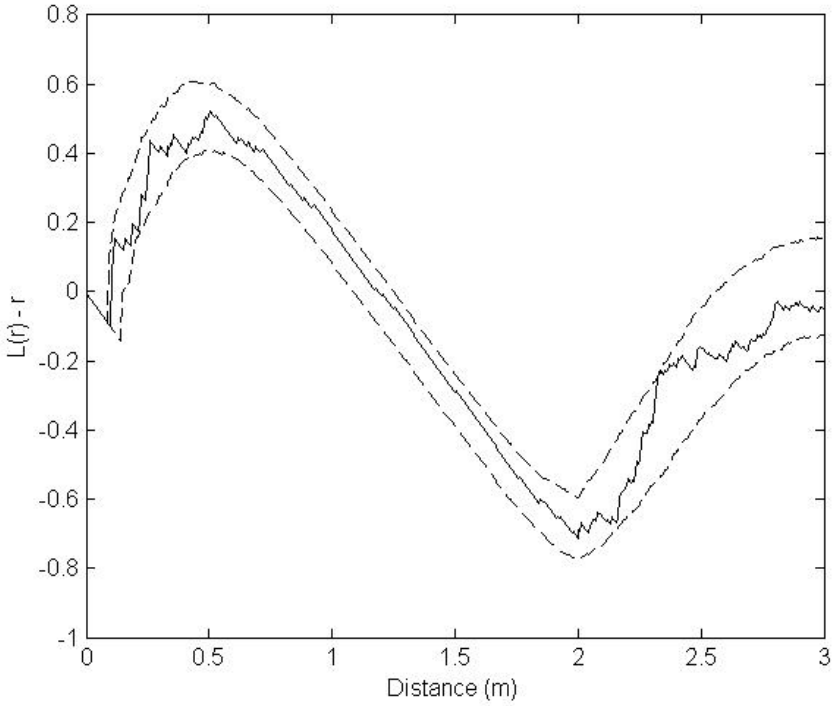

Figure 82. Monte Carlo envelope of *L*-function of SPTP model using the posterior mode for the 5,000 iteration burn-in against the simulated 10 x 10 "plot".

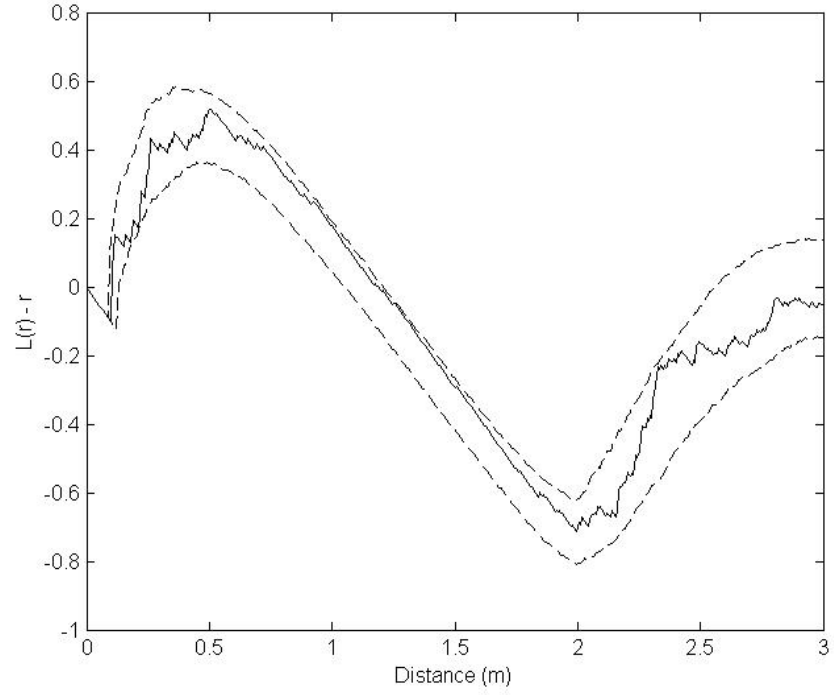

Figure 83. Monte Carlo envelope of *L*-function of SPTP model using the posterior mode for the 50,000 iteration burn-in against the simulated 10 x 10 "plot".

To further explore the effect of sub-chain burn-in on the performance of Bognar's algorithm, MATLAB code was developed to perform two sample *t*-tests on the differences of the estimated normalizing constant ratio and corresponding Metropolis-Hastings acceptance probabilities generated from moving from the true parameter vector of ($\beta = 2.5$, $\theta_1 = 2.5$, $\theta_2 = -1$, $\theta_3 = 2.5$) to other proposed parameter vectors. For each proposed parameter vector, Geyer and Møller's algorithm was executed 100 times to generate samples of the estimated normalizing constant ratios and acceptance probabilities of moving from the true parameter vector to the proposed parameter vector using a 5,000 iteration and 50,000 iteration burn-in. For both burn-in choices, 100 importance samples spaced 100 iterations apart were drawn from the next 10,000 iterations following burn-in. For comparison, the resulting mean estimates and two sample *t*-statistics (assuming equal variances) are grouped by the qualitative change in the 2nd-order potential (more or less attractive) and in the 3rd-order potential (more or less repulsive) by using the proposed parameter vector as compared to the target parameter vector (Table 6). When the proposed parameter vector has a more attractive 2nd-order potential (Groups 1 and 2), the mean normalizing constant ratio is significantly larger using a 5,000 iteration burn-in of Geyer and Møller's algorithm than with a 50,000 iteration burn-in, regardless of whether the 3rd-order potential was proposed to be more or less repulsive. Mean acceptance probabilities were also significantly larger using a 5,000 iteration burn-in when the parameter θ_2 was proposed as -1.1 (implying more attractiveness), regardless of the change in the 3rd-order potential through parameter θ_3. When the parameter θ_2 was decreased to -1.5, the mean acceptance probabilities were

Table 6. Mean log normalizing constant ratio, acceptance probabilities, and two sample t-tests for differences of means generated from 100 trials of Bognar's algorithm with sub-chain burn-in (SCBI) of 5,000 and 50,000 iterations. "Proposed" parameters compared against the "current state" $(\beta = 2.5,\ \theta_1 = 2.5,\ \theta_2 = -1,\ \theta_3 = 2.5)$.

Group	*Proposed parameter vector* $(\beta,\ \theta_1,\ \theta_2,\ \theta_3)$	*Mean log normalizing constant ratio* SCBI = 5,000	SCBI = 50,000	t-statistic (p-value)	*Mean acceptance probability* SCBI = 5,000	SCBI = 50,000	t-statistic (p-value)
1	(2.5, 2.5, -1.1, 2.6)	-12.39	-12.87	3.73 (0.0003)	1.1053	0.8065	2.56 (0.0113)
1	(2.5, 2.5, -1.5, 3.0)	-90.92	-98.24	8.77 (< .0001)	1.77×10^{-9}	3.87×10^{-12}	1.12 (0.2639)
2	(2.5, 2.5, -1.1, 2.4)	-16.55	-17.12	3.59 (0.0004)	0.2039	0.1396	2.17 (0.0316)
2	(2.5, 2.5, -1.5, 2.0)	-168.52	-196.00	16.14 (< .0001)	9.49×10^{-32}	3.92×10^{-42}	1.00 (0.3197)
3	(2.5, 2.5, -0.9, 2.6)	11.77	11.93	-1.38 (0.1697)	0.0955	0.1133	-1.14 (0.2553)
3	(2.5, 2.5, -0.5, 3.0)	33.37	33.07	1.06 (0.2912)	1.02×10^{-16}	1.13×10^{-16}	-0.14 (0.8862)
4	(2.5, 2.5, -0.9, 2.4)	9.96	10.19	-2.29 (0.0233)	0.1318	0.1898	-2.73 (0.0071)
4	(2.5, 2.5, -0.5, 2.0)	32.94	32.95	-0.03 (0.9787)	7.9×10^{-12}	1.73×10^{-11}	-0.96 (0.3368)

Groups: 1, 2nd order potential more attractive, 3rd order potential more repulsive;
2, 2nd order potential more attractive, 3rd order potential less repulsive;
3, 2nd order potential less attractive, 3rd order potential more repulsive;
4, 2nd order potential less attractive, 3rd order potential less repulsive.

larger using a 5,000 iteration burn-in, but since the acceptance probabilities were so close to zero that the differences were not statistically significant. In contrast, when the 2nd-order potential was proposed to be less attractive (Groups 3, 4), differences in the mean normalizing constant ratio and mean acceptance probabilities were not statistically significant, with the sole exception being the case with the proposed parameter vector (2.5, 2.5, -0.9, 2.4), implying less attractive 2nd-order potentials and less repulsive 3rd-order potentials. In this case, using a 5,000 iteration burn-in resulted in a significantly smaller mean normalizing constant ratio and mean acceptance probability. While this effect is not easily seen in the trace plots from the trials with the Simtest50 dataset, the results here indicate that using a small (5,000 iteration) burn-in with Geyer and Møller's algorithm to draw importance samples will be more likely to accept proposed parameter vectors with increased attraction in the 2nd-order potential than when a larger (50,000 iteration) burn-in is used.

Given the slow mixing of SPTP process with the SimTest50 data, the question was raised whether point density of the data was influential in the mixing of the Metropolis-Hasting sampler of the Gibbs posterior distribution. Preliminary trials (with 6,000 or fewer total iterations) with the SPTP process and the Coweeta Plot 318 data indicated very slow mixing as well, despite trying different values for the tuning parameters and the use of a multivariate normal proposal distribution with covariance as well as variance tuning parameters. It was questioned whether the low point density on Plot 318 (0.0545 trees per square meter) had an effect on the posterior sampler mixing. To investigate the possibility, a simulation study was conducted using the SPTP process using $\theta_1 = 2.5$, $\theta_2 = -1.0$, and $\theta_3 = 2.5$ and an interaction distance of 2m. As evidenced

from Ex. 2 of the simulation of the SPTP process (Figs. 60, 61), these parameter values result in a process with moderate 2nd-order attraction and moderate 3rd-order repulsion (Figs. 64, 65). Different point densities were obtained by choosing values of 0.75, 4.5, 15, and 100 for β to simulate the SPTP process involving 50 points on 12 x 12, 9 x 9, 8 x 8, and 7 x 7m plots. A post-burn-in sample configuration was obtained for each plot size using Geyer and Møller's algorithm. The four sample configurations were used together with the SimTest50 dataset, where the plot size was 10 x 10m, to test the behavior of Bognar's algorithm for inference with the SPTP process. For each of the five plot sizes, Bognar's algorithm was initialized with parameter values that had been used to generate each respective dataset. As before, 100 importance samples are drawn at a spacing of 100 iterations using a 50,000 iteration burn-in. The resulting trace plots (Figs. 83 – 86) showed slow mixing for all SPTP parameters across all point densities. Trace plots for the parameter β showed that at the lowest point densities tested, the trace plots did not appear to mix around any particular values, with the trace plot for the 9 x 9 dataset (target $\beta = 4.5$) crossing the trace plot for the 10 x 10 (target $\beta = 2.5$). While trace plots for parameter θ_1 had no clear trend present, the trace plots for θ_2 indicated that the lowest plot densities had the worst mixing, with the run using the 12 x 12 dataset generating sample values of θ_2 as small as -2.5, in comparison to the known target value of $\theta_2 = -1.0$ (Fig. 86). Trace plots for θ_3 also indicated that the lower plot densities mixed more slowly and sampled more extreme values than those at the higher plot densities tested (Fig. 87).

After Bognar's algorithm was completed for each plot size, all 20,000 iterations

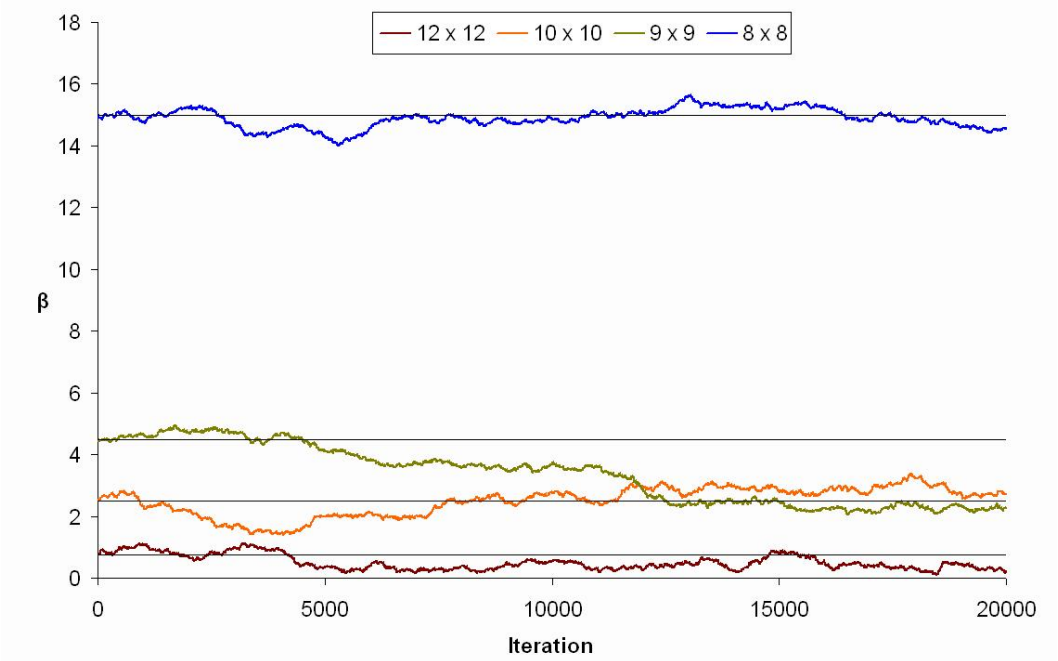

Figure 84. Trace plot of the parameter β of the SPTP model for simulated data with varying point density. Plot sizes displayed are 8 x 8, 9 x 9, 10 x 10, and 12 x 12m. Horizontal black lines indicate known target values of β.

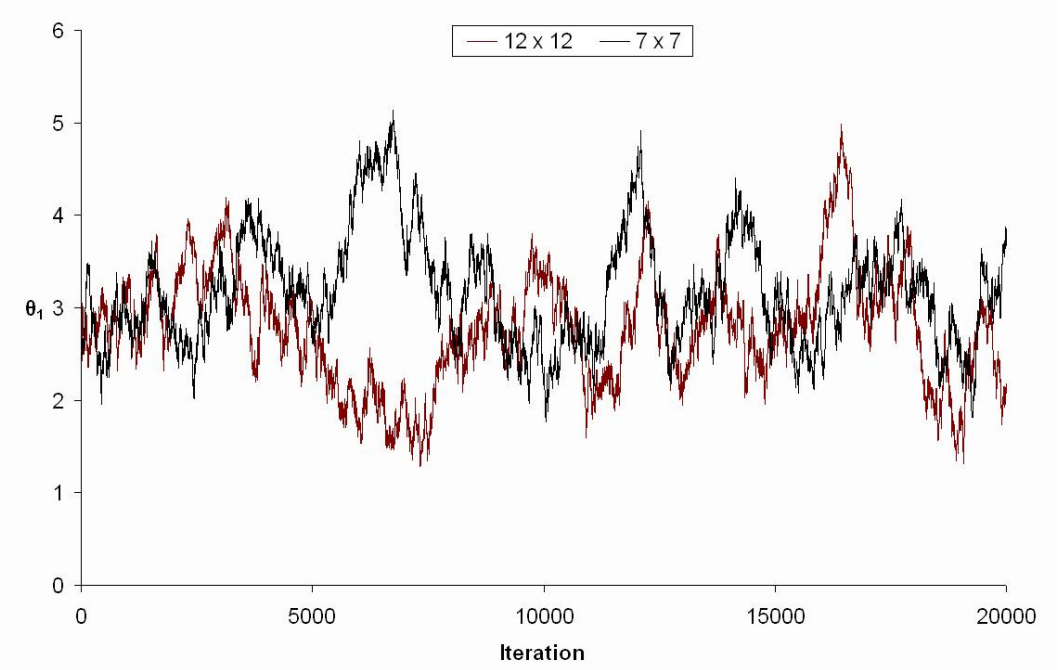

Figure 85. Trace plot of the parameter θ_1 of the SPTP model for simulated data with varying point density. Plot sizes displayed are 7 x 7 and 12 x 12m.

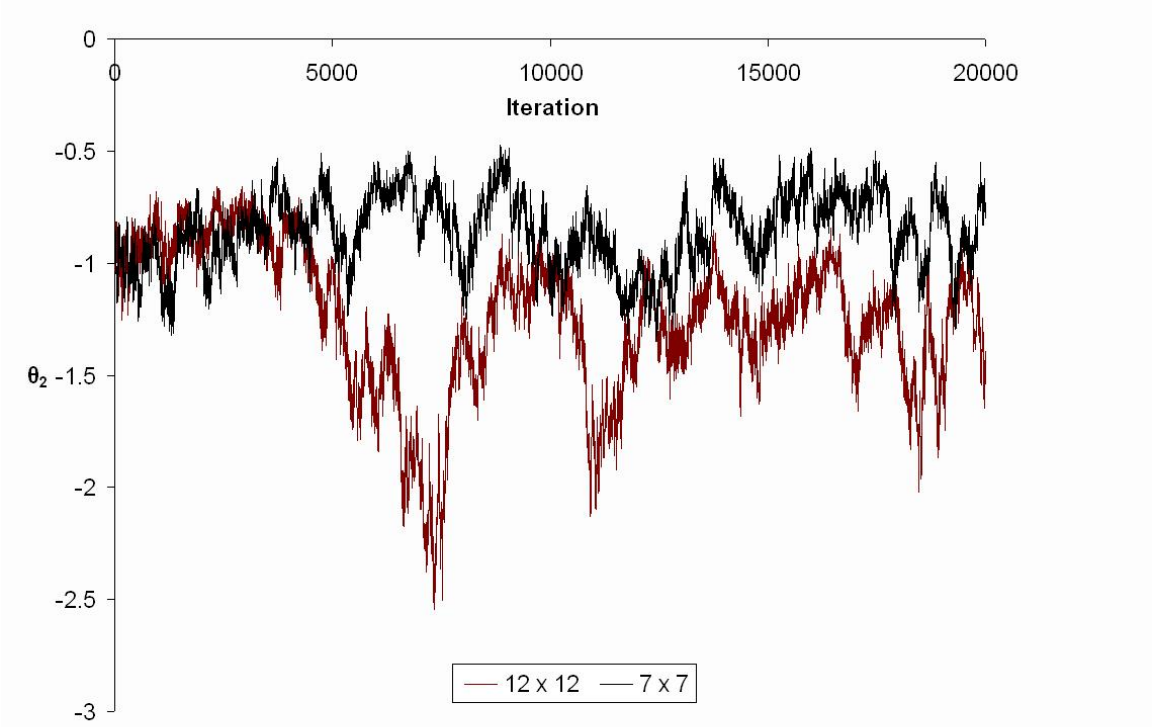

Figure 86. Trace plot of the parameter θ_2 of the SPTP model for simulated data with varying point density. Plot sizes displayed are 7 x 7 and 12 x 12m.

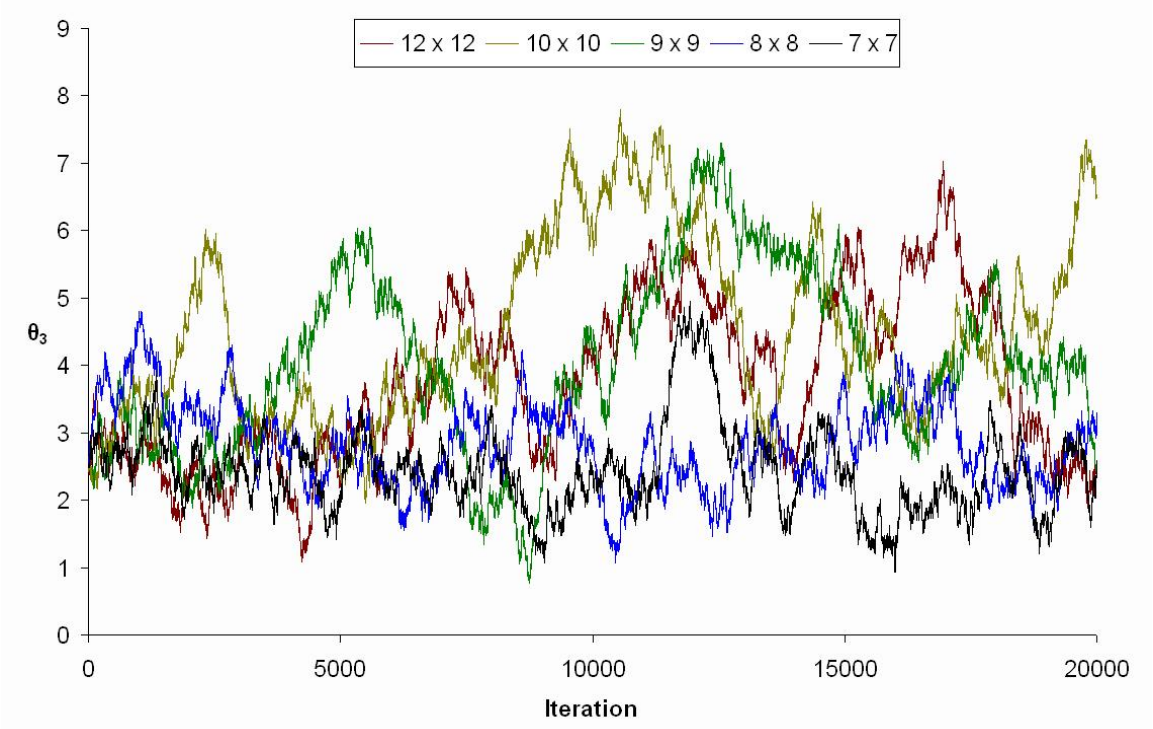

Figure 87. Trace plot of the parameter θ_3 of the SPTP model for simulated data with varying point density. Plot size ranges from 7 x 7 to 12 x 12m.

Table 7. Posterior modes and acceptance rates from Bognar's algorithm with SPTP model with varying point density. Plot sizes range from 7 x 7 to 12 x 12m.

Point density	*Plot size*	*SPTP parameter*				*Acceptance*
(Points/m^2)	*(m x m)*	β	θ_1	θ_2	θ_3	*Rate (%)*
0.3264	12 x 12	0.34	2.87	-1.10	2.59	68.0
0.5100	10 x 10	2.80	2.97	-1.08	3.99	67.9
0.5926	9 x 9	2.18	2.69	-1.24	3.87	68.0
0.7500	8 x 8	14.49	3.95	-0.78	2.87	64.1
1.0000	7 x 7	99.98	4.52	-0.68	2.43	63.8

from each run were used to estimate joint and marginal posterior densities for each parameter using the multivariate and univariate normal kernel smoothers. The joint posterior modes indicate that the lower point densities (12 x 12, 10 x 10, and 9 x 9 plots) led to greater under-estimation of θ_2 and greater overestimation of θ_3 (Table 7). Surprisingly, the greatest overestimation of θ_1 (4.52) occurred with the highest point density dataset. Plotting the estimated joint posterior density in the θ_1 − dimension against the marginal posterior for θ_1 revealed that while the marginal density is mound shaped with a mode at approximately 3, the joint posterior is bi-modal, with most of the sampled values for θ_1 occurring at a smaller mode around 3 rather than the greater mode at 4.5 (Fig. 88).

As in the simulation study of the effect of sub-chain burn-in on the SPTP process, the numerical differences between the joint posterior modes and the true parameter values

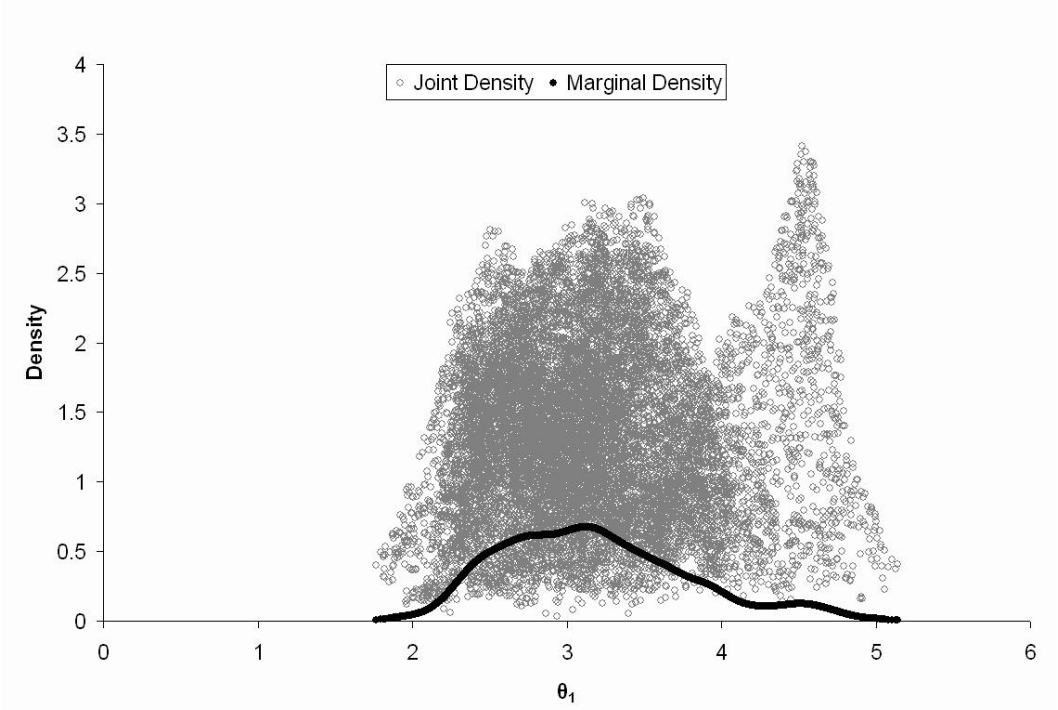

Figure 88. Joint and marginal posterior density estimation for the parameter θ_1 of the SPTP model. Simulation from the 7 x 7 plot.

did not lead to substantially different spatial structures between simulations of the SPTP process fitted with the posterior modes and the data that was used to test Bognar's algorithm. For all point densities except for the 7 x 7m dataset, the Monte Carlo envelope completely enclosed the *L*-function from the data used to test Bognar's algorithm, yet their was also a tendency for these Monte Carlo envelopes to become wider as point density was lower (Figs. 88 – 89). In the case of the 7 x 7m plot configuration, the Monte Carlo envelope simulated for the joint posterior mode was more attractive than the test configuration's *L*-function at distances less than 0.5m. It was also more repulsive at distances between 1.5 and 2.25m (Fig. 90) than the test configuration's *L*-function. The

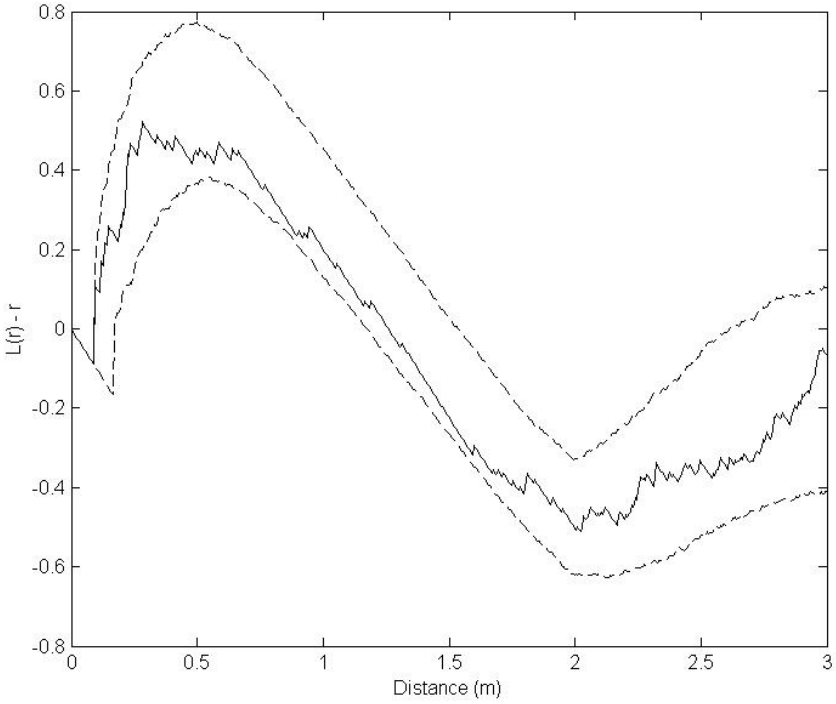

Figure 89. Monte Carlo envelope of *L*-function of SPTP model using the posterior mode against the simulated 12 x 12 plot.

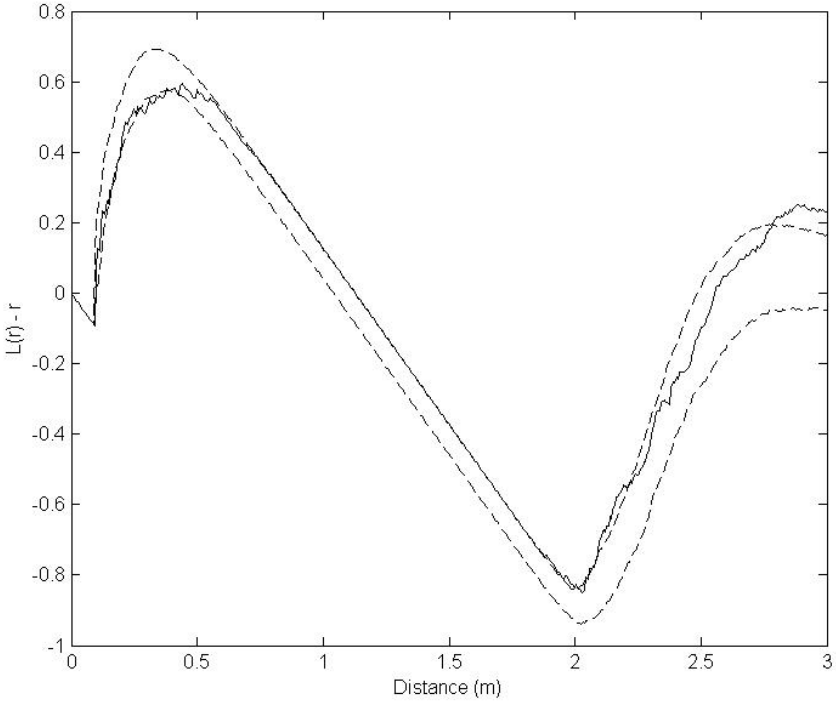

Figure 90. Monte Carlo envelope of *L*-function of SPTP model using the posterior mode against the simulated 7 x 7 plot.

joint posterior mode for the 7 x 7m plot (β = 99.98, θ_1 = 4.52, θ_2 = −0.68, θ_3 = 2.43) yielded a pair potential for the SPTP process that when plotted, showed greater repulsion and less attraction than did the SPTP pair potential plot with any other plots' posterior mode solution.

Following completion of the test of point density with the SPTP process, the performance of Bognar's algorithm was tested for spatial structures other than the relatively strongly clustered patterns given by the SPTP process with θ_1 = 2.5, θ_2 = -1.0, and θ_3 = 2.5. To this end the EXPTP process was chosen for testing Bognar's algorithm using two different parameter vectors; one representing weak localized point attraction (β = 0.0850, θ_1 = −3.5, θ_2 = 2) and the other representing purely repulsive point interactions (β = 10.0, θ_1 = 1, θ_2 = 1). As in the tests with the SPTP process, simulated data were used to generate datasets with approximately 50 points on a 10 x 10m plot with the interaction distance set to 2m. When simulated with Geyer and Møller's algorithm, the EXPTP process with weak localized attraction converged in less than 10,000 iterations to a process with about 60 – 65 points (Figs. D1 – D3). Although this point density was somewhat higher than the target density of 50 points, a sample with n = 64 points drawn at the 100,000th iteration (Fig. 91) was used as a test configuration, referred to here as the SimTest64 dataset. Given the rapid burn-in of the EXPTP process with these parameter values, Bognar's algorithm was set to use a 15,000 iteration sub-chain burn-in, with 100 importance samples being drawn at a spacing of 50 iterations from the next 5,000 iterations. Bognar's algorithm was initialized with the target parameter of (β = 0.0850, θ_1 = −3.5, θ_2 = 2) and tuning parameters set as $\sigma_\beta = 0.01$, $\sigma_{\theta_1} = 0.05$, and

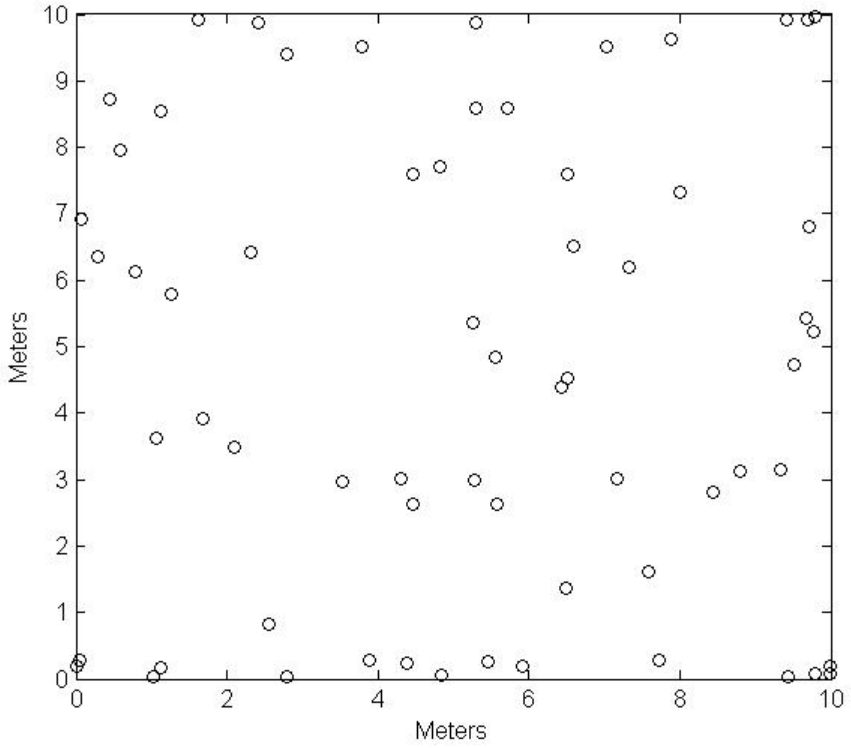

Figure 91. Configuration of the SimTest64 dataset, generated from the EXPTP process with parameter values $(\beta = 0.0850,\ \theta_1 = -3.5,\ \theta_2 = 2,\ R = 2)$.

$\sigma_{\theta_2} = 0.05$. After completing the 20,000 iteration run, the algorithm had attained a 72.9% acceptance rate. In spite of this high acceptance rate, mixing for all parameters was noticeably improved from previous trials with the SPTP process. The trace plots for θ_1, θ_2, and θ_3 all oscillated around their respective true parameter values (Figs. 91 – 93). While the trace plot for β showed reasonable mixing, it consistently overestimated this parameter (Fig. 92). Using all 20,000 iterations for analysis, the smoothed posterior mode was found as $\left(\widehat{\beta} = 0.1075,\ \widehat{\theta}_1 = -3.30,\ \widehat{\theta}_2 = 1.89\right)$. A comparison of the joint posterior

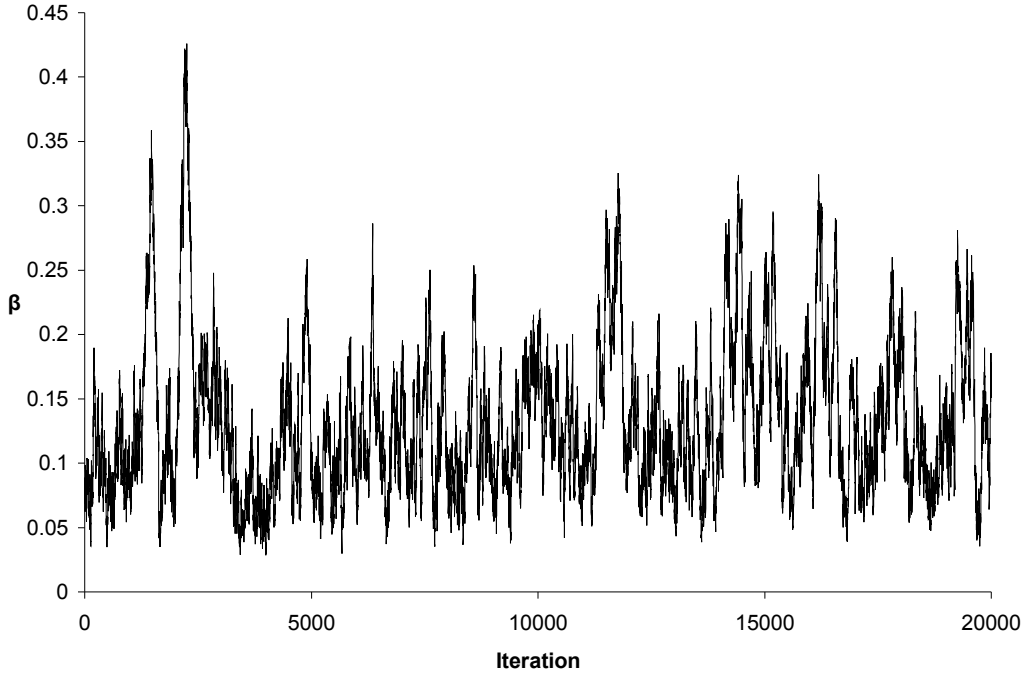

Figure 92. Trace plot of the parameter β of the EXPTP model for simulated data on a 10 x 10 plot. Known value of β is 0.0850.

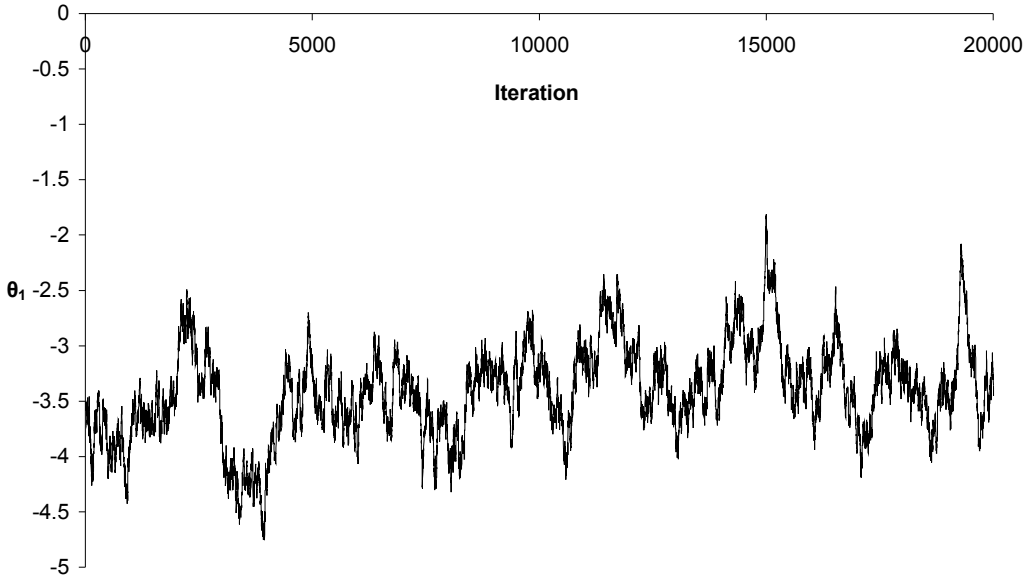

Figure 93. Trace plot of the parameter θ_1 of the EXPTP model for simulated data on a 10 x 10 plot. Known value of θ_1 is -3.5.

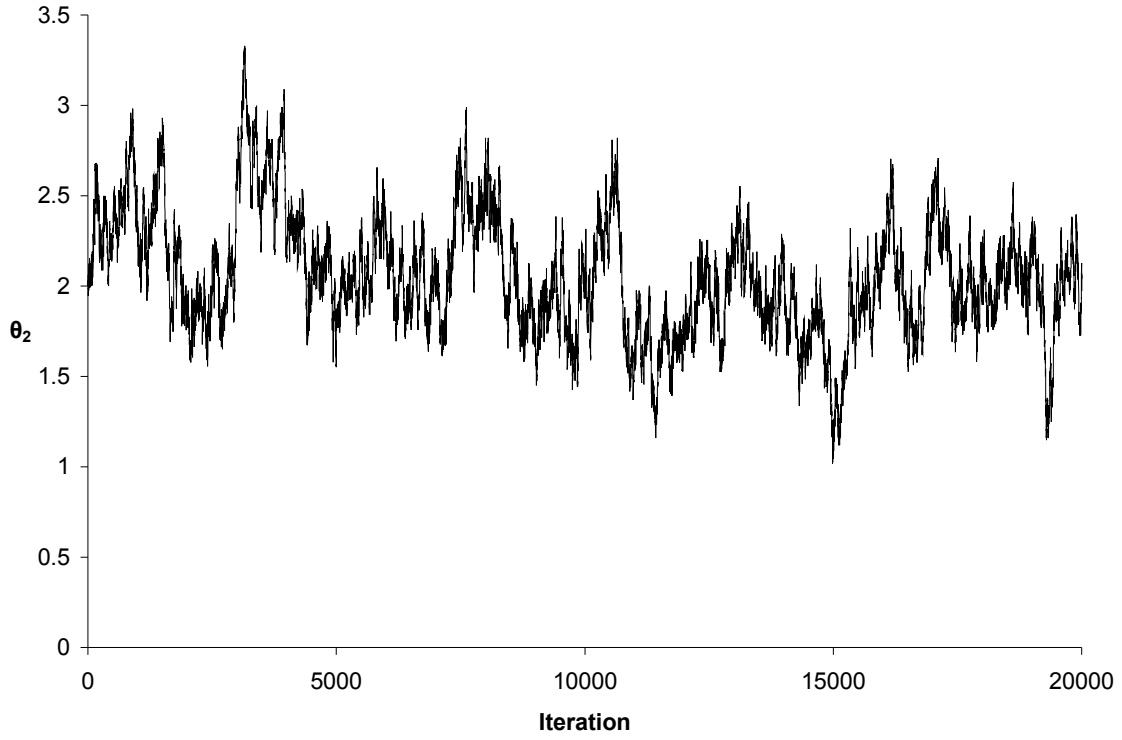

Figure 94. Trace plot of the parameter θ_2 of the EXPTP model for simulated data on a 10 x 10 plot. Known value of θ_2 is 2.0.

with the marginal posterior densities for the EXPTP process parameters shows consistency in joint and marginal mode estimates for all parameters (Figs. D4 – D6). Simulation of the EXPTP process with the joint posterior mode indicated consistent spatial structure with the SimTest64 dataset, as evidenced in the *L*-function, for distances up to the 2m interaction distance (Fig. 95). From approximately 2.25 to 2.75m, the *L*-function of the SimTest64 data shows greater repulsion in points than the Monte Carlo envelope for the EXPTP process based on estimates from the posterior mode (Fig. 95).

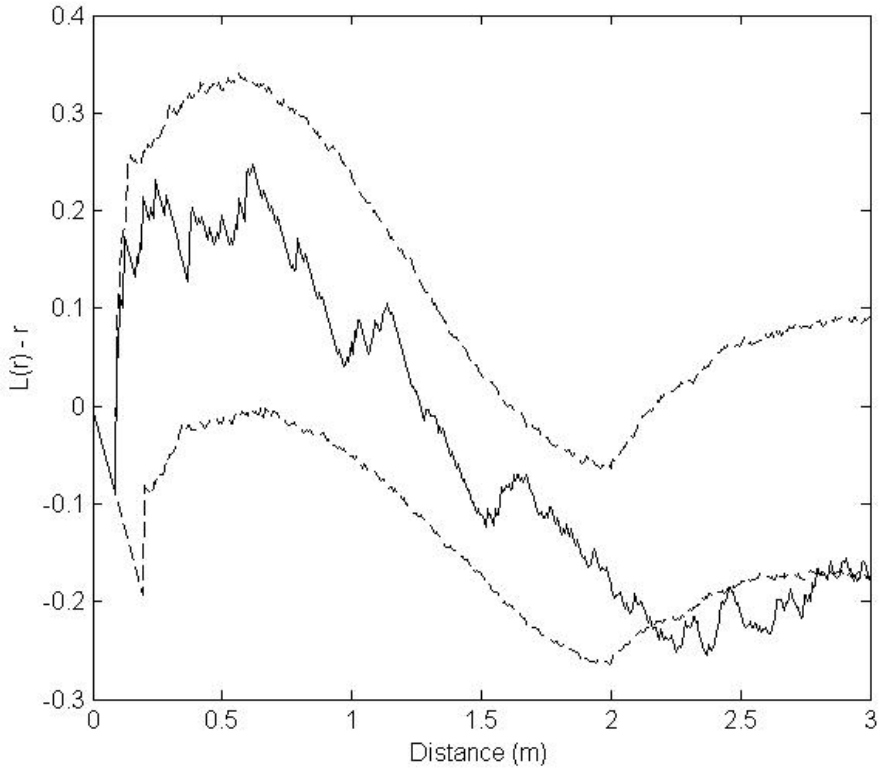

Figure 95. Monte Carlo envelope of *L*-function of EXPTP model using the posterior mode against the configuration SimTest64 with weak locally attractive spatial structure.

To consider the performance of Bognar's algorithm for Bayesian inference with a point configuration showing purely repulsive point interactions, a simulation trial was undertaken using the EXPTP process with the parameters $\beta = 10.0$, $\theta_1 = 1.0$, and $\theta_2 = 1.0$. The interaction distance was fixed at 2m with the EXPTP process parameters chosen to give a point density of approximately 50 points on a 10 x 10m plot. Simulation of the EXPTP process with these parameters indicated rapid convergence (< 5,000 iterations) to a process with approximately 50 points (Figs. D7 – D9). After a 10,000 iteration burn-in, a sample was drawn with a repulsive pattern containing $n = 51$ points (Fig. 96), denoted

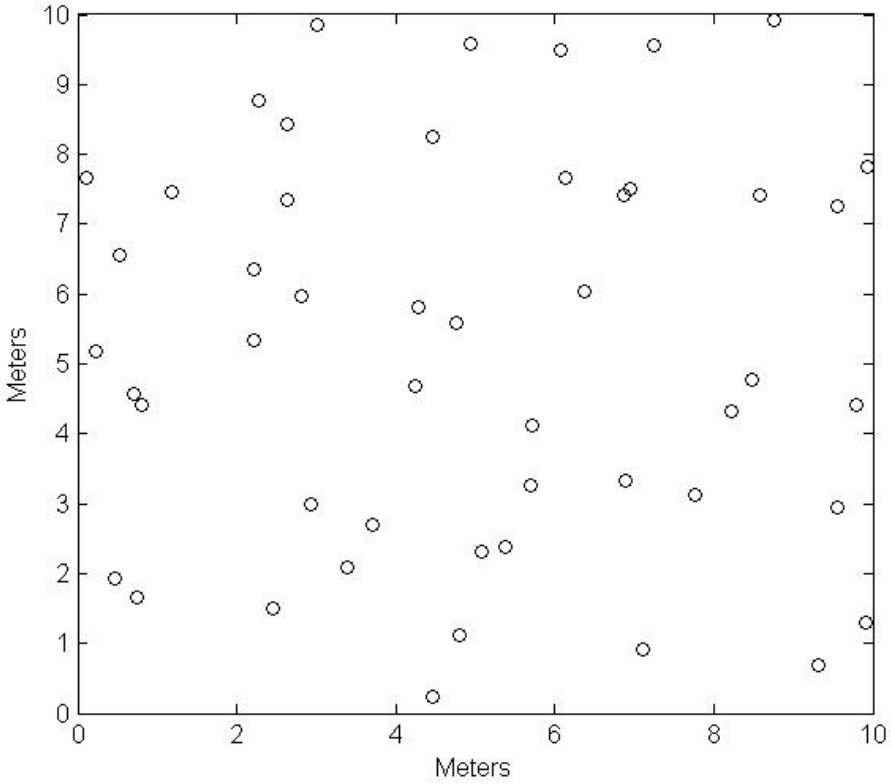

Figure 96. Configuration of the SimTest51 dataset, generated from the EXPTP process with parameter values $(\beta = 10,\ \theta_1 = 1,\ \theta_2 = 1,\ R = 2)$.

the SimTest51 dataset. To draw samples from the posterior distribution of the Gibbs process, Bognar's algorithm was initialized at the known parameter values (β = 10.0, θ_1 = 1, θ_2 = 1) and run for 20,000 iterations. For the trial, a 10,000 iteration sub-chain burn-in is used, with 100 importance samples being drawn at a spacing of 50 iterations from the next 5,000 iterations. As with the previous trial with the EXPTP process, tuning parameters were set as $\sigma_\beta = 0.01$, $\sigma_{\theta_1} = 0.05$, and $\sigma_{\theta_2} = 0.05$.

Trace plots from the 20,000 iteration run indicated slow mixing (Figs. 96 – 98),

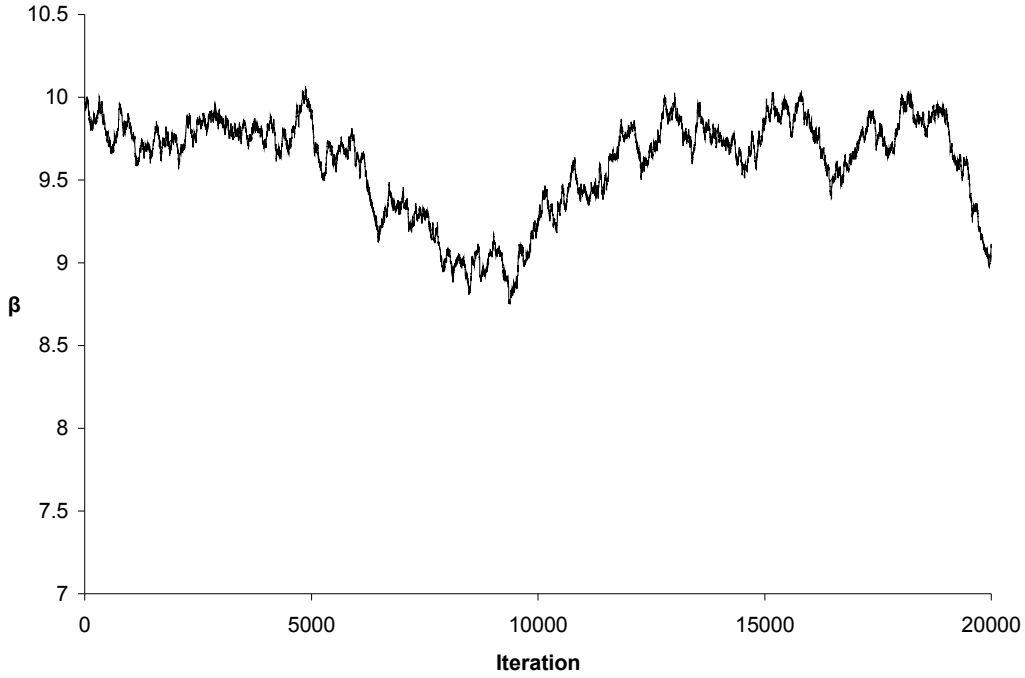

Figure 97. Trace plot of the parameter β of the EXPTP model for simulated data on a 10 x 10 plot with purely repulsive spatial structure. Known value of β is 10.

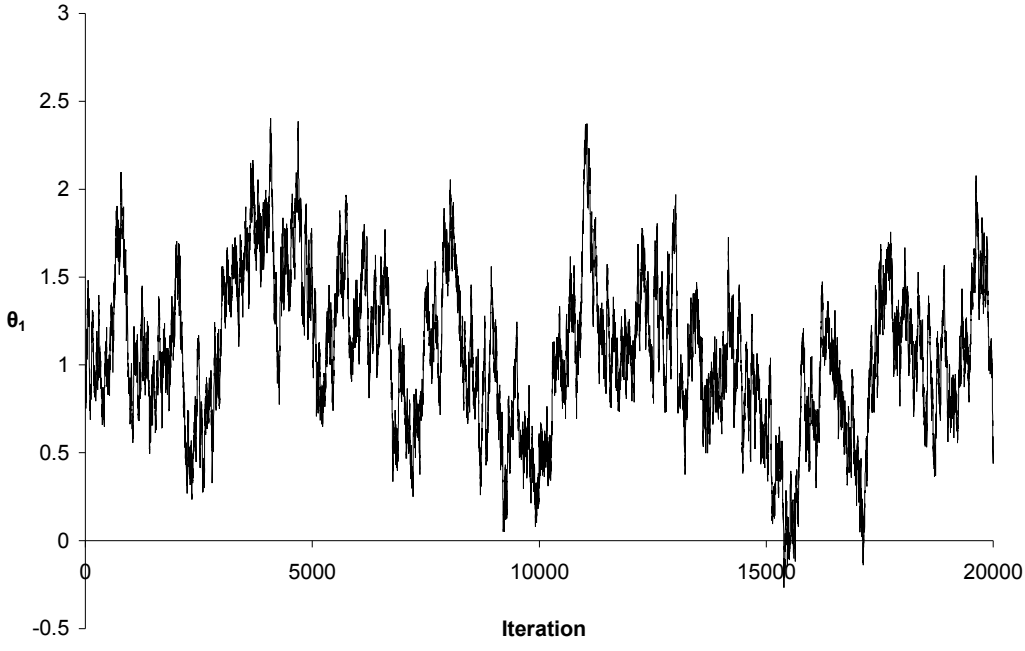

Figure 98. Trace plot of the parameter θ_1 of the EXPTP model for simulated data on a 10 x 10 plot with purely repulsive spatial structure. Known value of θ_1 is 1.

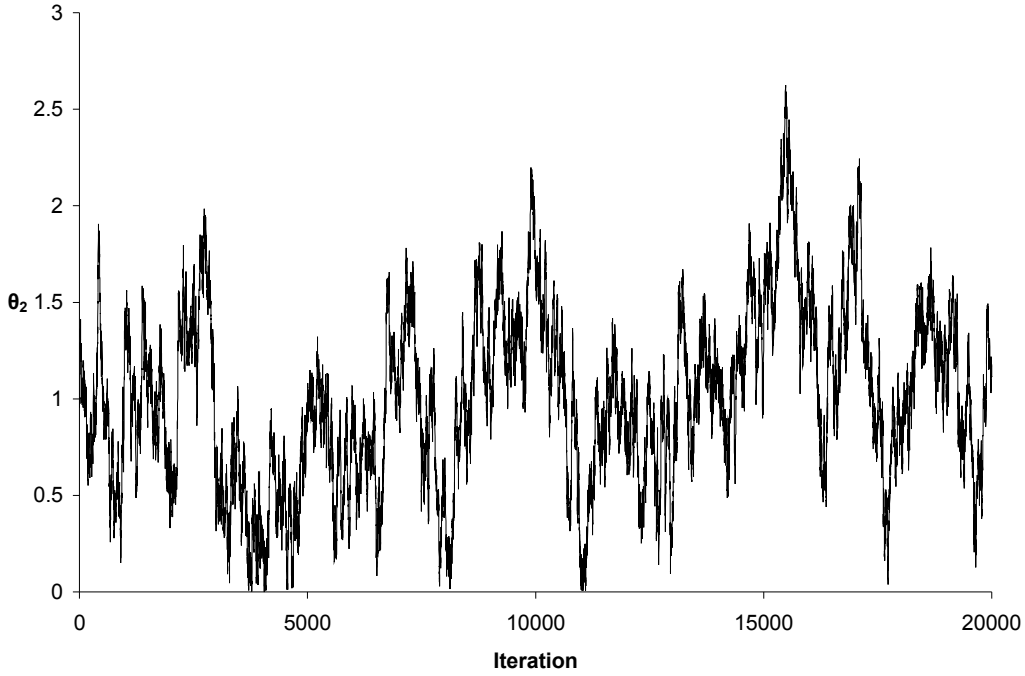

Figure 99. Trace plot of the parameter θ_2 of the EXPTP model for simulated data on a 10 x 10 plot with purely repulsive spatial structure. Known value of θ_2 is 1.

particularly with the intensity parameter, β. The run had an exceedingly high acceptance rate (90.53%), further indicating slow mixing, Nevertheless, the trace plots for θ_1 and θ_2 appeared to oscillate around the known values used to generate the test configuration Simtest51. Using the multivariate normal smoother, the estimated posterior mode is $\left(\widehat{\beta} = 9.7025,\ \widehat{\theta_1} = 1.10,\ \widehat{\theta_2} = 0.89\right)$. Estimated marginal modes are all consistent with the joint posterior mode (Figs. D10 – D12) and are numerically close to the true parameter values. Moreover, when the EXPTP process was simulated using the joint posterior mode, the Monte Carlo envelope of the *L*-function enclosed the *L*-function of the dataset under inference at practically all distances (Fig. 100).

The next simulation test undertaken was designed to look at the effect of the

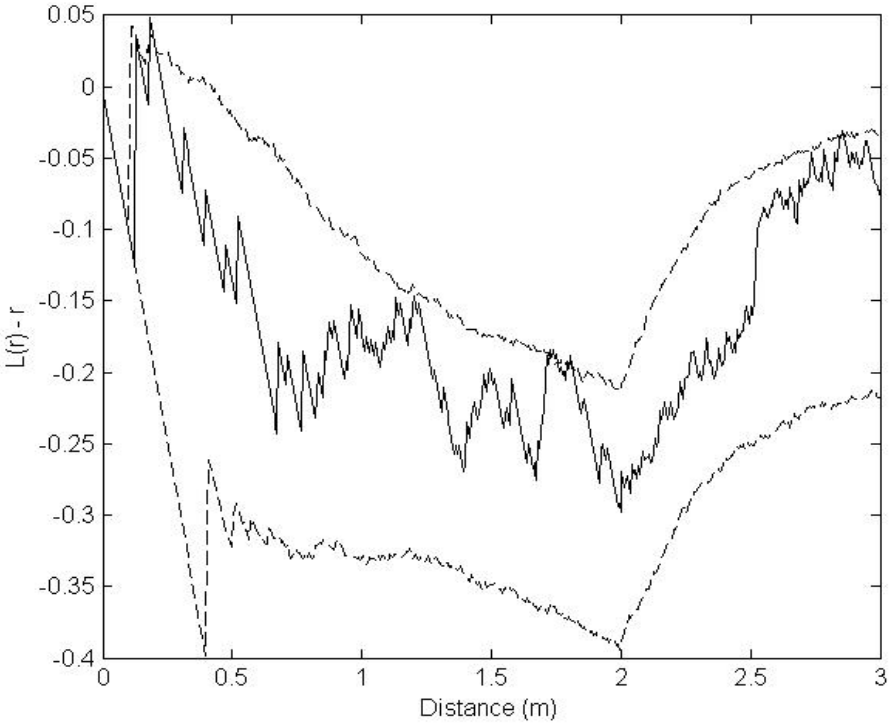

Figure 100. Monte Carlo envelope of *L*-function of EXPTP model using the posterior mode against the configuration SimTest51 with purely repulsive spatial structure.

choice of interaction distance on the performance of Bognar's algorithm. In the only published simulation study of triplets models, Mateu and Montes (2001) reported that Monte Carlo maximum likelihood, Monte Carlo Newton Raphson and stochastic approximation estimates of Geyer's triplets process had less bias and lower standard errors when interaction distances were small (0.05 and 0.10 units on a 1 unit square plot) than when large (0.15 units). In the simulation test here, a plot size of 30 x 30m is selected with interaction distances of 4, 6, 8, and 10m. Since the center 30 x 30m piece of Coweeta Plot 318 contains 60 points, the EXPTP process was selected with parameters $\beta = 0.0625$, $\theta_1 = -2.5$, and $\theta_2 = 3.5$ that would give sample configurations with approximately 60 points across the range of interaction distances studied combined with

an attractive spatial structure similar to that of Plot 318. The interaction distances were chosen based on a search of the competition literature in other hardwood forests in the eastern USA; one implication of this choice is that these interaction distances are comparatively much larger than what Mateu and Montes used in their simulation study.

For each chosen interaction distance, Geyer and Møller's algorithm was used to draw a sample from the EXPTP process. The resulting samples are referred to as the ST55_IR4, ST55_IR6, ST55_IR8, and ST55_IR10 datasets and when plotted contain pairs and triples of points scattered within an otherwise over-dispersed point pattern (Figs. D13, D15, D17, D19). In each case, Geyer and Møller's algorithm appeared to converge rapidly, with trace plots suggesting a burn-in of less than 5,000 iterations (Figs. D14, D16, D18, D20). However the EXPTP process with smaller interaction distances gives samples with larger number of points, where the 4m interaction distance gives samples with 60 – 65 points and the 10m interaction distance gives samples with about 40 points. To limit the effect of different point density on the simulation test, the samples for the ST55_IR datasets were chosen to have approximately 55 points, subject to the condition that they were representative samples, i.e. each test configurations' *L*-function fell entirely within the 95% Monte Carlo envelope of the configurations generated for each process.

To implement Bognar's algorithm, each run was initialized using the true parameter vectors that generated the ST55_IR data being analyzed. As before 100 importance samples were drawn at a spacing of 50 iterations following a 15,000 iteration burn. Tuning parameters for the EXPTP process were set at $\sigma_{\beta} = 0.01$, $\sigma_{\theta_1} = 0.05$, and $\sigma_{\theta_2} = 0.05$. As before, Bognar's algorithm was run for 20,000 iterations. During the

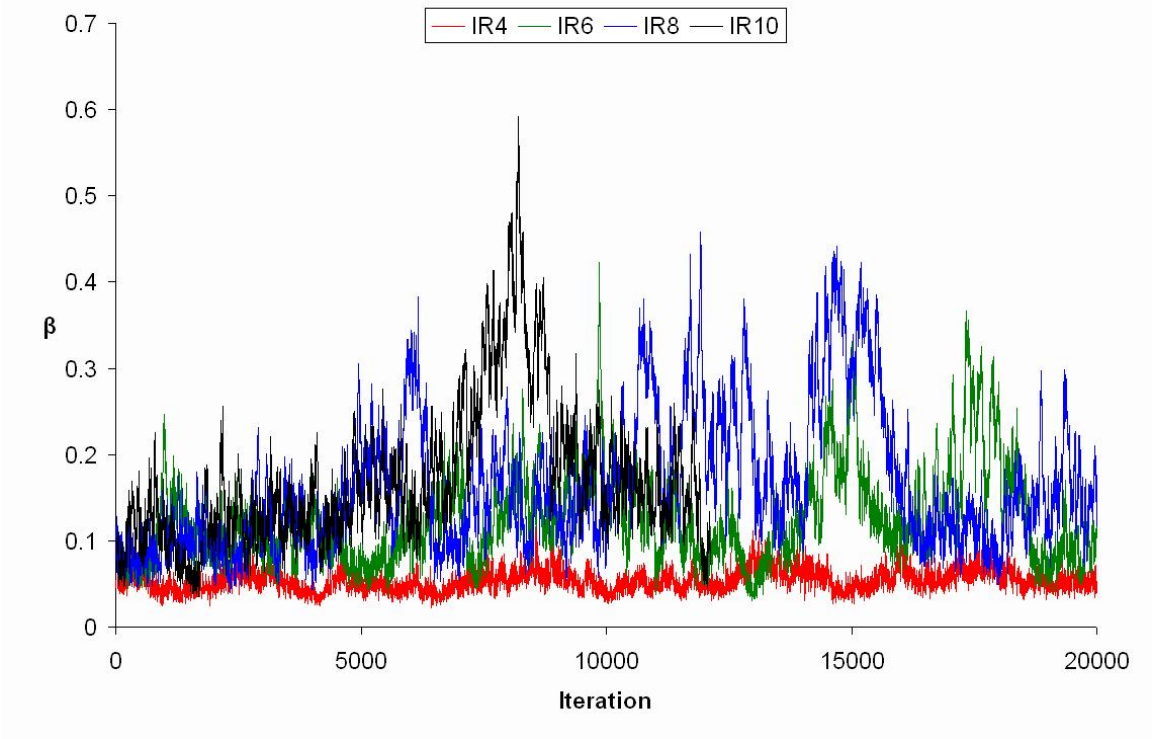

Figure 101. Trace plot of the parameter β of the EXPTP model for simulated data on a 30 x 30m plot. Interaction distance ranges from 4m (red) to 10m (black).

course of the trials, very large differences were found in the iteration times of Bognar's algorithm, increasing from 27.00 seconds with the 4m interaction distance to over 179.12 seconds with the 10m interaction distance. Because the full 20,000 iteration run of the 10m interaction distance trial would require a total of almost 1,000 hours of continuous running time (41.5 days), this trial was intentionally terminated after 12,122 iterations.

Trace plots of the EXPTP parameters obtained from each interaction distance trial indicated slow mixing for all trials (Figs. 100 – 102) . Larger interaction distance trials showed slower mixing, and for the parameter θ_2, only the 4m trial showed a trace plot that repeatedly explored the area around the true value of 3.5 (Fig. 103).

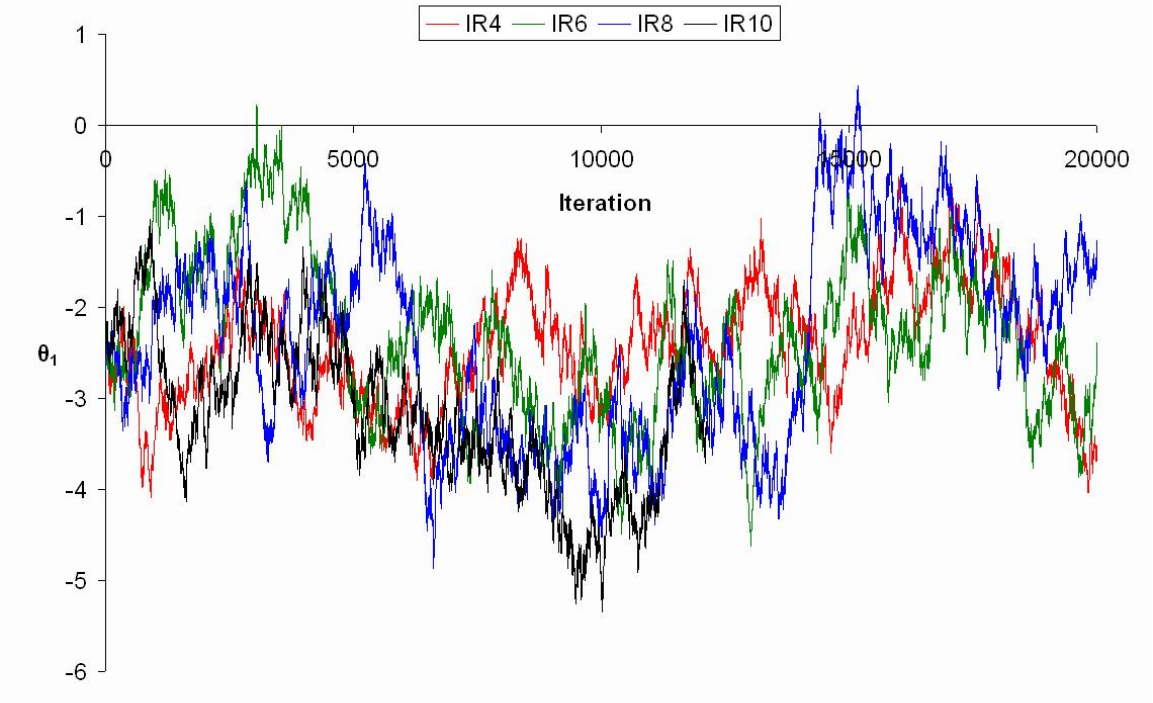

Figure 102. Trace plot of the parameter θ_1 of the EXPTP model for simulated data on a 30 x 30 plot. Interaction distance ranges from 4m (red) to 10m (black).

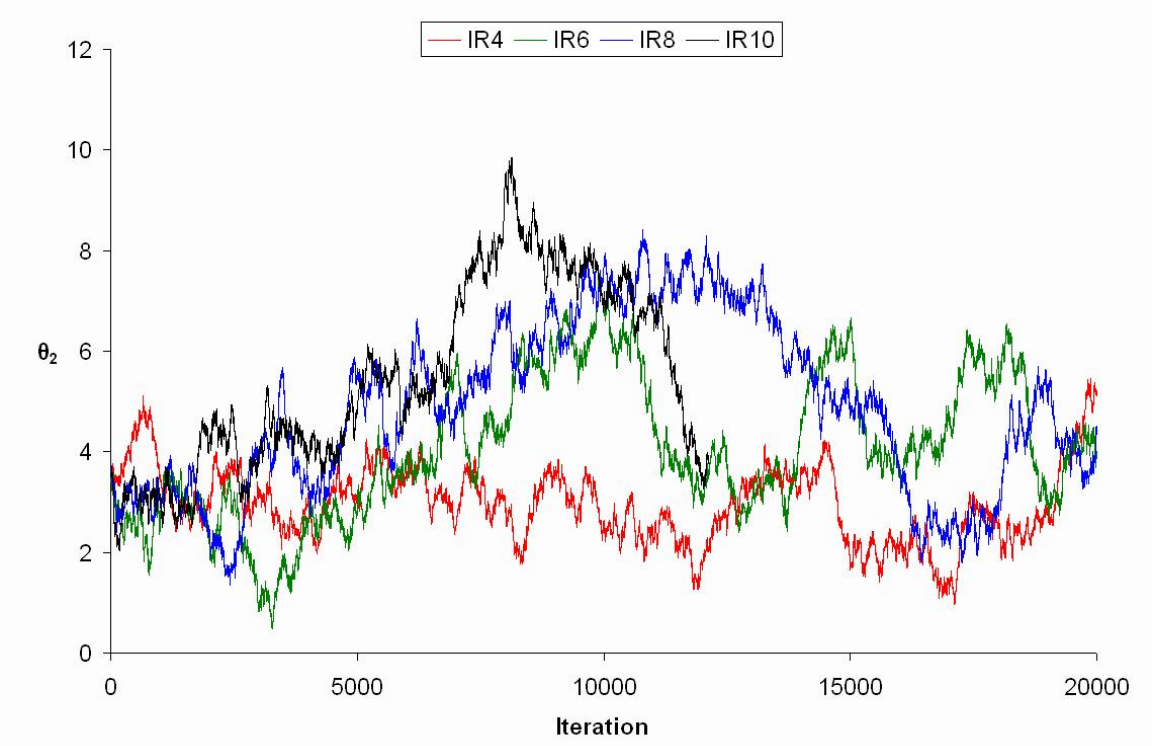

Figure 103. Trace plot of the parameter θ_2 of the EXPTP model for simulated data on a 30 x 30 plot. Interaction distance ranges from 4m (red) to 10m (black).

Table 8. Posterior modes and acceptance rates from Bognar's algorithm with EXPTP model for simulated data on 30 x 30 plot. Interaction distance ranges from 4m to 10m.

Interaction	*SPTP parameter*			*Acceptance*
distance (m)	β	θ_1	θ_2	*Rate (%)*
4	0.0464	-2.72	2.80	62.9
6	0.1098	-2.43	3.77	84.5
8	0.1051	-1.79	3.16	88.8
10	0.1212	-2.77	4.33	89.9

After completion of Bognar's algorithm for each interaction distance trial, the coefficient of variation (CV) of the importance samples were obtained and compiled into histograms which indicate that larger CVs were obtained with the smaller interaction distance trials (Figs. D21 – D24). For the 4m interaction distance trial, 3,894 iterations had CVs greater than 100% (19.47%), whereas the 6, 8, and 10m interaction distance trials had 215, 49, and 30 iterations with CVs greater than 100%, corresponding to 1.075, 0.245, and 0.2475% of the iterations, respectively, for these trials. In contrast to the increase in CVs with decreasing interaction distance, the acceptance rates of Bognar's algorithm increased with increasing interaction distance, with a nearly 90% acceptance rate for the 10m interaction distance trial (Table 8). As higher CVs imply higher variance of the importance sampling weights and consequently poorer estimation of the normalizing constant ratio, there appears to be an inverse relationship between the quality of estimation of the normalizing constant ratio, and the quality of the mixing of Bognar's algorithm, where high acceptance rates imply very slow (undesirable) mixing.

Using all available (20,000 or 12,122) iterations, a multivariate normal kernel density estimator was used to obtain joint and marginal posterior modes for each interaction distance trial. Of all the trials, the 6m interaction distance trial had an estimated posterior mode that was numerically closest to the known target parameter values of $\beta = 0.0625$, $\theta_1 = -2.5$, and $\theta_2 = 3.5$ (Table 8). The estimated mode for the 8m trial contained the smallest underestimate of θ_1, whereas the estimated mode for the 10m trial had the largest overestimate of both θ_1 and θ_2. In both the 8m and 10m trials, the smoothed posterior distributions (not shown) were highly bi- or multimodal, reflecting the poorer mixing of the trace plots. In comparison, the joint posterior densities for the 4m and 6m interaction distances were much more strongly unimodal in character in all parameter dimensions.

As in the case of previous trials, numerical differences in the estimated posterior modes and the known parameter values did not prevent the Monte Carlo envelopes of the simulated EXPTP process from enclosing the *L*-function of the ST55_IR datasets (Figs. 103 – 106). Only a few isolated values of the *L*-functions of the ST55_IR datasets fell slightly outside of the Monte Carlo envelope, such as the 4m trial at a distance of approximately 8m and the 10m trial at distance of approximately 10m. Although *L*-functions are not recommended to be plotted at distance greater than 25% of the side of a square plot (Baddeley and Turner 2008), the result for the Monte Carlo envelope for the 10m trial (Fig. 107) was computed out to 12m, and was found to enclose the ST55_IR10 *L*-function up to 12m as well.

As the last study of the performance of Bognar's algorithm for Bayesian inference with triplets processes, inference was undertaken for the observed data on Coweeta Plot

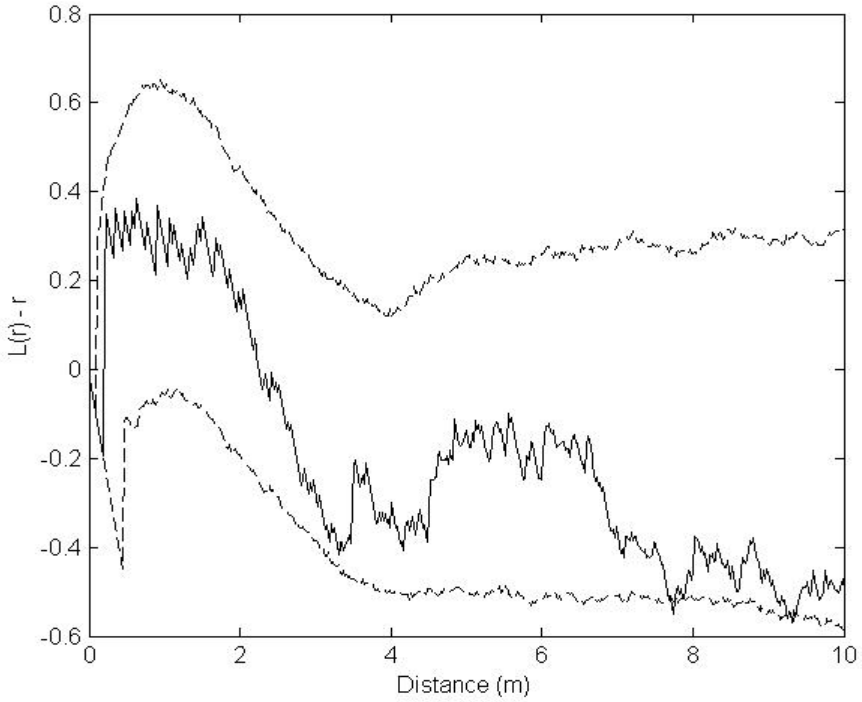

Figure 104. Monte Carlo envelope of *L*-function of EXPTP model using the posterior mode and a 4m interaction distance against the ST55_IR4 dataset.

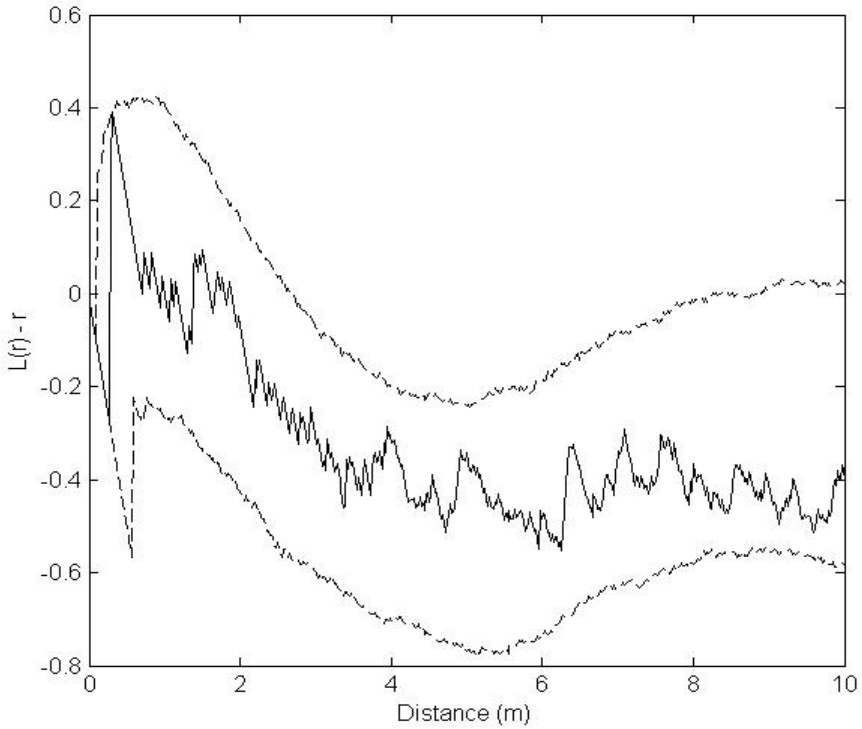

Figure 105. Monte Carlo envelope of *L*-function of EXPTP model using the posterior mode and a 6m interaction distance against the ST55_IR6 dataset.

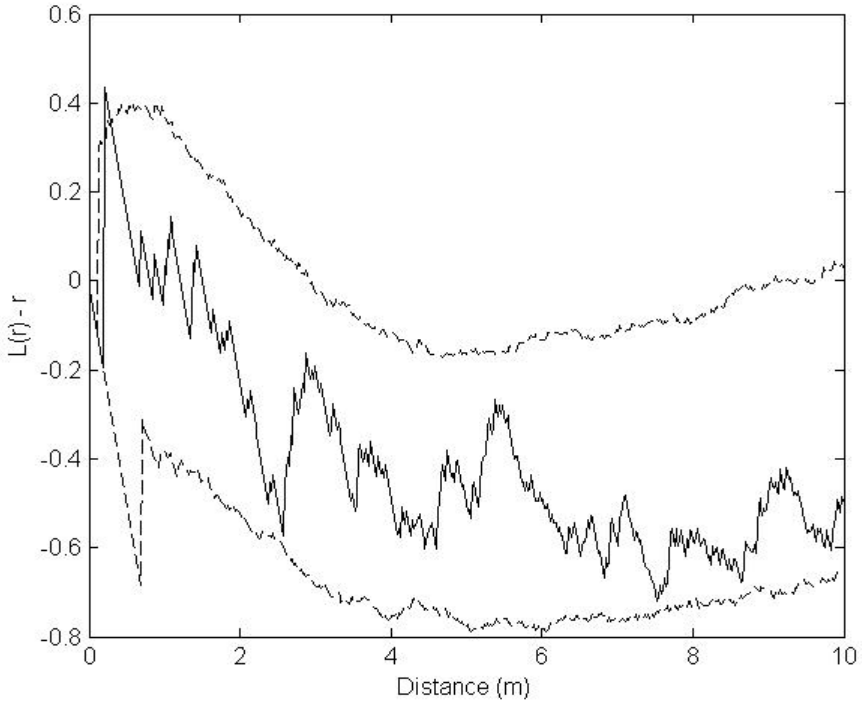

Figure 106. Monte Carlo envelope of *L*-function of EXPTP model using the posterior mode and a 8m interaction distance against the ST55_IR8 dataset.

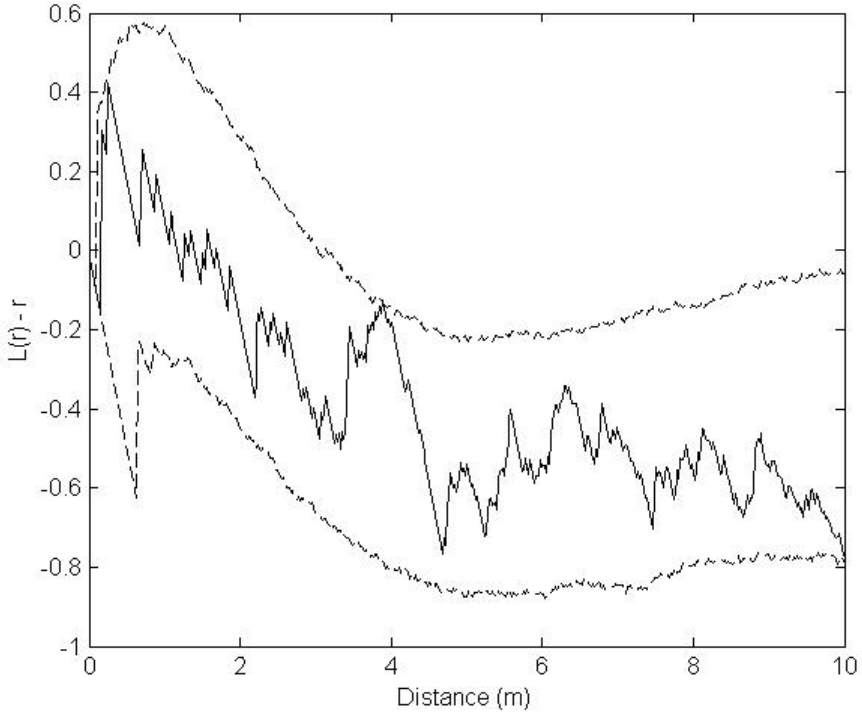

Figure 107. Monte Carlo envelope of *L*-function of EXPTP model using the posterior mode and a 10m interaction distance against the ST55_IR10 dataset.

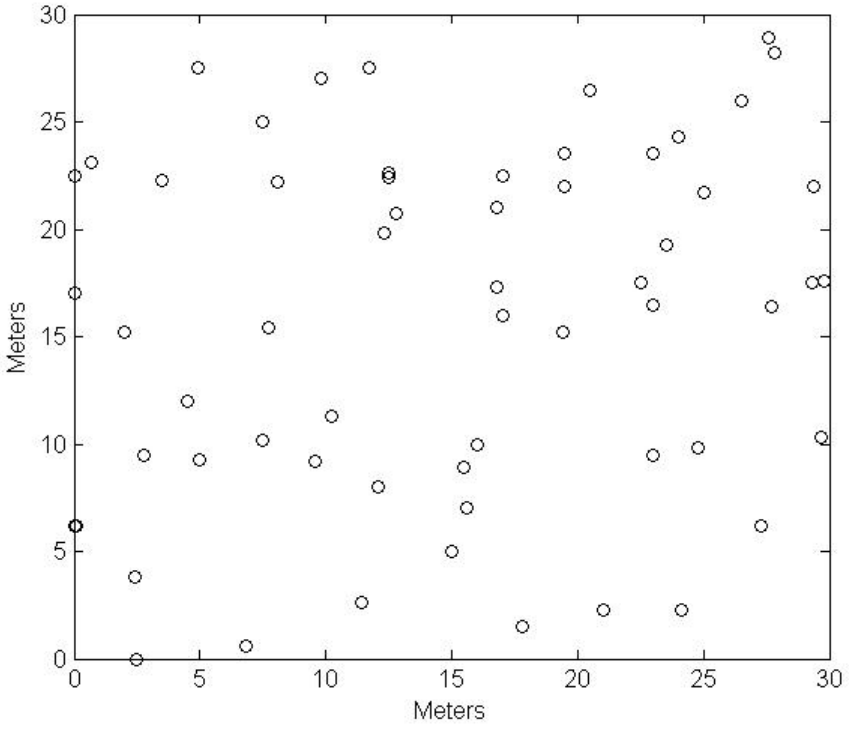

Figure 108. Stem map of the center 30 x 30m of Coweeta Plot 318.

318 using the EXPTP, GTP, and SPTP processes. As observed with the simulation trial of the EXPTP process using the ST55_IR10 dataset, iteration times of over 200 seconds can occur with Bognar's algorithm with triplets processes and large interaction distances, even when a small subset of the data (n = 50 points) was used. Since Coweeta Plot 318 has a much larger number of points (349), the iteration time of Bognar's algorithm was excessively long regardless of the choice of interaction distances when the whole 80 x 80m plot was used. Therefore, the center 30 x 30m piece from the center of Plot 318, with n =60 points, was used instead (Fig. 108). The choice of an interaction distance within the range of 4 to 10m is not obvious. One way to estimate the interaction distance statistically is the so called "cusp point" method, which takes the interaction distances as

the point of discontinuity in the *K*-function (Illian et al. 2008). While a computation of the *K*- and *L*- functions indicated that for the center 30 x 30m of plot 318 the cusp point of the *L* function occurs at 4.26m, a slightly larger 5m interaction distance was used.

In previous trials with the EXPTP and SPTP processes, Bognar's algorithm was executed using simulated data and a starting state of the parameter vector corresponding to the true parameter values that generated the simulated data. For inference on Plot 318, the starting parameter values of each model were set as follows: 1) the intensity parameter (β) being set to 0.05, 2) the fixed hard core and interaction distances set to 0.09 and 5.0m, respectively, and 3) all other parameters set to zero. For the trial of each triplets model, Bognar's algorithm was run for 20,000 total iterations. For the EXPTP and GTP models, the sub-chain burn-in was set to 15,000 iterations, with 100 importance samples drawn from the next 5,000 iterations of Geyer and Møller's algorithm. Out of concern that the SPTP process may be slower to burn in, the sub-chain burn-in for this model was set to 50,000 iterations, with 100 importance samples drawn from the next 5,000 iterations at a spacing of 50 iterations. After a few preliminary trials, the tuning parameters for the triplets models were selected as follows: 1) For the EXPTP process, $\sigma_{\beta} = 0.01$, $\sigma_{\theta_1} = 0.05$, and $\sigma_{\theta_2} = 0.05$; 2) for the GTP process $\sigma_{\beta} = 0.01$, $\sigma_{\theta_1} = 0.10$, and $\sigma_{\theta_2} = 0.10$, and 3) for the SPTP process $\sigma_{\beta} = 0.01$, $\sigma_{\theta_1} = 0.05$, $\sigma_{\theta_2} = 0.05$,and $\sigma_{\theta_3} = 0.05$.

Owing to the much larger sub-chain burn-in, the trial using the SPTP process with the Plot 318 data had much slower average iteration times (117.59 seconds) than did the trials with the EXPTP (38.29 seconds) and GTP processes (44.98 seconds). As a consequence, the trial with the SPTP process required 27 days and 5.27 hours to complete

the 20,000 iteration run of Bognar's algorithm, in comparison to the total time of 8 days, 20.75 hours for the EXPTP and 10 days, 9.88 hours for the GTP runs. Trace plots from the EXPTP, GTP , and SPTP runs (Figs. E1 – E10) indicate slow mixing across all models, with perhaps the best mixing present in the trace plot for θ_2 of the GTP model (Fig. E6). The trace plots for the SPTP process indicates very poor mixing with the Markov chain appearing to be slowly exploring several different modes, where the signs of the parameters θ_1 and θ_2 switched several times within the first 10,000 iterations (Figs. E8, E9). Despite apparent better mixing in the trace plots of the GTP process and an acceptance rate of 32.5%, the GTP process had surprisingly large CVs of the importance sampling weights. The histograms of the CVs for both the GTP and SPTP process (Figs E12, E13) indicate poor estimation of the normalizing constant ratio, with 50.25% and 35.69%, respectively, of the 20,000 iterations of Bognar's algorithm having CVs greater than 100%. Only the run for the EXPTP model obtained low CVs of the importance sampling weights, with only 1.13% of the 20,000 iterations having CVs greater than 100%.

An examination of the summary statistics of the MCMC output using the R-language contributed package BOA (Smith 2005) indicated the presence of high autocorrelation in the time series of the parameter values for all model runs of Bognar's algorithm (Table 9). The SPTP process for example, had a serial correlation of 72.57% for the intensity parameter β, implying that the values for this parameter are still 72.57% correlated with each after 100 iterations of Bognar's algorithm have been computed. Not only is this a high degree of correlation by itself, but for the SPTP process the parameters

Table 9. Lag – 100 autocorrelation and acceptance rates from Bognar's algorithm with the EXPTP, GTP, and SPTP models for the center 30 x 30m of Coweeta Plot 318.

Gibbs point	*Parameter*				*Acceptance*
process model	β	θ_1	θ_2	θ_3	*Rate (%)*
EXPTP	0.2066	0.4466	0.6198	–	84.5
GTP	0.7053	0.4079	0.1193	–	32.5
SPTP	0.7257	0.8336	0.9710	0.7954	49.0

controlling the spatial interaction all had even higher autocorrelations. A summary of the correlation between different parameters across the 20,000 iteration runs indicated for both the EXPTP and GTP models the presence of strong negative correlation between the parameters θ_1 and θ_2, which control the strength of the attractive 2nd order and repulsive 3rd order spatial interaction (Table 10). The SPTP process showed a rather complex correlation structure with the 2nd order potential parameters θ_1 and θ_2 being negatively correlated with each other, but where θ_1 was negatively correlated and θ_2 was positively correlated with the 3rd order potential parameter θ_3. The presence of such high autocorrelations reflect the presence of slowly mixing Markov chains, whereas the presence of cross-correlation may reflect the need to reparameterize the model (Cowles and Carlin 1996).

Given the presence of these correlations and the slow mixing evident in the trace plots, the choice of a burn-in for assessing summary statistics of the posterior distributions is problematic. For such slow mixing chains such as the SPTP process run, it

Table 10. Cross-correlations of EXPTP, GTP, and SPTP model paramters from Bognar's algorithm with the center 30 x 30m of Coweeta Plot 318.

	EXPTP Process Parameters			
	β	θ_1	θ_2	
β	1			
θ_1	0.3117	1		
θ_2	0.2896	-0.6403	1	
	GTP Process Parameters			
	β	θ_1	θ_2	
β	1			
θ_1	0.7489	1		
θ_2	-0.3095	-0.7730	1	
	SPTP Process Parameters			
	β	θ_1	θ_2	θ_3
β	1			
θ_1	0.7146	1		
θ_2	-0.3026	-0.4250	1	
θ_3	-0.0191	-0.3983	0.2394	1

is possible that even after all 20,000 iterations have completed the chain has not converged to the stationary distribution. It was determined from visual inspection that a burn-in of 2,0000 iterations was a suitable choice for testing Bognar's algorithm, given that no clearer evidence of a suitable burn-in interval could be discerned from the trace plots (Figs. E1 – E10). The remaining 18,000 iterations for each run were used to estimate the joint posterior density via the multivariate normal kernel density estimator, along with the smoothed marginal densities for each model's parameters. For the EXPTP

and GTP processes, marginal densities for parameters were mound-shaped with peaks that were similar to the joint densities for the respective models' parameters (Figs E14 – E19). For the SPTP process on the other hand, there are large differences in the shapes of the marginal densities and the joint densities (Figs E20 – E23) showing a more complex posterior surface. For the parameter θ_2 for example, the marginal density has an irregular bimodal shape with the largest mode at approximately -1. The joint posterior when viewed in the θ_2 dimension, is left-skewed with a mode at 8.89. (Fig. E22, Table 11)

Since an important advantage of Bayesian inference for point processes is the ability to generate interval estimates without recourse to asymptotic theory on the behavior of point process parameter estimators, 95% highest posterior density intervals for the marginal distribution for each parameter of the EXPTP, GTP, and SPTP models were generated through the program BOA, which uses the method of Chen and Shao (1999). These are presented along with the modes of the joint posterior distribution for each triplets process (Table 11). It is important to note that these interval estimates have a much more direct interpretation than comparable (frequentist) confidence intervals. Specifically these intervals state that there is a 95% probability that the parameter in question lies in the interval itself, unlike confidence intervals which give probability statements about the upper and lower bounds of an interval. Furthermore, the highest probability density intervals have the smallest interval width of all possible 95% Bayesian credible intervals that could be formed from the MCMC output (Chen and Shao 1999), and consequently do not need to correspond to the 2.5% and 97.5% percentiles of the parameter's respective marginal sample density. Another important point to bear in mind is that the highest probability density intervals are constructed from the unsmoothed

Table 11. Joint posterior modes and highest 95% probability density intervals (in parentheses) for the (unsmoothed) marginal distribution of each parameter of the EXPTP, GTP, and SPTP models for the center 30 x 30m of Coweeta Plot 318 from Bognar's algorithm.

Gibbs point	*Parameter*			
process model	β	θ_1	θ_2	θ_3
EXPTP	0.0918	-2.55	4.81	–
	(0.0551, 0.1791)	(-3.92, -0.33)	(2.06, 7.69)	
GTP	0.1232	-0.04	0.17	–
	(0.0308, 0.2824)	(-0.41, 0.26)	(0.00, 0.41)	
SPTP	0.0999	-0.09	8.89	2.85
	(0.0494, 0.2746)	(-1.09, 2.43)	(-2.53, 11.82)	(0.59, 6.61)

(marginal) sample path of each parameter, whereas the joint posterior modes are found as the maximum value of the smoothed joint posterior density. The joint posterior mode point estimates of each parameter do not lie at the center of the highest probability density intervals and are constructed independently from them.

One immediate feature of the interval estimates is that for the parameter θ_1 of both the GTP and SPTP models, and the parameter θ_2 for the SPTP model, the intervals include both positive and negative values. For the GTP process, this is problematic since when θ_1 is positive the process is entirely repulsive, which is not consistent with the Plot 318 data. For the SPTP process, the uncertainty in the signs of θ_1 and θ_2 does not necessarily imply that the model would be purely repulsive, since negative values of θ_1 combined with positive values of θ_2 can give partially attractive 2^{nd}-order interactions just

as well as positive values of θ_1 combined with negative values of θ_2. Given the negative cross correlation between these parameters (-0.4250), it is likely that the Markov chain for the SPTP posterior is exploring modes where different combinations of θ_1 and θ_2 are giving similar shapes of the SPTP process pair potential.

After estimation of the joint posterior modes, the EXPTP, GTP, and SPTP processes were simulated via Geyer and Møller's algorithm using the estimated joint posterior mode. For each model, the 95% Monte Carlo envelopes of the *L*-function were computed and compared against the *L*-function from the center 30 x 30m piece of Coweeta Plot 318 (Figs 108 – 110). While none of the Monte Carlo envelopes enclosed the observed *L*-function for all distances, all processes enclosed the observed *L*-function for all distances greater than 0.5m. Of the three models, the EXPTP process gives the best fit, with the observed *L*-function being slightly more attractive at distances less than 0.5 meters. The envelope for the GTP process provides the poorest fit, where the observed *L*-function has a sharp decrease from clustering to repulsion within the interval of 0 – 2m, whereas the Monte Carlo envelope shows instead a gradual decrease from random to repulsive interaction in the same interval. Aside from this deficiency in modeling attraction at very short distances, the GTP process envelope is otherwise quite consistent with the observed *L*-function and the envelopes of the EXPTP and SPTP processes, despite having a very different functional form of the pair potential than either of these models.

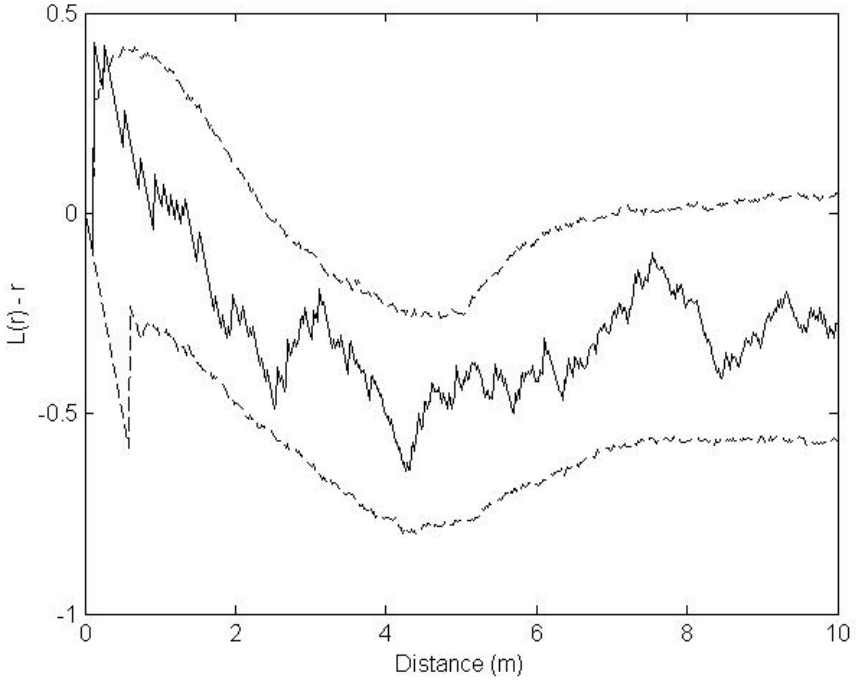

Figure 109. Monte Carlo envelope of *L*-function of the EXPTP model using the posterior mode from Bognar's algorithm against the center 30 x 30m of Plot 318.

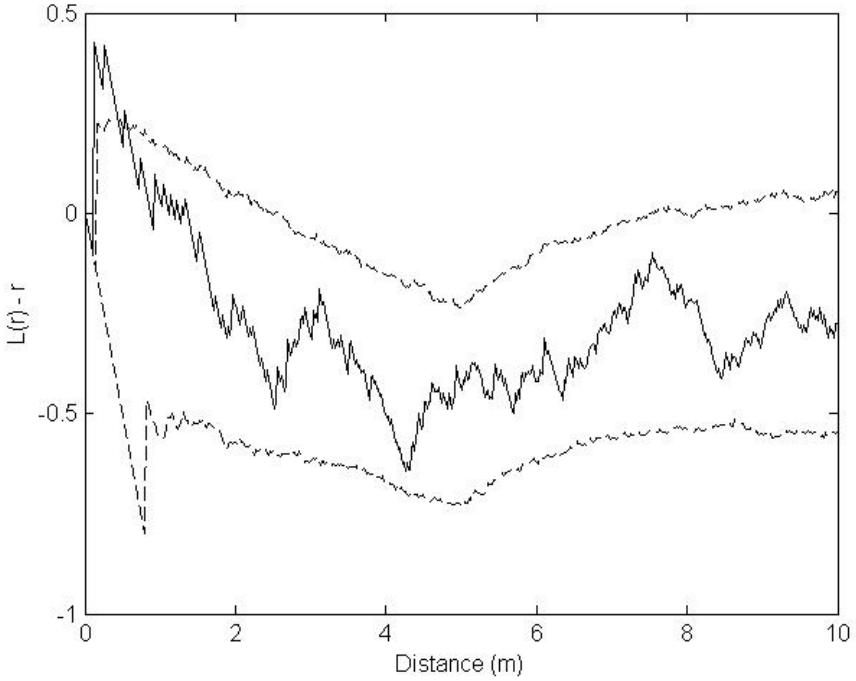

Figure 110. Monte Carlo envelope of *L*-function of the GTP model using the posterior mode from Bognar's algorithm against the center 30 x 30m of Plot 318.

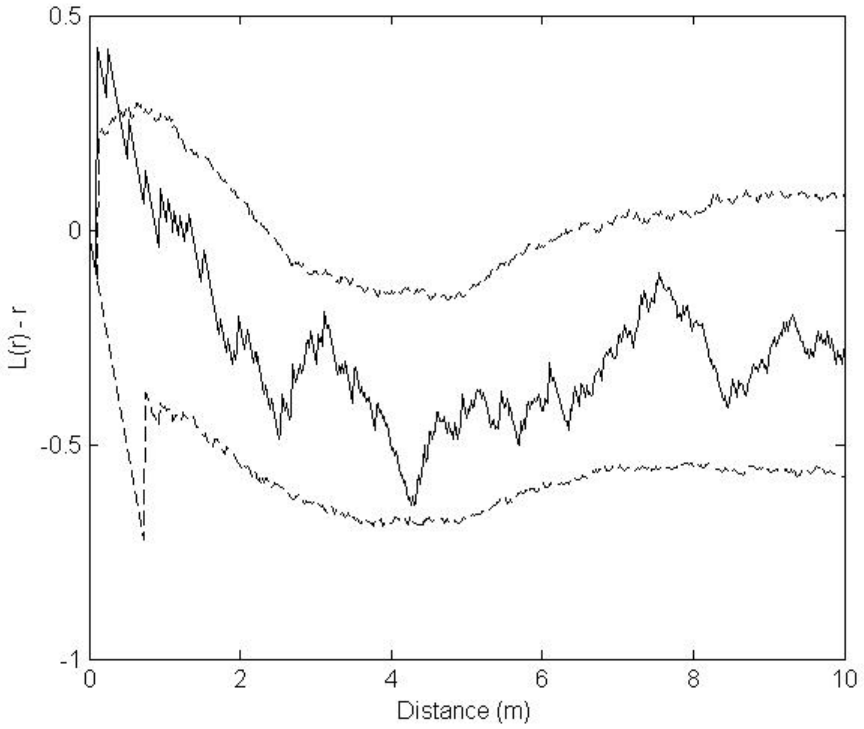

Figure 111. Monte Carlo envelope of *L*-function of SPTP model using the posterior mode from Bognar's algorithm against the center 30 x 30m of Plot 318.

4.4 Gibbs Point Process Inference Using Auxiliary Variables

As an alternative to approximating ratios of normalizing constants in Metropolis-Hastings algorithms that sample from the posterior distribution of a Gibbs or Markov point process, Møller et al. (2006) and Berthelsen and Møller (2006) present a novel approach to cancel out the normalizing constants by introducing an auxiliary variable that is proposed with same distribution as the Gibbs or Markov point process likelihood. As pointed out by Møller et al. (p. 457), perfect samples of the auxiliary are needed to "avoid introducing additional undesirable stochasticity."

From the discussion of stability in Sec. 3.73, the use of dominated CFTP requires the dominating process to be a Poisson process with intensity measure $\Lambda(B) = K\ \mu_L(B)$, for Borel sets *B*, where *K* is the upper bound of the locally stable point process' conditional intensity function. To initialize the algorithm, a sample is needed from this Poisson process which depends upon *K*. The upper bound *K* depends upon the model and its parameters, so consider the EXPTP process with the parameters $\beta = 0.0918$, $\theta_1 = -2.55$, $\theta_2 = 4.81$, $r_0 = 0.09$, $R = 5$ (Table 11), corresponding to the solution from Bognar's algorithm on the center 30 x 30m of Coweeta Plot 318. In Sec. 3.5, the upper bound of the log-conditional intensity function for the EXPTP model was found as:

$$\ln\left[\lambda^*(u;\mathbf{x})\right] \leq= -2\theta_2 \exp(-R) n_R(\mathbf{x})^2 + \left[2\theta_2 \exp(-R) - 4\theta_1 \exp(-r_0)\right] n_R(\mathbf{x}) + \ln(\beta)$$

Substituting in the parameter values gives a quadratic function with a maximum of

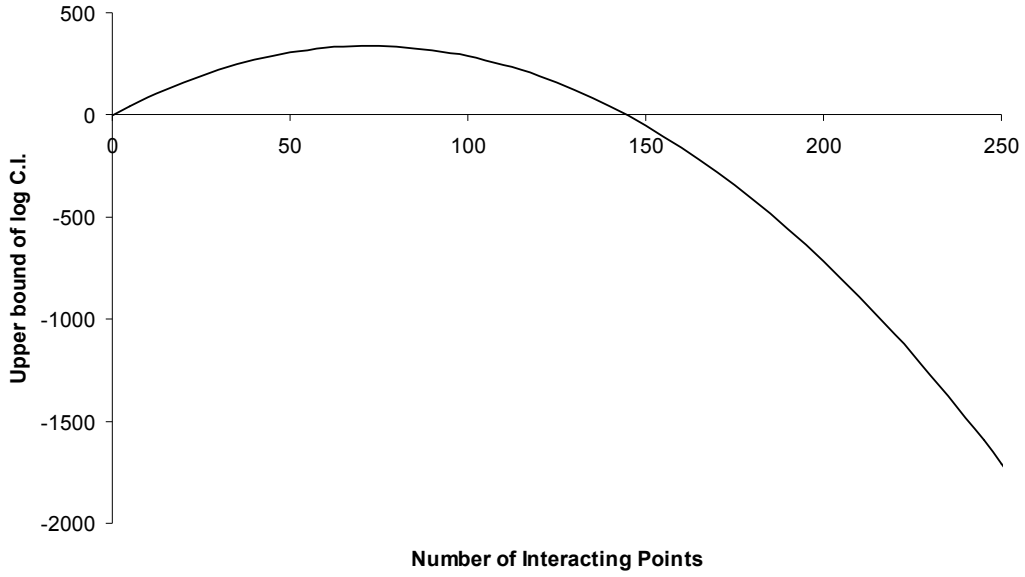

Figure 112. Plot of the upper bound of the log of the conditional intensity function, $\ln\left[\lambda^*(u;\mathbf{x})\right]$, against the number of interacting points using the EXPTP process and the parameter values $(\beta = 0.098,\ \theta_1 = -2.55,\ \theta_2 = 4.81,\ r_0 = 0.09,\ R = 5)$.

approximately 337.4479 corresponding to 72 interacting points (Fig. 112). Since this is the maximum of the log of the conditional intensity the maximum of the conditional intensity itself is $K = \exp(337.4479) \approx 3.6 \text{ x } 10^{146}$. Consequently, for a 30 x 30m plot the dominating process of the dominated CFTP algorithm must be initialized as a sample from a Poisson process with a mean of $K\ \mu_L(F) = (3.6 \text{ x } 10^{146})(30^2) \approx 3.2 \text{ x } 10^{149}$ points. Since the dominated CFTP algorithm proceeds by adding to and deleting points from this initial configuration, one at a time, until the empty configuration is reached, it is clear that dominated CFTP would take too much time to be practical in the context of Møller et al.'s (2006) auxiliary variable algorithm when applied to the EXPTP process. Further experimentation with other triplets models and parameter values indicated that all of the

triplets processes had very high upper bounds on their conditional intensity functions, and that these upper bounds increased dramatically with increases in the interaction distance of the model. Moreover, both the dominated CFTP algorithm using upper and lower processes and the clan of ancestors algorithm (Berthelsen and Møller 2002; Van Lieshout and Stoica 2006) depend upon initializing a dominating Poisson process constructed with intensity measure $\Lambda(B) = K\ \mu_L(B)$. Consequently neither algorithm could be used, and perfect simulation of triplets processes is viewed as impractical.

Instead of using perfect simulation, Møller et al.'s algorithm was implemented where imperfect samples of the auxiliary variable are drawn via Geyer and Møller's shift-birth-death Metropolis Hastings algorithm, where one sample configuration is drawn after burn-in. To provide a comparison with Bognar's algorithm, Møller et al.'s algorithm was tested for the EXPTP, GTP, and SPTP processes on the center 30 x 30m of Coweeta Plot 318. To draw the imperfect samples of the auxiliary, the burn-in for Geyer and Møller's algorithm is set to 15,000 iterations for the EXPTP and GTP trials, and 50,000 iterations with the SPTP trial. For each trial, the auxiliary variable parameter vector $\tilde{\theta}$ was set to the posterior mode solutions from Bognar's algorithm from Plot 318 (Table 11). The auxiliary variable was initialized with a sample from a binomial point process where n = 60 points are selected with point locations independently and uniformly distributed in a 30 x 30 square. The initial state of the parameter vector of the posterior for each triplets process was kept the same from the trials with Bognar's algorithm, where the intensity parameter β is set to 0.05, and all other parameters of the 2nd- and 3rd-order potentials are zero. As before, the hard-core distance is kept at 0.09m and the interaction distance at 5m. The tuning parameters for the independent normal parameter

proposal distributions for each triplets model are the same as used with Bognar's algorithm: 1) For the EXPTP process, $\sigma_\beta = 0.01$, $\sigma_{\theta_1} = 0.05$, and $\sigma_{\theta_2} = 0.05$; 2) for the GTP process $\sigma_\beta = 0.01$, $\sigma_{\theta_1} = 0.10$, and $\sigma_{\theta_2} = 0.10$, and 3) for the SPTP process $\sigma_\beta = 0.01$, $\sigma_{\theta_1} = 0.05$, $\sigma_{\theta_2} = 0.05$,and $\sigma_{\theta_3} = 0.05$.

During the implementation of Møller et al.'s algorithm with the GTP process, occasionally parameter vectors would be proposed with near zero repulsion in the 3rd-order potential, which can occur when θ_2 is very small. When $\theta_1 < 0$ is proposed, this resulted in Geyer and Møller's algorithm attempting to draw configurations using parameters that were so attractive that the number of points in the sampled configurations exceeded 1,000. In less than 200 iterations, the C++ code running Møller et al.'s algorithm crashed because the interacting pairs and triples from configurations with over 1,000 points became too large to hold in memory. To avoid this problem, the prior of the GTP process was modified so that $\theta_2 > 0.05$, and any proposed parameter vectors with too weak repulsion were automatically rejected.

After completion of each trial, the average iteration times for the EXPTP (31.31 seconds), GTP (32.23 seconds), and SPTP processes (107.29 seconds) were 6.98, 12.75, and 10.30 seconds faster, respectively, than those from Bognar's algorithm. The long iteration time for the SPTP process coincides with the 50,000 iteration sub-chain burn-in used with this model, and this resulted in the total time to complete the 20,000 iteration run to be 24 days and 20 hours.

Trace plots from the trials (Figs. F1 – F13) indicated the presence of long runs of rejection; for example the EXPTP process had a run of 3,372 iterations where the Markov

Table 12. Cross-correlations of EXPTP, GTP, and SPTP model parameters from Møller et al.'s algorithm with the center 30 x 30m of Coweeta Plot 318.

	EXPTP Process Parameters			
	β	θ_1	θ_2	
β	1			
θ_1	0.2358	1		
θ_2	0.4224	-0.6673	1	
	GTP Process Parameters			
	β	θ_1	θ_2	
β	1			
θ_1	0.7519	1		
θ_2	-0.2689	-0.7284	1	
	SPTP Process Parameters			
	β	θ_1	θ_2	θ_3
β	1			
θ_1	0.4770	1		
θ_2	-0.0183	0.2568	1	
θ_3	0.2416	-0.3140	-0.5992	1

chain was "stuck" at the parameter values $\beta = 0.2270$, $\theta_1 = -1.25$, and $\theta_2 = 4.90$. All models show slowly mixing trace plots, with the SPTP process showing the poorest mixing, where the trace plots for θ_1 and θ_2 are sampling erratically positive values of θ_1 and negative values of θ_2. The presence of cross correlation between parameters and autocorrelation for a given process' parameter is evident for all three triplets processes. Strong negative cross correlation occurred between the parameters θ_1 and θ_2 of the 2nd and 3rd order potentials for the EXPTP and GTP processes (Table 12). The SPTP process

Table 13. Lag – 100 autocorrelation and acceptance rates from Møller et al.'s algorithm with the EXPTP, GTP, and SPTP models for the center 30 x 30m of Coweeta Plot 318.

Gibbs point	*Parameter*				*Acceptance*
process model	β	θ_1	θ_2	θ_3	*Rate (%)*
EXPTP	0.8759	0.8562	0.9022	−	21.6
GTP	0.8079	0.5539	0.1611	−	13.8
SPTP	0.7812	0.8308	0.9261	0.8486	19.3

showed a complex correlation structure where the parameters θ_1 and θ_2 were positively correlated with each other and both negatively correlated with θ_3. In contrast, the parameters θ_1 and θ_2 were negatively correlated with each other and were negatively and positively correlated with θ_3, respectively, when using Bognar's algorithm (Table 10) for the SPTP process and the same data.

As a consequence of the long runs of rejection in Møller et al.'s algorithm, the acceptance rates for all models were quite low, with the EXPTP process having the largest rate at 21.6%. Another consequence of the long runs of rejection was high autocorrelation, particularly for the EXPTP and SPTP processes. For example, autocorrelation for the EXPTP process the smallest lag-100 autocorrelation for the EXPTP process was 85.62% for the parameter θ_1, and the smallest lag-100 autocorrelation for the SPTP process was 78.12% for the parameter β (Table 13).

The presence of autocorrelations reinforce the visual notion that Møller et al.'s Metropolis-Hastings sampler is slowly mixing around the support for each triplets process' posterior distribution. An undesirable consequence of this slow mixing is that it

becomes quite difficult to determine what an appropriate burn-in is for the determination of point and interval estimates of each posterior distribution. As in the analysis of the Plot 318 data with Bognar's algorithm, visual inspection of the trace plots was used to choose 3,500, 2,000 and 2,500 iterations for the EXPTP, GTP, and SPTP processes, respectively, in the absence of any other evidence from the trace plots. The post-burn-in iterations from each trial were then smoothed to estimate the joint and marginal densities for each triplets process' parameters (Figs F14 – F26). For the GTP process, the joint density and the marginal densities appeared as mound shapes with similar values for the marginal and joint posterior mode for each parameter. For the EXPTP process, the 3,372 run of rejection at $\beta = 0.2270$, $\theta_1 = -1.25$, and $\theta_2 = 4.90$ caused the smoothed joint posterior density to be discontinuous and unusually large at this point (Figs F14 – F16). All marginal densities have modes that correspond to this point, resulting in a bimodal marginal density for the parameter β, where the largest mode occurs at 0.2270 and a second occurs at 0.1183. For the SPTP process, the presence of multiple long runs of rejection resulted in the smoothed SPTP posterior density being rather irregular in appearance. Moreover, the joint posterior mode for the SPTP process corresponds to $\beta = 0.1635$, $\theta_1 = 0.26$, $\theta_2 = -2.11$, and $\theta_3 = 4.38$, a point where Møller et al.'s algorithm was stuck in a run of non-acceptance for 466 iterations.

Because Møller et al.'s algorithm was run for only 20,000 iterations, the 3,372 run where no proposal is accepted in the trial of the EXPTP process represents 15.8% of the entire run of the algorithm. Consequently the joint posterior mode for the EXPTP process corresponds to the parameters values where the Markov chain is stuck, even though this

Table 14. Joint posterior modes and highest 95% probability density intervals (in parentheses) for the (unsmoothed) marginal distribution of each parameter of the EXPTP, GTP, and SPTP models for the center 30 x 30m of Coweeta Plot 318 from Møller et al.'s algorithm.

Gibbs point process model	*Parameter*			
	β	θ_1	θ_2	θ_3
EXPTP	0.1183	-2.15	5.35	−
	(0.0615, 0.2278)	(-3.50, -0.97)	(2.68, 7.29)	
GTP	0.1692	0.03	0.16	−
	(0.0401, 0.2501)	(-0.35, 0.23)	(0.05, 0.37)	
SPTP	0.1635	0.26	-2.11	4.38
	(0.0849, 0.2574)	(-0.09, 2.59)	(-4.47, 0.00)	(0.81, 4.80)

point is extreme with respect to the parameter β. Instead of using this point as the posterior mode solution for the EXPTP process, the point that corresponds to the mode of the majority of the mass of the joint density was used instead, with β = 0.1183, θ_1 = -2.15, and θ_2 = 5.35 (Table 14) and a joint posterior density of 2.41. For the GTP and SPTP processes, which did not have such long runs of non-acceptance, the posterior modes as obtained from the output of the kernel density estimator were used as point estimates (Table 14).

The results of the highest 95% probability density intervals show large variability in the parameter estimates for all models. As an example, there's a 95% chance that the parameter θ_1 of the GTP model lies in the interval (-0.35, 0.23), but this interval covers both positive and negative values. Similarly both the interval estimates for parameter θ_1

and θ_2 of the SPTP model include positive and negative values, implying considerable uncertainty in the form of the SPTP pair potential, which depends critically upon the sign of these parameters.

Using the respective mode of the smoothed posterior distribution, Geyer and Møller's algorithm was used to simulate each triplets model. Following a 50,000 iteration burn-in, 1,000 samples were drawn at a spacing of 200 iterations for each 250,000 iteration run. Ripley's *K*- and *L*-functions were calculated for the 1,000 samples from each model to form 95% Monte Carlo envelopes and were compared against the observed *L*-function from the center 30x30m of Coweeta Plot 318 (Figs. 112 – 114). Of the three models, the EXPTP process provided the best fit, where the observed *L*-function was enclosed by the EXPTP envelope except at distances less than 0.5 m, where the short distance clustering is underestimated. This underestimation is more pronounced for both the GTP and SPTP processes. In particular, the SPTP process appears to underestimate point attraction throughout distances less than 1m, which is the consequence of the joint posterior mode giving weak 2^{nd}-order attraction ($\theta_1 = 0.26$, and $\theta_2 = -2.11$) and strong 3^{rd}-order repulsion ($\theta_3 = 4.38$). Since the posterior mode for the SPTP process corresponds to a 466 iteration run of non-acceptance, it seems that the posterior mode is not providing a good basis for point inference in favor of other measures of the center of the SPTP posterior distribution, such as the posterior mean or median.

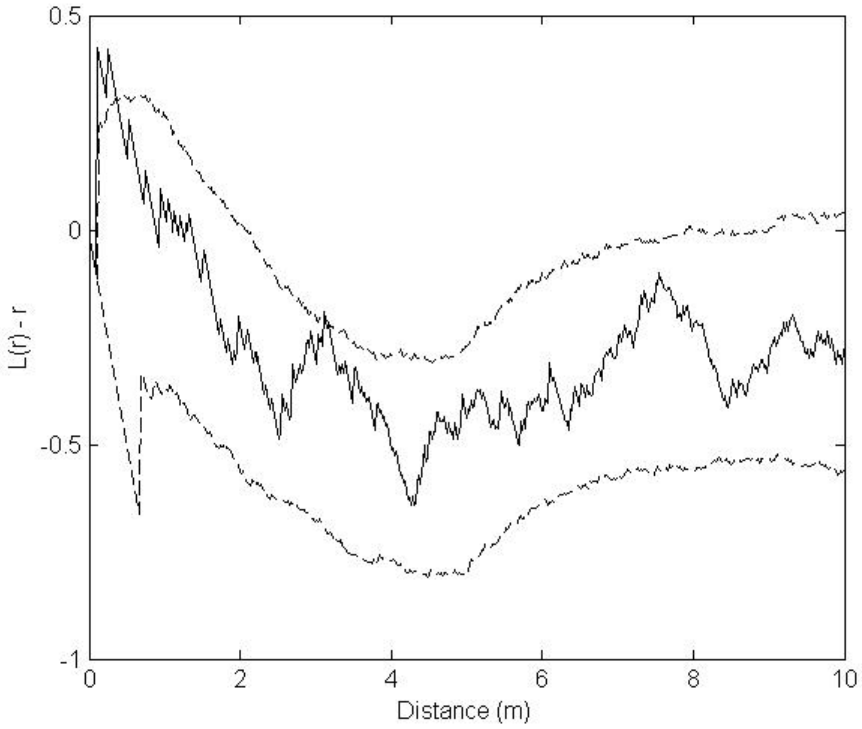

Figure 113. Monte Carlo envelope of *L*-function of the EXPTP model using the posterior mode from Møller et al.'s algorithm against the center 30 x 30m of Plot 318.

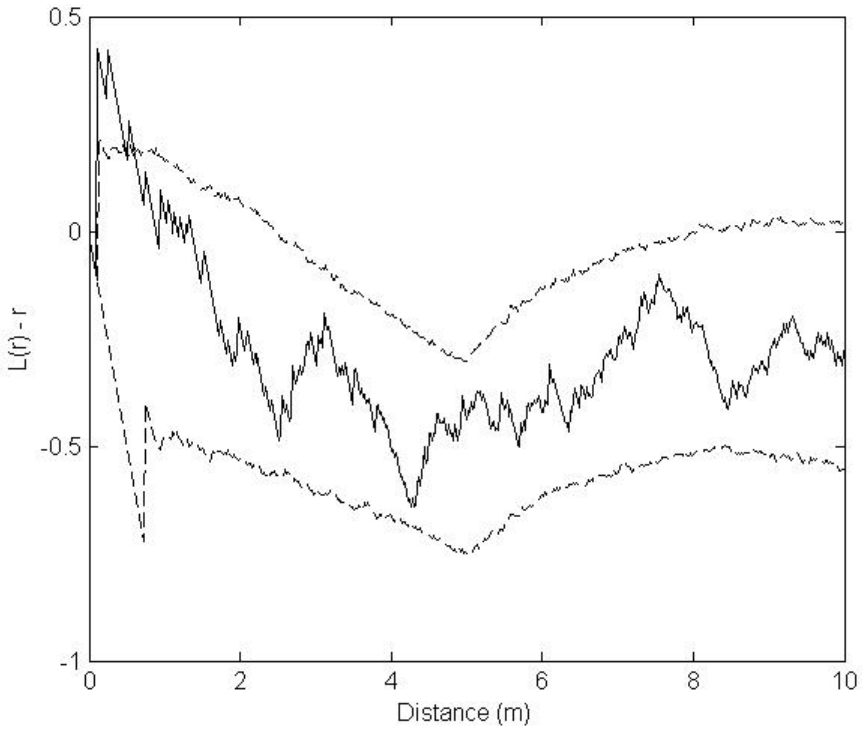

Figure 114. Monte Carlo envelope of *L*-function of the GTP model using the posterior mode from Møller et al.'s algorithm against the center 30 x 30m of Plot 318.

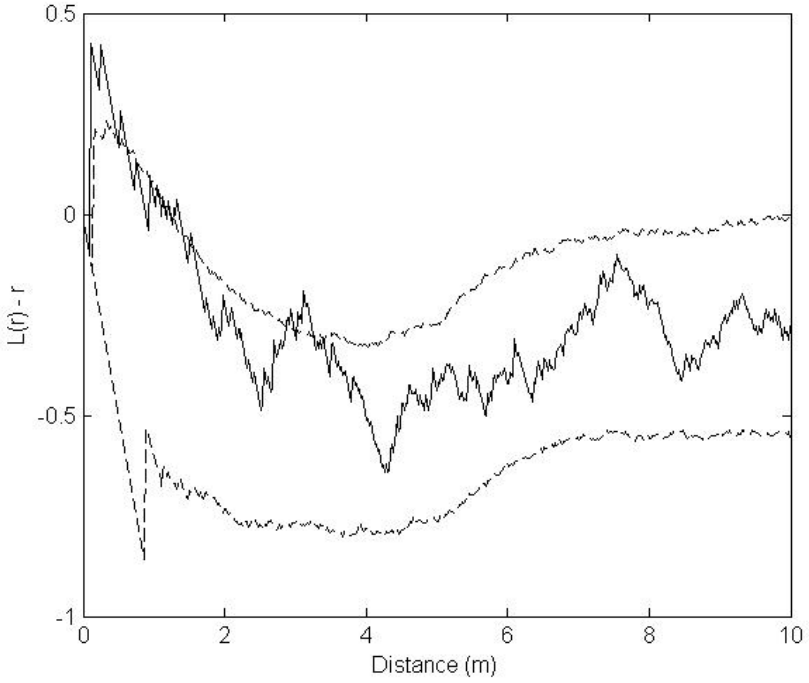

Figure 115. Monte Carlo envelope of *L*-function of SPTP model using the posterior mode from Møller et al.'s algorithm against the center 30 x 30m of Plot 318.

5. Discussion

The formulation and evaluation of Gibbs point processes as a means of simulating the patterns of tree stem locations in mixed-species Appalachian forests was a primary objective of this research. Prior research with Gibbs point processes on forest tree pattern (Penttinen et al. 1992; Mateu et al. 1998; Stoyan and Stoyan 1998; Degenhardt 1999; Kokkila et al. 2002; Neeff et al. 2005) has focused almost exclusively on purely repulsive patterns, owing in no small part to the fact that Gibbs processes were designed for modeling repulsive particle interactions in statistical thermodynamics. In the context of mixed Appalachian hardwood forests, stump sprouting is a common means of reproduction (Wendel and Trimble 1968; Wendel 1975; Beck 1977; Beck and Hooper 1986) that results in tree stem locations being very close to each other, owing to their origin from a common root system. It follows that Appalachian hardwood tree patterns may have many small clusters of trees interspersed in otherwise random or repulsive patterns. While there exist many Gibbs and Markov point process models for clustering such as the area interaction process (Baddeley and Van Lieshout 1995) and nearest-neighbor Markov processes (Baddeley and Møller 1989; Baddeley et al. 1996; Van Lieshout 2000; Grabarnik and Särkkä 2001), the triplets models developed here allow for modeling localized clustering and longer range repulsion simultaneously.

In recent reviews of point processes by Møller and Waagepeterson (2007) and Comas and Mateu (2007), existing Gibbs point processes are not considered adequate models for attractive interaction between points. Perhaps the most commonly cited work to support this view is that of Gates and Westcott (1986), who demonstrated that the pair

potential models of Strauss (1975), Ogata and Tanemura (1981, 1984), and Ripley (1977) violate Ruelle stability when the pair potentials of these models are allowed to give attractive interactions. In addition to providing evidence that the Mayer cluster expansion approximation of Ogata and Tanemura (1981) is poor when an attractive pair potential is used, Gates and Westcott argued that unstable potentials produce simulations from these models much more tightly clustered than reported by Ogata and Tanemura (1981) and Ripley (1977). This phenomenon is also reported by Geyer and Thompson (1995) and Geyer (1999), who looked at the Strauss process with attractive pair potential conditional upon a fixed number of points. Geyer and Thompson found that when the parameter of the Strauss process models increasing attractiveness, the number of neighboring pairs of points sharply increases towards a maximum of $\binom{n}{2}$, which occurs when all points interact with each other in one clump. Geyer (1999) confirmed that when the conditional Strauss process has very weak attraction, the process is Poisson-like. When the Strauss process models instead strong attractiveness, it becomes the "one-clump" process, where all points lie within a circle with a radius corresponding to the interaction distance of the model.

The results here support the work of Gates and Westcott (1986) that Gibbs point processes with only purely attractive 2nd-order interactions produce simulations that are highly clustered. The three models with only 2nd-order interactions (the Fiksel, HCME, and HCSP) models all simulated configurations that appeared to be realizations from the "one-clump" process, despite the fact that all three models were designed with a non-zero hard-core parameter and finite range of interaction to ensure that these models are both Ruelle and locally stable. An interesting result of this research is that neither local nor

Ruelle stability was sufficient to prevent sample configurations from Gibbs processes with attractive 2nd-order interactions from degenerating into realizations from the one-clump process. Furthermore models such as the HCME and HCSP processes with only partially attractive pair potentials showed a tendency to rapidly degenerate into "a multiple clump" process, particularly when attraction was modeled at short distances. The inherent instability of these models demonstrated that neither the presence of a non-zero hard-core distance nor conditioning on the number of points is necessarily sufficient to prevent Metropolis-Hastings simulations of attractive or partially pairwise Gibbs models from giving excessively clustered sample configurations. While the stepwise potential of Penttinen's (1984) multiscale pairwise interaction process was not specifically tested here, the results here suggest that allowing such a pair potential to be partially attractive as in Särkka and Tomppo (1998) and Tomppo (1986) can also result in instable models that will be excessively clustered when simulated.

Despite the poor performance of the Fiksel, HCME and HCSP models, the triplets models studied provided a viable means of modeling the point attraction and repulsion found on Coweeta Plot 318. While the EXPTP, GTP and SPTP models are also locally stable, their better performance can be seen in how the triplets models satisfy the local stability bound in comparison to the Fiksel, HCME ,and HCSP models with only pairwise interactions. In the pairwise models, the conditional intensity is bounded due to the combination of the hard-core and interaction distances. Within a finite range of interaction, the hard-core parameter assures that all points interacting with any given point u must all be separated by the non-zero hard-core distance from each other and u, creating at its limit a honeycomb-like grid of points (Fig. 12). When simulated with

Geyer and Møller's (1994) algorithm, points will be drawn into this highly clustered arrangement provided that no pair of points violates the hard-core condition. With a sufficiently small hard-core distance, such as the 0.09m distance used here, it is not difficult to draw a very large number of points into such a cluster. For example, with a 0.09m hard-core distance and a 5m interaction distance, the number of points within a circle of 5m radius of any point *u* must satisfy

$$n_R(\mathbf{x}) \le 6\left(\frac{R}{r_0}\right) = 6\left(\frac{5}{0.09}\right) \approx 333$$

which implies, that for a configuration with fixed $n = 349$ in a 80 x 80m plot, all but 16 points can be located in one single clump of 5m in radius. A triplets process, on the other hand, arrives at the upper bound of its conditional intensity function through the interplay of the attractive 2nd-order potential and the repulsive 3rd-order potential. For $n > 5$ points within a given point's interaction distance, the number of interacting triples $\binom{n}{3}$ will exceed the number of interacting pairs $\binom{n}{2}$ by a factor of $(n - 2)/3$. Consequently as the number of points grows in a cluster around any given point *u*, the sum contribution of the attractive 2nd-order potential evaluated on the pairs of point in the cluster will eventually be overwhelmed by the sum contribution of the repulsive 3rd-order potential evaluated on the much larger number of triples of points in the cluster. In effect the magnitude of the 3rd-order potential relative to the 2nd-order potential at very small distances, along with this combinatorial feature of the 3rd-order interaction, provides a much stronger break on

the formation of large clusters in triplets processes than what occurs in locally stable pairwise models with attractive interactions.

The results of this research indicate that triplets processes can simulate configurations of points that retain the short range attractive features found in the Coweeta data, whether caused by stump sprouting or other reasons. While the triplets processes here show much promise, they do not completely avoid "the pitfall that ensnared the Strauss process." (Geyer 1999, p. 110). The triplets process studied here, including the slight modification of the triplets process proposed by Geyer, can produce simulations that are excessively clustered. When the repulsion in the 3rd-order potential is relatively weak and the attraction in the 2nd-order potential is relatively strong, the EXPTP and SPTP process gave sample configurations with multiple large clusters arranged in fairly regular patterns (Figs. 73, 74). The GTP process in these circumstances gave sampled configurations where points formed linear clusters spaced at distances mirroring the interaction distance of the model (Fig. 73). Provided there is sufficient balance between the attractive 2nd-order and repulsive 3rd-order potentials, the resulting process can model localized attraction of points that do not explode into unrealistic "superclusters". While the results here do not provide a theoretical formulation for quantifying this balance point, heuristically it appears that if the 3rd-order potential has absolute value at the average pairwise distance of 0 greater in magnitude than the absolute value of the attractive 2nd-order potential at 0 distance, then simulations of the resulting process will not be realizations of the single-clump or multiple clump processes.

One question of interest was the how the form of the potential function(s) influenced the samples from the resulting Gibbs distribution, given the broad flexibility

in choosing the potentials subject only to satisfying Ruelle stability and parameter identifiability requirements (Baddeley 2001). The simulation results indicate that if a potential is allowed to model attraction, the magnitude of this attraction and the means by which the upper bound of the conditional intensity function is achieved is much more important than the form of the potential used. For example, both the HCSP process and the SPTP processes were simulated with $\beta = 0.0545$, pair potential parameter values of $\theta_1 = -2.5$ and $\theta_2 = -1$, and a 5m interaction distance. When plotted (Fig. 34), the pair potential indicates repulsive interactions up to 1m and seemingly weak attractive interactions beyond this up to the interaction distance. Simulations of the HCSP process with these parameter values gave configurations with massive attraction and unrealistically large numbers of points (Fig. 39). If the number of points is conditioned upon, then sample configurations appear as realizations of the one-clump process (Fig. 40). In comparison, the SPTP process was simulated with the same pair potential parameter values, but with relative strong third order repulsion with $\theta_3 = 2.5$. Now sample configurations appear completely repulsive (Fig. 71), which is in fact confirmed by the *L*-function plot for this model (Fig. 72). Without the repulsive third order interaction, the longer range attractiveness overwhelms the short range repulsion of the pair potential, but when the repulsive third order interaction is in place the longer range attractiveness of the pair potential is not even detectable. How attractiveness is controlled in Gibbs point processes is essential to what sample configurations the process will give, in spite of the form of the potentials used. These results suggest that efforts to develop very complex pair potentials that incorporate attractiveness (Degenhardt and Pofahl 2000) may not be justified. Since the Gibbs likelihood is proportional to the exponential

of the *sum* contribution of all interacting points' potentials, complex features of potential functions appear to get lost in this sum when evaluating the likelihood ratio of two particular configurations in a Metropolis-Hastings sampler.

Aside from specifying the attractive and repulsive interactions of Gibbs point processes, a few other points about Gibbs point process models are worthy to note. First, the use of the edge correction to the Gibbs likelihood (Diggle et al. 1994) did not appear to sufficient to prevent edge-effect bias when strong spatial attraction was modeled in the GTP (Figs. 46, 48, 72), EXPTP (Fig. 74), and SPTP (Fig. 75, 76) processes. The bias can be seen for these simulations by the presence of a higher point density at the plot edge as opposed to the plot interior. Apparently the powerful attraction in the pair potential leads to a greater contribution to the Gibbs process likelihood by forming more clusters than is lost by the edge correction weights. In cases of strong attraction, it seems advisable to either collect point pattern data outside the study plot to act as a guard region or alternatively to use other edge corrections with the Gibbs process models, such as the periodic correction. A price to be paid by using a periodic correction with the Bayesian inference used here is that necessarily more triples and pairs will have to be kept track of by the Gibbs posterior sampler as points interact from opposite plot edges, slowing down iteration times.

A second feature of the Gibbs models was that the determination of the interaction distance was problematic. There was no compelling biological reason to choose one particular value of this parameter, such as the choice of the 5m interaction distance for the inference of the center 30 x 30m of Coweeta Plot 318. For the observed data at Coweeta, no measurements were recorded on tree height, crown class, or crown

size that could provide a basis for establishing a particular interaction distance. Studies of the effect of inter-tree competition on other North American hardwood forests provided a range of possible values to use for the interaction distance. In studies of Mississippi Valley bottomland harwoods, Dimov et al. (2008) found that competitor trees that were equal in height or greater than desired red oak (*Quercus pagoda* Raf., *Quercus nigra* L., *Quercus nuttallii* Palmer) crop trees and were within 11 meters of the subject tree had the strongest negative correlation with the 5-year radial growth of crop trees. In contrast trees located from 14m to 17m away from the crop trees showed positive correlation with the 5-year radial growth of crop trees. In the stands observed by Dimov et al. (2008), which had on average 31.68 m^2 ha^{-1} of basal area and were 94 – 139% stocked, the 11m interaction distance observed corresponded to approximately 2.4 times the mean overstory crown radius. In comparison, Lorimer (1983) argued that for the purposes of identifying competitor trees for distance-dependent competition indices, a variable search radius corresponding to 3.5 times the mean overstory crown radius should be used. Other studies of competition in northern hardwood forests that employed fixed search radii used 7.62m (Holmes and Reed 1991), 1.2 – 10.6m (Shi and Zhang 2003), and 10m (Kittredge Jr. 1988) distances to identify competitor trees. In an ecological study of mixed white oak (*Quercus alba* L.) and tulip poplar (*Liriodendron tulipiferai* L.) stands in eastern Tennessee, Doyle (1983) used a search radius of 8m. In contrast to these studies, Zhao et al. (2006) proposed using a fixed 25m search radius to identify competitors for a distance-dependent individual tree diameter growth and survival model developed from a mixed (29 species including *Magnolia grandiflora* L., *Nyssa sylvatica* Marsh., and *Pinus glabra* Walt.) temperate forest in the Southeastern USA.

Even when a fixed interaction distance can be determined, the interaction distance in an unmarked Gibbs point process applies equally for all trees, regardless of the size and distribution of their crowns and root systems. A more realistic model would be a marked Gibbs process with some measure of tree size as a mark variable, such as a discrete crown class rating (e.g. dominant, co-dominant, etc.), where the interaction distances would depend upon the marks (crown classes) of the interacting points. One way that these interaction distances could be chosen is through a scheme based upon horizontal point sampling:

1) Determine an estimate of the average tree DBH by crown class, perhaps through prior knowledge/inventories of stands with similar structure to that being analyzed.
2) Use a desired angle gauge, such as an optical prism to determine the radius of influence r_c of a tree in each crown class c through

$$r_c = \frac{\overline{DBH_c}}{2\sin(\alpha/2)}$$

where $\overline{DBH_c}$ is the average DBH of trees in crown class c and α is the angle projected by the angle gauge.
3) The interaction distance between trees of crown classes c and c' is then the sum of the respective radii of influence, i.e. $R_{c,c'} = r_c + r_{c'}$

As a numerical illustration, suppose that the average co-dominant tree had a DBH of 40 cm and the average intermediate tree had a DBH of 20cm. For a 4.60 m^2 ha^{-1} basal area factor (BAF) prism (equivalent to 20 ft^2 ac^{-1} in English units) the projected angle is α = 2.46 degrees. The radius of influence for the co-dominant class is 9.33m; similarly the radius of influence for the intermediate class is 4.66m. Therefore for interaction between co-dominant and intermediate trees, an interaction distance of approximately 14m could be used. Conversely, for the co-dominant – intermediate interaction distance to be 5m (the choice of interaction distance for all trees in the analysis of the center 30x30m of Coweeta Plot 318), a projected angle of 6.88 degrees is needed, equivalent to a 36 m^2 ha^{-1} BAF prism (156.82 ft^2 ac^{-1} in English units), if in fact a prism with such large BAF exists. The idea of a tree's radius of influence being determined by a horizontally projected angle is not new; Lemmon and Schumacher (1962), Opie (1968), and Daniels (1976) all have discussed this idea in the context of tree competition indices. Furthermore the literature on competition indices offers many ideas on how different measures of tree size can be used to determine a tree's zone of influence (Biging and Dobbertin 1992). One possibility for example would be to project a vertical angle (perpendicular to the ground) to the top height of a class of trees to determine a region of influence as proposed by Biging and Wensel (1990).

While a system for choosing interaction distances like the one above offers the advantage of specifying a series of interaction distances that are proportional to tree size, the problem of choosing an interaction distance is simply replaced with a new problem of choosing an appropriate angle to determine the radius of influence. Furthermore, the choice of interaction distance from a system such as this may not be consistent with a

statistical analysis of the data such as the cusp point method (Stoyan and Grabarnik 1991). The interaction distance can be determined from the statistical inference methodology along with all other parameters of the Gibbs likelihood, for example Bognar (2005, 2008) imposed a prior on the interaction distance and estimated a marginal posterior distribution for this parameter. In the work here however, the interaction distance was kept fixed for analytical and computational simplicity. One drawback to estimating the interaction distance in the context of Bognar's algorithm is that for each interaction distance proposed, the list of interacting pairs and triples (from the data) must be determined for multiple interaction distances. In the implementation of Bognar's algorithm here, this implies that at each iteration of the posterior sampler, pairs and triples will be needed for the current value of the interaction distance, the proposed interaction distance, and also the average of the two interaction distances, which is used for drawing the importance samples from Geyer and Møller's (1994) algorithm. Moreover, to evaluate the normalizing constant ratio, the list of interacting pairs and triples must be generated for each of the these three interaction distances for each importance sample, substantially increasing the computational burden of the algorithm. Nor is this difficulty limited to Bognar's algorithm, since the list of pairs and triples using each of these three interaction distances would be needed to evaluate the auxiliary point configuration at each iteration of the Gibbs posterior sampler in the context of Møller et al.'s algorithm. Given the very slow iteration times of both of these algorithms when applied to triplets processes, having fixed interaction distances prior to any statistical analysis still seems the most practical choice.

One key result from Bayesian inference with triplets processes, whether utilizing Bognar's algorithm or Møller et al.'s algorithm, is that while it is possible to do, it is very computationally demanding. In order to approximate ratios of normalizing constants appearing in the likelihood ratios of a Metropolis-Hastings posterior sampler, another Metropolis-Hastings algorithm (e.g. that of Geyer and Møller 1994) was implemented. In effect both algorithms use a nested-MCMC structure, which is viewed by some as intractable (Illian et al. 2008, p. 476). As a consequence of this structure, the Gibbs posterior sampler, using either Bognar's or Møller et al.'s algorithm had very long iteration times. Even for a plot with few points, such as the center 30 x 30m of Coweeta Plot 318 with 60 trees, iteration times ranged from approximately 30 seconds to two minutes, leaving runs of only 20,000 iterations needing from 10 days to nearly a month to complete. Compounding this problem is a number of specific issues with the triplets models analyzed here which promote even longer iteration times. First, the choice of burn-in of Geyer and Møller's algorithm depends upon the degree of attractiveness present in the data being analyzed. Gibbs parameter values that reflect strong localized clustering need longer burn-ins for Geyer and Møller's algorithm to generate a sample with these features. For all of the triplets models tested here, when strong 2^{nd}-order attraction is joined with weak 3^{rd}-order repulsion, Geyer and Møller's algorithm can easily require a burn-in of 50,000 iterations or more (Figs. C1 – C3, C7 – C12). One important consequence in practice of trying to use too small a burn-in in order to save computation time was that normalizing constant ratios and acceptance probabilities were significantly larger when moving to parameter vectors with stronger 2^{nd}-order attraction. While the true normalizing constant ratios are unknown, the t-test results from the SPTP

process with the SimTest50 data indicated that the Gibbs posterior sampler is moving in a biased way towards parameter vectors with more extreme attraction when Geyer and Møller's algorithm is not properly burned in (Table 6). Given the negative cross correlation found between the strength of attraction in the 2nd-order potential with the strength of repulsion in the 3rd-order potential in all models (Tables 10, 12), the bias towards parameters with strong 2nd-order attraction will promote large values of the parameter of the repulsive 3rd-order potential as well.

Another issue which promotes longer iteration times for Bognar's algorithm is the apparent inverse relationship between the quality of CVs of the importance samples and the quality of mixing in the parameter trace plots. As pointed out by Geyer (1999) and Bognar (2004, 2005), in order for importance sampling to provide a good estimate of a ratio of normalizing constants, the two parameter vectors which define the ratio must not be too far apart numerically. Proposing parameter vectors that are incrementally close to the current state would give good estimates of the normalizing constant ratio, but incremental proposals necessarily means that the Markov chain must explore the support of the posterior incrementally. If instead rapid mixing is desired, proposed parameter vectors could be quite far from the current state of the chain and the normalizing constant ratio would be compromised (Bognar 2004). Results from Bognar's algorithm here indicate that better mixing (better trace plots, lower autocorrelation) lead to higher CVs of importance sampling weights and vice versa. Deliberately using slowly mixing chains in order to achieve good normalizing constant ratios combined with a very large number of iterations the algorithm is run is one solution in principle, though not a practical one given the iteration times found here. Instead moderately mixing chains using larger

numbers of importance samples might be able to reduce excessively large CVs of importance sampling weights. Of course, the price for this is significantly longer iteration times for the Bognar's algorithm, since Geyer and Møller's (1994) algorithm would need to be run for a much greater number of iterations.

Within the context of Bognar's algorithm, a number of influential factors were explored, in particular the effect of point density, interaction distance, and choice of triplets models on the mixing of the chains for each parameter and on the fit of the model to the simulated or real data that was used as the basis for inference. Of all of these factors, the choice of the triplets model had the most noticeable effect on the trace plots. Given the limited tuning of the chains for the various simulation trials, the SPTP process repeatedly showed poor mixing, with chains that often did not oscillate around the known target parameter values and, when used for inference on the observed data at Coweeta, did not oscillate around any common value at all. The likely cause for this poor mixing is the great flexibility of the serpentine pair potential used in the SPTP process. Different combinations of values for the parameters of this pair potential can given similar shapes, even though some combinations may have positive values for θ_1 and negative values for θ_2 and other combinations will have negative values for θ_1 and positive values for θ_2. One obvious way to improve this would be to restrict the parameter space of the potential so that, for example, θ_1 could only take on positive values. However, even when parameters of the serpentine potential do not change sign, poor mixing of the SPTP process still occurs, as evidenced by the trace plots of the density test from the SimTest50 dataset (Figs. 83 – 86). When all three models were used to draw inference from the Coweeta

data, better mixing was obtained from the EXPTP and GTP processes (Figs. E1 – E6) than from the SPTP process (Figs. E7 – E9).

Despite the differences in mixing, both the EXPTP and SPTP models showed the ability to reproduce the spatial structure of the simulated or real data for which inference was desired (Figs. 81, 82, 88, 89, 94, 99, 103 – 106, 108, 110). The poor mixing of the SPTP process often left large numerical differences between the estimated posterior modes and the known target parameters of simulated plots (Tables 5, 7). Despite this, the Monte Carlo envelopes for the various tests seemed to be insensitive to these numerical differences, particularly for excessively large values of the parameter of the repulsive 3rd-order potential (Figs. 81, 82). With regard to the GTP process, one feature of this model that became apparent through both simulation of and inference with this process was the inability of this model to provide the same type of localized clustering as found in Plot 318 from Coweeta (Fig. 48, 50, 110). The pair potential of this model allows for the same level of point attraction throughout its interaction range, and simulations from this model with strong attractiveness show the tendency to develop long "chains" of clustered points (Figs. 46, 48). Since it is of interest to model clustering from stump sprouting, it seems more realistic that the point attraction should be able to decrease rapidly after distance greater than 1m. Given the limitations of an inflexible pair potential in the GTP process and a too flexible potential in the SPTP process, the EXPTP process is considered the best model for the types of spatial point patterns observed in the Coweeta data. Certainly other model forms of triplets processes can be considered, but relatively simple 2nd- and 3rd-order potentials with few parameters seem preferable over more complex potentials.

In contrast to the large qualitative differences between the different triplets models used to do inference, effects of point density and the value of the (fixed) interaction distance were much more difficult to ascertain. No clear trends were evident from the study of point density using the SPTP process with strong attraction and repulsion and plot sizes of 7 x 7m through 12 x 12m (Figs. 83 – 86, 88, 89). While lower plot densities showed slower mixing with more extreme parameter values being sampled (Figs. 83 – 86), the highest plot density tested (using the 7 x 7m plot) had the worst fit of the Monte Carlo envelope of the *L*-function (Fig. 90). The test of the interaction distance with the EXPTP process with weak attraction showed slower mixing with larger interaction distances (Figs. 100 – 102) but inference with smaller interaction distances showed noticeably larger CVs of importance sample weights, given the same set of tuning parameter values (Figs. D21 – D24). All of the interaction distance trials gave reasonable parameter estimates (Table 8) corresponding to Monte Carlo envelopes that enclosed the *L*-functions generated from the simulated datasets (Figs. 103 – 106).

One particular challenge to testing variables such as point density and interaction distance was being able to design a test that could isolate one variable from the effects of others. Point density can me modified in one of two ways, either by changing the number of points or by changing the size of the plot within which the points lie. Keeping the plot size and interaction distance fixed and changing the point density requires a sufficiently large plot size so that at the low end of point density there is still a sufficient number of points for inference. Yet at the same time, at high point densities, large plot sizes imply a very large number of points, which will lead to very large numbers of interacting pairs and triples and very slow iteration times. If the iteration time is desired to be kept under

control, then variable point density can be studied by selecting a target number of points and changing the plot size. For a fixed interaction distance however, decreasing plot size can lead to increased edge-effect bias, as more and more points' interaction distances extend beyond the plot boundary. Testing the interaction distance for a desired number of points has unintended consequences. First, for partially attractive processes, smaller interaction distances result in greater point density for the same set of potential parameters as do larger interaction distances. This occurs because the limiting action of the repulsive 3rd-order potential on cluster formation is restricted to smaller areas, allowing more points and clusters to form in a given fixed area. So for a given value of the intensity parameter (β), it becomes progressively difficult to keep the number of points constant with increasing range of interaction distances being tested. Additionally, larger interaction distances necessitate the calculation of more interacting pairs and triples for any given point. In spite of the tendency for large interaction distances to result in sample configurations with smaller numbers of points than configurations generated from smaller interaction distances, the total number of interacting pairs and triples generated from large interaction distances is much larger, leading to very slow iteration times. The 10m interaction distance trial, for example, had an average iteration time of 179.12 seconds, implying that almost 6 weeks (41.5 days) of continuous running time would have been needed to complete the 20,000 iterations of Bognar's algorithm for a plot that contained only 50 points. In contrast, the 4m interaction distance trial had an average iteration time of 27.0 seconds, needing a total of 6.25 days to complete 20,000 iterations using a configuration that contained 58 points.

Regarding the choice between Bognar's algorithm using importance sampling and Møller et al.'s auxiliary variable algorithm some observations can be made. When perfect simulation of Gibbs point processes is feasible, Møller et al.'s algorithm is theoretically sounder. The approximation of the normalizing constant ratio by importance sampling prevents the algorithm from achieving detailed balance, and there is no theory to assure the analyst that the posterior samples being drawn are from the correct distribution. That said, in this study of triplets processes, perfect simulation was not feasible because of the high upper bound of the conditional intensity function of the triplets processes. In particular, the ability of the triplets processes to have partially or completely attractive 2^{nd}-order interactions results in a high upper bound on the conditional intensity function (Fig. 112). Had purely repulsive 2^{nd}-order interactions been used, the conditional intensity of the triplets process would be bounded by the processes' intensity parameter, and perfect simulation would be feasible, as demonstrated by Møller et al. (2006) and Berthelsen and Møller (2006) for the purely repulsive Strauss process.

In the absence of perfect simulation, the primary theoretical benefit of Møller et al.'s algorithm is muted. The imperfect scheme of using Geyer and Møller's (2004) algorithm to draw samples of the auxiliary variable allowed the possibility of Møller et al.'s algorithm to be implemented for the triplets processes considered here, and it was faster to use than Bognar's algorithm. However, Møller et al.'s algorithm repeatedly generated runs of non-acceptance lasting hundreds and occasionally even thousands of iterations, a phenomenon that was also noted by Møller et al. (2006) (Figs. F1 – F13). In the context of the relatively short 20,000 iteration runs of the algorithm used for inference on Coweeta Plot 318, such runs of non-acceptance forced the joint posterior mode

solution to match the longest run obtained and distorted the appearance of the smoothed posterior density (Figs. F14 – F16, F24 – F26). One possibility for addressing this issue would be to run Møller et al.'s algorithm for such a long time that any particular run would not overly influence the joint posterior. Møller et al. (2006) chose to use tuning parameters that generated incremental proposals to keep the number and length of runs of non-acceptance to a minimum, although this implies that mixing will be necessarily slow.

For the analysis of mixed attractive and repulsive spatial structure like that found in the Coweeta data, Bognar's algorithm appears to be preferable to Møller et al.'s algorithm. Given the lack of feasible perfect simulation for triplets processes, Bognar's algorithm offers an alternative that avoids the problems associated with long runs of non-acceptance. Additionally, there is no software available for estimating parameters of triplets processes, so the fixed parameter estimate needed for the conditional distribution of the auxiliary variable is not easy to come by. Møller et al. (2006) and Berthelsen and Møller (2006) do propose alternate choices for the conditional distribution of the auxiliary variable, including partially ordered Markov models (POMMs) proposed by Cressie and Davidson (1998). However given the already heavy computational burden inherent with Bayesian inference for triplets processes, no attempt was made to utilize POMMs for this study.

At the outset of this research, one primary objective was to formulate and evaluate Gibbs point processes that could be flexible enough to handle the types of patterns encountered in mixed-species Appalachian forests, particularly the presence of localized clustering due to stump sprouting. If such processes were formulated, could the Gibbs processes simulate patterns of mixed-species Appalachian hardwood forests distribution

and how tractable would such processes be as a forest stand structure generator? The results of this research indicate that Gibbs point processes with suitably repulsive 3rd-order interactions can realistically simulate patterns of such forests (Figs. 64, 65, 108, 110). While the spatially random pattern of Coweeta Plot 218 (Fig. 14) does not appear to be difficult to simulate by a triplets processes, the patterns of Coweeta Plots 118, 427, and 527 (Figs. 13, 16, 17) show clustered but no repulsive interaction. This clustering is due both to stump sprouting and to inhomogeneity of pattern, as evidenced for example in the stem map of Plot 527 (Fig.17), where the tree density is noticeably lower in the lower right-hand corner of the plot. It is unlikely that these patterns could be realistically simulated with a triplets model with a spatially constant first order interaction, as those proposed here. To model these patterns inhomogeneous Gibbs processes would be needed. Spatial inhomogeneity can be incorporated in Gibbs point processes with repulsive 2nd-order interactions (Stoyan and Stoyan 1998; Bognar 2005) through the use of a non-constant intensity function, an approach which can be extended to the triplets models studied here. For Markov point processes generally, inhomogeneity can be introduced by thinning of homogeneous Markov point processes (Baddeley et al. 2000), transformation of homogeneous Markov point processes (Jensen and Nielsen 2000; Nielsen and Jensen 2004), or construction of a locally scaled Markov point process (Hahn et al. 2003). The primary reason that inhomogeneous Gibbs processes were not considered is that fluctuations in tree density across one research plot are specific to that plot and do not necessarily fit spatial inhomogeneity found in other research plots. In absence of a direct biological mechanism for the inhomogeneity, building a point process model to conform to observed spatial inhomogeneity on a particular plot would then

undesirably restrict the application of the model to that particular plot. Indeed, from the point of view of building a stand structure generator for Appalachian hardwoods that could be applied generally in the Southern Appalachian mountains, it seems necessary to either assume spatial homogeneity, or alternatively, to have the user of the stand structure generator input sufficient information to quantify the location and intensity of any inhomogeneity that is desired in the forest simulation. In contrast, the features of tree clustering from stump sprouts and tree repulsion from competition may reasonably be considered as features found in the spatial pattern of Appalachian hardwood forests in general.

The question of the tractability of Gibbs point processes for use with a stand structure generator depends upon how such a generator might be constructed. Analogous to such a spatial stand structure generator, tree diameter distribution models give many examples of how continuous probability density functions (pdfs) can be used to serve as a stand table generator (Packard 2000), where the distribution of the number of trees into diameter classes can be obtained once the parameters of the pdfs have been estimated through other forest stand information. In general the methods of parameter prediction, parameter recovery, and percentile prediction are the most popular means for parameter estimation. In the first approach, k parameters are predicted through linear or nonlinear regression models using various forest stand covariates. In parameter recovery, moments of the tree diameter distribution, such as the mean and quadratic mean, are predicted for a stand and the pdf parameters are "recovered" by using the method of moments to solve a system of equations containing k unknown pdf parameters and k predicted moments. In

percentile prediction, *k* order statistics of the tree diameter data are predicted and used to solve a system of equations for the *k* unknown parameters in a similar way.

The principles of diameter distribution methodology could work to build a stand structure generator. However, Gibbs point process densities are more difficult to work with, particularly because they are defined on classes of subsets (σ-fields) of finite point configurations in $\mathbb{R}^d$ and not on the Borel sets of $\mathbb{R}$ (such as intervals), as are continuous univariate pdfs used with diameter distribution models. Percentile prediction methodologies are not applicable, because without any natural ordering of point configurations there is no meaningful way to define order statistics for a Gibbs point process. While the parameter prediction approach can be applied to any parameterized density, several objections limit the usefulness of this approach. First, there are no meaningful biological relationships between pdf parameters and forest stand variables. Secondly, there is an inherent statistical inconsistency in using linear regression models to predict parameters. If one adopts the classical statistical inference point of view, parameters are fixed constants that do not have probability distributions. Yet the parameter prediction approach stipulates that the parameters do have a probability density; for linear regression models, for example, the response variable (in this case a pdf parameter) follows a normal distribution (Rencher 2000). Suppose instead that a Bayesian inference point of view is adopted, as was the case here. The Gibbs point process parameters do follow a probability distribution, the joint posterior distribution. But this posterior density is not the result of linear or non-linear combination of fixed covariates and a normally distributed error term, so standard methodology for linear regression models does not apply.

The best methodology for building a forest stand structure generator with a spatial point process pdf seems to be the method of moments. As noted by Fiksel (1984), the Takacs-Fiksel method (see Sec. 2.31) is an application of the method of moments, where the sample moments are estimates from the second-moment measures of the Gibbs point process, under the assumption of a stationary and isotropic Gibbs process. Tomppo (1986) proposed a straightforward way to estimate these moments from data collected from the forest itself. In particular, a series of randomly located trees and a separate sample of randomly located circular fixed radius plots are installed within the forest of interest. For each randomly located tree, the distance to its nearest neighbor is recorded. For each sample plot the number of trees are counted in m circles of increasing radius that match the n steps of a multiscale pairwise interaction process (Penttinen 1984). These counts and nearest neighbor measurements would then be combined to solve a system of m equations for the $n + 1$ unknown Gibbs point process parameters. As noted by Tomppo, the Takacs-Fiksel method does not require a multiscale pairwise interaction process model. However, Tomppo's use of field plots would need to be modified slightly. In particular, a smooth potential function would require that for each of the m sample plots, all trees would need to be stem-mapped so that the potential could be evaluated. Furthermore, Tomppo recommended that a sample of at least 60-100 plots and 60-100 trees be used so that the variance of the moment estimates is not too large. All of the plots and trees would need to be sufficiently located away from the edge of the stand, and the plots would need to be larger in diameter than the interaction radius of the Gibbs point process model.

Clearly, such an approach is expensive to carry out, and in small stands the sampling effort might rival the time needed to stem map the stand, eliminating the need for a stand structure generator altogether. A further limitation is that the method is only applicable to homogeneous Gibbs point processes; there is currently no published work on how an inhomogeneous Gibbs process' parameters might be estimated through the method of moments. In fact, even methods for spatial summary functions that use 2^{nd}-order moment measures of inhomogeneous point processes have themselves only recently developed (Baddeley et al. 2000).

6. Conclusions

Although Gibbs point processes are particularly well suited to model repulsive point patterns (Illian et al. 2008), Gibbs point processes with 3rd-order interactions can be developed that simulate features of clustering and overdispersion as found in Appalachian hardwood forests. In this thesis, the spatial structure of five stem-mapped plots of mixed Appalachian hardwoods was explored with *L*-functions and pair correlation functions, where each plot was evaluated over a range of minimum tree diameter limits. Spatially non-random clustering at distances less than 2m was found in four of the five plots that was consistent with the presence of clumps of stump sprouts. In all five plots, non-random point overdispersion was detected. In three plots, non-random overdispersion increased with increasing tree diameter limits, occurring at distances from 2 to 10m for Plots 218 and 427 and at distances from 4 to 6m for Plot 527. While this is consistent with the notion of overdispersed patterns resulting from intertree competition over time (Christensen 1977; Ward et al. 1996; Kenkel et al. 1997), no such causal mechanism need be assumed. The analysis of the spatial structure of the Coweeta data indicated that models for the spatial pattern of Appalachian hardwoods should, however, account for both short range attractive and intermediate-range repulsive interactions.

Models for the 2nd- and 3rd-order potential functions of locally stable Gibbs point processes were proposed as a means of simulating the patterns founding in the Coweeta data. When partly or completely attractive 2nd-order potentials were specified in a pairwise interaction process, simulations from Geyer and Møller's (1994) Metropolis-Hastings algorithm typically gave excessively clustered patterns. Introducing purely

repulsive 3rd-order interactions allowed for the simulation of short range clustering and intermediate range overdispersion simultaneously. However, when the 3rd-order potential has weak repulsion relative to the attraction in the 2nd-order potential, simulations from the proposed triplets models can result in samples from a "multi-clump" process that is not realistic of Appalachian hardwood forests. This result emphasizes the need for a balance between attractive and repulsive potentials so that the clustering introduced in the Gibbs process is not excessive.

One feature observed in the Coweeta dataset, particularly in Plot 527, was the presence of longer range (10 − 20m) point attraction due to spatial inhomogeneity. While it was not the objective here to model or simulate such inhomogeneity, the presence of it in a small dataset such as the five Coweeta plots indicates that models for the spatial pattern of mixed Appalachian hardwoods will need to take this into account. Certainly the Gibbs point processes proposed here can be adapted to model inhomogeneity by allowing the constant first-order interaction (given by the parameter β) to be a function of location, as demonstrated by the inhomogeneous pairwise interaction processes proposed by Bognar (2005). Doing so may result in the 1st-order interaction be tailored specifically to the observed data (Stoyan and Stoyan, 1998) however, and may limit the application of the model to other datasets.

While Bayesian inference was pursued as a method for estimating Gibbs point process parameters, it was found to be very computationally expensive. One to four weeks of continuous running time was required for relatively short (20,000 iterations) runs of the tested algorithms, despite the programming of the algorithms in C++ and repeated attempts to exploit computational shortcuts. Both the importance sampling

based approach of Bognar (2004, 2005) and the auxiliary variable approach of Møller et al. (2006) necessitated the use of nested MCMC methods, which remains as the primary obstacle to improving computational speed. While Møller et al.'s auxiliary variable approach offers significant theoretical advantages when perfect simulation of Gibbs point processes is practical, the partially attractive models here have such large upper bounds on their respective conditional intensity functions that perfect simulation via dominated coupling from the past (CFTP) was not practical. Regardless of the algorithm used, the trace plots for the proposed triplets models indicated slow mixing. When Bognar's algorithm was used, slow mixing became a consequence of the need for incremental parameter proposals to accurately estimate the normalizing constant ratio, whereas in Møller et al.'s algorithm, incremental proposals were used to minimize the effects of long runs of non-acceptance. Tests of Bognar's algorithm with configurations representing different point density, spatial structures, and model interaction distances failed to generate clear trends with poor mixing found in all conditions. The poor mixing combined with short (20,000 iterations or less) runs requiring long computation times was a major obstacle to drawing clear inferences. When Bognar's and Møller et al.'s algorithms were applied to a subset of one of the five Coweeta plots (Plot 318), the obtained posterior modes from each of the three tested triplets models provided *L*-function Monte Carlo confidence envelopes that nearly entirely contained the observed *L*-function. While this offers some promise that Bayesian inference could be used to estimate parameters of partially attractive Gibbs point processes, the heavy computational burden of the algorithms tested leads to the conclusion that such methods are not practical

for the construction and application of Gibbs point process models in forest management settings.

The theory and applications of Gibbs point processes and spatial point processes more generally continues to rapidly evolve. Areas of particular interest of current research include the development of residual analyses for point processes (Baddeley et al. 2005; Baddeley et al. 2008), models and inference for inhomogeneous Markov processes (Berthelsen and Møller 2008; Møller and Helisová 2008), and spatio-temporal marked point processes (Särkkä and Renshaw 2006; Comas and Mateu 2007). With continuing advancement of point process theory, it is expected that new point process models and methods for their inference will allow for greater realism in modeling the complex spatial and temporal dynamics of mixed-species forests.

7. References

Allen, M.P., and Tildesley, D.J. 1987. Computer simulation of liquids. Clarendon Press, Oxford, UK. pp.

Avery, T.E., and Burkhart, H.E. 2002. Forest Measurements. 5th ed. McGraw-Hill, Inc, New York. 456 pp.

Baddeley, A. 2001. Likelihoods and pseudolikelihoods for Markov spatial processes. *In* State of the Art in Probability and Statistics: Festschrift for Willem R. van Zwet. *Edited by* M. de Gunst, C. Klaassen, and A. van der Vaart. Institute of Mathematical Statistics, Beachwood, Ohio. pp. 21-49.

Baddeley, A. 2007. Spatial point processes and their application. *In* Stochastic geometry: Lectures given at the C.I.M.E. Summer School held in Martina Franca, Italy, Sept. 13-18, 2004. *Edited by* W. Weil. Springer, New York, NY. pp. 1-75.

Baddeley, A., and Møller, J. 1989. Nearest-neighbour Markov point processes and random sets. International Statistics Review **57**(2): 89-121.

Baddeley, A., Møller, J., and Pakes, A.G. 2008. Properties of residuals for spatial point processes. Annals of the Institute of Statistical Mathematics **60**: 627-649.

Baddeley, A., Møller, J., and Waagepetersen, R.P. 2000. Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. Statistica Neerlandica **54**(3): 329-350.

Baddeley, A., and Turner, R. 2000. Practical maximum pseudolikelihood for spatial point patterns. Australia and New Zealand Journal of Statistics **42**(3): 283-322.

Baddeley, A., and Turner, R. 2005. Spatstat: an *R* package for analyzing spatial point patterns. Journal of Statistical Software **12**: 1-42.

Baddeley, A., and Turner, R. 2008. The Spatstat package [online]. Available from http://www.spatstat.org/spatstat [cited May 10, 2008].

Baddeley, A., Turner, R., Møller, J., and Hazelton, M. 2005. Residual analysis for spatial point processes. Journal of the Royal Statistical Society. Series B **67**(5): 617-666.

Baddeley, A., and Van Lieshout, M.N.M. 1995. Area-interaction point processes. Annals of the Institute of Statistical Mathematics **47**(4): 601-619.

Baddeley, A., Van Lieshout, M.N.M., and Møller, J. 1996. Markov properties of cluster properties. Advances in Applied Probability **28**(2): 346-355.

Batista, J.L.F., and Maguire, D.A. 1998. Modeling the spatial structure of tropical forests. Forest Ecology and Management **110**: 293-314.

Beck, D.E. 1977. Growth and development of thinned versus unthinned yellow-poplar sprout clumps. USDA For. Serv. Res. Pap. SE-173.

Beck, D.E., and Hooper, R.M. 1986. Development of a Southern Appalachian hardwood stand after clearcutting. South. J. Appl. For. **10**: 168-172.

Berman, M., and Turner, R. 1992. Approximating point process likelihoods with GLIM. Applied Statistics **41**(1): 31-38.

Berthelsen, K.K., and Møller, J. 2002. A primer on perfect simulation for spatial point processes. Bulletin of the Brazilian Mathematical Society **33**(3): 351-367.

Berthelsen, K.K., and Møller, J. 2003. Likelihood and non-parametric Bayesian MCMC inference for spatial point processes based on perfect simulation and path sampling. Scandinavian Journal of Statistics **30**: 549-564.

Berthelsen, K.K., and Møller, J. 2006. Bayesian analysis of Markov point processes. *In* Case studies in spatial point process modeling. *Edited by* A. Baddeley, P. Gregori, J. Mateu, R. S. Stoica, and D. Stoyan. Springer-Verlag, New York, NY. pp. 85-97.

Berthelsen, K.K., and Møller, J. 2008. Non-parametric Bayesian inference for inhomogeneous Markov point processes. Australia and New Zealand Journal of Statistics **50**(3): 257-272.

Besag, J.L.F. 1975. Statistical analysis of non-lattice data. The Statistician **24**: 179-195.

Besag, J.L.F. 1977. Some methods of statistical analysis for spatial data. Bulletin de l'Institut Internationale de Statistique **47**(2): 77-92.

Besag, J.L.F., Milne, R., and Zachary, S. 1982. Point process limits of lattice processes. Journal of Applied Probability **19**(1): 210-216.

Biging, G.S., and Dobbertin, M. 1992. A comparison of distance-dependent competition measures for height and basal area growth of individual conifer trees. Forest Science **38(3)**: 695-720.

Biging, G.S., and Wensel, L.C. 1990. Estimation of crown form for six conifer species of northern California. Can. J. For. Res. **20**(8): 1137-1142.

Billingsley, P. 1995. Probability and measure. 3rd ed. John Wiley & Sons, New York, NY. 593 pp.

Bognar, M.A. 2002. Bayesian Estimation of a Potential Function in a Pairwise Interacting Point Process. PhD, University of Iowa, Iowa City, IA.

Bognar, M.A. 2004. Bayesian inference for pairwise interacting point processes. Statistics and Computing **14**: 109-117.

Bognar, M.A. 2005. Bayesian inference for spatially inhomogeneous pairwise interacting point processes. Computational Statistics and Data Analysis **49**: 1-18.

Bognar, M.A. 2008. Bayesian modeling of continuously marked spatial point patterns. Computational Statistics **23**(3): 361-379.

Bormann, F.H. 1953. The statistical efficiency of sample plot size and shape in forest ecology. Ecology **34**(3): 474-487.

Casella, G., and Berger, R.L. 2002. Statistical inference. 2nd ed. Duxbury, Pacific Grove, CA. 660 pp.

Casella, G., Lavine, M., and Robert, C.P. 2001. Explaining the perfect sampler. The American Statistician **55**(4): 299-305.

Chen, M.-H., and Shao, Q.-M. 1997. On Monte Carlo methods for estimating ratios of normalizing constants. The Annals of Statistics **25**(4): 1563-1594.

Chen, M.-H., and Shao, Q.-M. 1999. Monte Carlo estimation of Bayesian credible and HPD intervals. Journal of Computational and Graphical Statistics **8**(1): 69-92.

Christensen, N.L. 1977. Changes in structure, pattern, and diversity associated with climax forest in Piedmont, North Carolina. American Midland Naturalist **97**(1): 176-188.

Clark, J.S., Macklin, E., and Wood, L. 1998. Stages and spatial scales of recruitment limitation in southern Appalachian forests. Ecol. Monogr. **68**(2): 213-235.

Clark, P.J., and Evans, F.C. 1954. Distance to nearest neighbor as a measure of spatial relationships in populations. Ecology **35**(4): 445-453.

Comas, C., and Mateu, J. 2007. Modelling forest dynamics: A perspective from point process methods. Biometrical Journal **49**(2): 176-196.

Copenheaver, C.A., Matthews, J.M., Showalter, J.M., and Auch, W.E. 2006. Forest stand development patterns in the southern Appalachians. Northeastern Naturalist **13**(4): 477-494.

Cowles, M.K., and Carlin, B.P. 1996. Markov chain Monte Carlo convergence diagnostics: A comparative review. Journal of the American Statistical Association **91**(434): 883-904.

Cox, D.R., and Isham, V. 1980. Point processes. Chapman and Hall, New York, NY. 188 pp.

Cressie, N.A.C. 1993. Statistics for spatial data. Revised ed. John Wiley & Sons, New York, NY. 900 pp.

Cressie, N.A.C., and Davidson, J.L. 1998. Image analysis with partially ordered Markov models. Computational Statistics and Data Analysis **29**(1): 1-26.

Dagley, C.M. 2008. Spatial pattern of coast redwood in three alluvial flat old-growth forests in northern California. Forest Science **54**(3): 294-302.

Daley, D.J., and Vere-Jones, D. 1988. An introduction to the theory of point processes. Springer-Verlag, New York, NY. 702 pp.

Daniels, R.F. 1976. Simple competition indices and their correlation with annual loblolly pine tree growth. Forest Science **22**(4): 454-456.

Degenhardt, A. 1999. Description of tree distribution patterns and their development through marked Gibbs processes. Biometrical Journal **41**(4): 457-470.

Degenhardt, A., and Pofahl, U. 2000. Simulation of natural evolution of stem number and tree distribution pattern in a pure pine stand. Envirometrics **11**: 197-208.

Diggle, P.J. 1986. Parametric and non-parametric estimation for pairwise-interaction point processes. *In* Proceedings of the 1st World Congress of the Bernoulli Society. Tashkent, USSR. 8-14 September 1986. *Edited by* Y. A. Prohorov and V. V. Sazonov. VNU Science Press, Utrecht, The Netherlands. pp. 101-111.

Diggle, P.J. 2003. Statistical analysis of spatial point patterns. 2nd ed. Arnold, London, UK. 159 pp.

Diggle, P.J., Fiksel, T., Grabarnik, P., Ogata, Y., Stoyan, D., and Tanemura, M. 1994. On parameter estimation for pairwise interaction point processes. International Statistics Review **62**(1): 99-117.

Dimov, L.D., Chambers, J., and Lockhart, B.R. 2008. Five-year radial growth of red oaks in mixed bottomland hardwood stands. Forest Ecology and Management **255**: 2790-2800.

Doležal, J., Miroslav, Š., Hara, T., Sumida, A., and Penttilä, T. 2006. Neighborhood interactions influencing tree population dynamics in nonpyrogenous boreal forest in northern Finland. Plant Ecology **185**: 135-150.

Doyle, T.W. 1983. Competition and growth relationships in a mixed-aged, mixed-species forest community. PhD, University of Tennessee, Knoxville, TN.

Ek, A.R., and Monserud, R.A. 1974. FOREST: A computer model for simulating the growth and reproduction of mixed species forest stands. Univ. of Wisconsin, School of Natural Resources. Res. Report R2635. 13 pp.

Ferrari, P.A., Fernández, R., and Garcia, N.L. 2002. Perfect simulation for interacting point processes, loss networks and Ising models. Stochastic Processes and their Applications **102**: 63-88.

Fiksel, T. 1984. Estimation of parametrized pair potentials of marked and non-marked Gibbsian point processes. Elektronische Informationsverabeitung und Kybernetik **20**(5/6): 270-278.

Fiksel, T. 1988. Estimation of interaction potentials of Gibbsian point processes. Statistics **19**(1): 77-86.

Fralish, J.S. 2003. The Central Hardwood Forest: it's boundaries and physiographic provinces. Pages 1-20 *in* J. W. Van Sambeek, J. O. Dawson, F. J. Ponder, E. F. Loewenstein, and J. S. Fralish, editors. 13th Central Hardwood Forest conference. USDA For. Serv. Gen Tech. Rep. NC-234, Urbana, IL.

Gates, D.J., and Westcott, M. 1986. Clustering estimates for spatial point distributions with unstable potentials. Annals of the Institute of Statistical Mathematics **38**: 123-135.

Gavin, D.G., and Peart, D.R. 1997. Spatial structure and regeneration of Tetramerista glabra in peat swamp rain forest in Indonesian Borneo. . Plant Ecology **131**: 223-231.

Gavrikov, V., and Stoyan, D. 1995. The use of marked point processes in ecological and environmental forest studies. Environmental and Ecological Statistics **2**: 331-344.

Gelfand, A.E., and Smith, F.M. 1990. Sampling-based approaches to calculating marginal densities. Journal of the American Statistical Association **85**(410): 398-409.

Gelman, A., and Meng, X.L. 1994. Path sampling for computing normalizing constants: identities and theory. Dept. of Statistics, Univ. of Chicago, Chicago, IL. Technical Report 377. pp.

Gelman, A., and Meng, X.L. 1998. Simulating normalizing constants: from importance sampling to bridge sampling to path sampling. Statistical Science **13**: 163-185.

Geman, S., and Geman, D. 1984. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence **6**: 721-741.

Getis, A., and Franklin, J. 1987. Second-order neighborhood analysis of mapped point patterns. Ecology **68**: 473-477.

Geweke, J. 1989. Bayesian inference in econometrics models using Monte Carlo integration. Econometrica **57**: 1317-1340.

Geyer, C.J. 1996. Estimation and optimization of functions. *In* Markov chain Monte Carlo in practice. *Edited by* W. R. Gilks, S. Richardson, and D. J. Spiegelhalter. Chapman & Hall, New York, NY. pp. 241-258.

Geyer, C.J. 1999. Likelihood inference for spatial point processes. *In* Stochastic geometry: likelihood and computation. *Edited by* O. Barndorff-Nielsen, W. S. Kendall, and M. N. M. Lieshout. Chapman & Hall, New York, NY. pp. 79-140.

Geyer, C.J., and Møller, J. 1994. Simulation procedures and likelihood inference for spatial point processes. Scandinavian Journal of Statistics **21**: 359-373.

Geyer, C.J., and Thompson, E.A. 1992. Constrained Monte Carlo maximum likelihood for dependent data. Journal of the Royal Statistical Society. Series B **54**: 657-699.

Geyer, C.J., and Thompson, E.A. 1995. Annealing Markov chain Monte Carlo with applications to ancestral inference. Journal of the American Statistical Association **90**(431).

Gignoux, J., Duby, C., and Barot, S. 1999. Comparing the performances of Diggle's tests of spatial randomness for small samples with and without edge-effect correction: application to ecological data. Biometrics **55**: 156-164.

Gilks, W.R., Richardson, S., and Spiegelhalter, D.J. 1996. Introducing Markov chain Monte Carlo. *In* Markov chain Monte Carlo in practice. *Edited by* W. R. Gilks, S. Richardson, and D. J. Spiegelhalter. Chapman & Hall, New York, NY. pp. 1-19.

Grabarnik, P., and Särkkä, A. 2001. Interacting neighbour point processes: some models for clustering. Journal of Statistical Computation and Simulation **68**(2): 103-125.

Green, P.J. 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika **82**(4): 711-732.

Gregori, P., and Mateu, J. 2002. Spatial point processes: an overview *In* Spatial statistics through applications. *Edited by* J. Mateu and F. Montes. WIT Press, Southampton, UK. pp. 41-65.

Gu, M.G., and Zhu, H.-T. 2001. Maximum likelihood estimation for spatial models by Markov chain Monte Carlo stochastic approximation. Journal of the Royal Statistical Society. Series B **63**(2): 339-355.

Haase, P. 1995. Spatial pattern analysis in ecology based on Ripley's *K*-function: Introduction and methods of edge correction. Journal of Vegetation Science **6**: 575-582.

Häggström, O. 2002. Finite Markov chains and algorithmic applications. Cambridge University Press, Cambridge, UK. 114 pp.

Häggström, O., Van Lieshout, M.N.M., and Møller, J. 1999. Characterization results and Markov chain Monte Carlo algorithms including exact simulation for some spatial point processes. Bernoulli **5**(4): 641-658.

Hahn, U., Jensen, E.B.V., Van Lieshout, M.N.M., and Nielsen, L.S. 2003. Inhomogeneous spatial point processes by location-dependent scaling. Advances in Applied Probability **35**(2): 319-336.

Hanus, M.L., Hann, D.W., and Marshall, D.D. 1998. Reconstructing the spatial pattern of trees from routine stand examination measurements. Forest Science **44**(1): 125-133.

Hastings, W.K. 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika **57**: 97-109.

Heikkinen, J., and Penttinen, A. 1999. Bayesian smoothing in the estimation of the pair potential function of Gibbs point processes. Bernoulli **5**(6): 1119-1136.

Holmes, M.J., and Reed, D.D. 1991. Competition indices for mixed species northern hardwoods. Forest Science **37**(5): 1338-1349.

Illian, J., Penttinen, A., Stoyan, H., and Stoyan, D. 2008. Statistical analysis and modelling of spatial point patterns. John Wiley & Sons, Chichester, England. 534 pp.

Jensen, E.B.V., and Nielsen, L.S. 2000. Inhomogeneous Markov point processes by transformation. Bernoulli **6**(5): 761-782.

Jensen, J.L. 1993. Asymptotic normality of estimates in spatial point processes. Scandinavian Journal of Statistics **20**: 97-109.

Jensen, J.L., and Künsch, H.R. 1994. On asymptotic normality of pseudo likelihood estimates for pairwise interaction processes. Annals of the Institute of Statistical Mathematics **46**(3): 475-486.

Kallenberg, O. 1984. An informal guide to the theory of conditioning in point processes. International Statistics Review **52**(2): 151-164.

Kelly, F.P., and Ripley, B.D. 1976. A note on Strauss's model for clustering. Biometrika **63**(2): 357-360.

Kendall, W.S. 1998. Perfect simulation for the area-interaction point process. *In* Probability towards 2000. *Edited by* L. Accardi and C. C. Heyde. Springer-Verlag, New York, NY. pp. 218-234.

Kendall, W.S., and Møller, J. 2000. Perfect simulation using dominating processes on ordered spaces, with application to locally stable point processes. Advances in Applied Probability **32**: 844-865.

Kendall, W.S., and Thönnes, E. 1999. Perfect simulation in stochastic geometry. Pattern Recognition **32**: 1569-1586.

Kenkel, N.C., Hendrie, M.L., and Bella, I.E. 1997. A long-term study of *Pinus banksiana* population dynamics. Journal of Vegetation Science **8**: 241-254.

Kittredge Jr., D.B. 1988. The influence of species composition on the growth of individual red oaks in mixed stands in southern New England. Can. J. For. Res. **18**: 1550-1555.

Kokkila, T., Mäkelä, A., and Nikinmaa, E. 2002. A method for generating stand structures using marked Gibbs point process. Silva Fennica **36**(1): 265-277.

Lamson, N.I. 1983. Precommercial thinning increases diameter growth of Appalachian hardwood stump sprouts. South. J. Appl. For. **7**(2): 93-97.

Lemmon, P.E., and Schumacher, F.X. 1962. Stocking density around ponderosa pine trees. Forest Science **8**(4): 397-402.

Li, F. 2004. Analysis on the spatial structures of spruce-fir stands in the Northwest, Maine, State University of New York College of Environmental Science and Forestry Syracuse, NY.

Lorimer, C.G. 1983. Tests of age-independent competition indices for individual trees in natural hardwood stands. Forest Ecology and Management **6**(4): 343-360.

Martin, G.L., Ek, A.R., and Monserud, R.A. 1977. Control of plot edge bias in forest stand growth simulation models. Can. J. For. Res. **7**: 100-105.

Mason, W.L., Connolly, T., Pommerening, A., and Edwards, C. 2007. Spatial structure of semi-natural and plantation stands of Scots pine (*Pinus sylvestris* L.) in northern Scotland. Forestry **80**(5): 564-583.

Matérn, B. 1960. Spatial variation. Meddelanden fran Statens Skogsforskningsinstitut **49**(5): 1-144.

Mateu, J., and Montes, F. 2001. Likelihood inference for Gibbs processes in the analysis of spatial point patterns. International Statistics Review **69**(1): 81-104.

Mateu, J., Uso, J.L., and Montes, F. 1998. The spatial pattern of a forest ecosystem. Ecological Modelling **108**(1-3): 163-174.

McDonald, R.I., Peet, R.K., and Urban, D.L. 2003. Spatial pattern of *Quercus* regeneration limitation and *Acer rubrum* invasion in a Piedmont forest. Journal of Vegetation Science **14**(3): 441-450.

Meng, X.L., and Wong, W.H. 1996. Simulating ratios of normalizing constants via a sample identity: a theoretical explanation. Statistica Sinica **6**: 831-860.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., and Teller, E. 1953. Equations of state calculations by fast computing machine. Journal of Chemical Physics **21**: 1087-1091.

Metsaranta, J.M., and Lieffers, V.J. 2008. A fifty-year reconstruction of annual changes in the spatial distribution of *Pinus banksiana* stands: does pattern fit competition theory. Plant Ecology **199**: 137-152.

Meyn, S., and Tweedie, R. 1993. Markov chains and stochastic stability. Springer-Verlag, New York, NY. pp.

Moeur, M. 1993. Characterizing spatial patterns of trees using stem-mapped data. Forest Science **39**(4): 756-775.

Møller, J., and Helisová, K. 2008. Power diagrams and interaction processes for unions of discs. Advances in Applied Probability **40**: 321-347.

Møller, J., Pettitt, A.N., Reeves, R., and Berthelsen, K.K. 2006. An efficient Markov chain Monte Carlo method for distributions with intractable normalising constants. Biometrika **93**(2): 451-458.

Møller, J., and Waagepetersen, R.P. 2003. An introduction to simulation-based inference for spatial point processes. *In* Spatial statistics and computational methods. *Edited by* J. Møller. Springer-Verlag, New York, NY. pp. 143-198.

Møller, J., and Waagepetersen, R.P. 2004. Statistical inference and simulation for spatial point processes. Chapman & Hall / CRC, Washington, DC. 300 pp.

Møller, J., and Waagepetersen, R.P. 2007. Modern statistics for spatial point processes. Scandinavian Journal of Statistics **34**: 643-684.

Monserud, R.A., and Ek, A.R. 1974. Plot edge bias in forest stand growth simulation models. Can. J. For. Res. **4**: 419-423.

Moyeed, R.A., and Baddeley, A. 1991. Stochastic approximation of the MLE for a spatial point pattern Scandinavian Journal of Statistics **18**: 39-50.

Neeff, T., Biging, G.S., Dutra, L.V., Freitas, C.C., and dos Santos, J.R. 2005. Markov point processes for modeling of spatial forest patterns in Amazonia derived from interferometric height. Remote Sensing of Environment **97**: 484-494.

Nguyen, X.X., and Zessin, H. 1979. Integral and differential characterizations of the Gibbs process. Mathematische Nachrichten **88**: 105-115.

Nielsen, L.S., and Jensen, E.B.V. 2004. Statistical inference for transformation inhomogeneous point processes. Scandinavian Journal of Statistics **31**(1): 131-142.

North, M., Chen, J., Oakley, B., Song, B., Rudnicki, M., Gray, A., and Innes, J. 2004. Forest stand structure and pattern of old-growth western hemlock/Douglas-fir mixed-conifer forests. Forest Science **50**(3): 299-311.

Ogata, Y., and Tanemura, M. 1981. Estimation of interaction potentials of spatial point patterns through the maximum likelihood procedure. Annals of the Institute of Statistical Mathematics **33**(B): 315-338.

Ogata, Y., and Tanemura, M. 1984. Likelihood analysis of spatial point patterns. Journal of the Royal Statistical Society. Series B **46**(3): 496-518.

Ohser, J., and Stoyan, D. 1981. On the second-order and orientation analysis of planar stationary point processes. Biometrical Journal **23**(6): 523-533.

Okabe, A., Boots, B., and Sugihara, K. 1992. Spatial tessellations: Concepts and applications of Voronoi diagrams. John Wiley & Sons, New York, NY. 532 pp.

Opie, J.E. 1968. Predictability of individual tree growth using various definitions of competing basal area. Forest Science **14**(3): 314-323.

Packard, K.C. 2000. Modelling tree diameter distributions for mixed-species conifer forests in the northeastern United States. MS Thesis, SUNY College of Environmental Science and Forestry, Syracuse, NY.

Packard, K.C., and Radtke, P.J. 2007. Forest sampling combining fixed- and variable-radius sample plots. Can. J. For. Res. **37**(8): 1460-1471.

Papangelou, F. 1974. The conditional intensity of general point processes and an application to line processes. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete **28**: 207-226.

Pélissier, R. 1998. Tree spatial patterns in three contrasting plots of a southern Indian tropical moist evergreen forest. Journal of Tropical Ecology **14**(1): 1-16.

Penttinen, A. 1984. Modelling interactions in spatial point patterns: parameter estimation by the maximum likelihood method. PhD diss., University of Jyväskylä, Jyväskylä, Finland.

Penttinen, A., Stoyan, D., and Henttonen, H.M. 1992. Marked point processes in forest statistics. Forest Science **38**(4): 806-824.

Picard, N., Bar-Hen, A., Mortier, F., and Chadoeuf, J. 2009. Understanding the dynamics of an undisturbed tropical rain forest from the spatial pattern of trees. Journal of Ecology **97**(1): 97-108.

Pielou, E.C. 1959. The use of point-to-plant distances in the study of the pattern of plant populations. Journal of Ecology **47**: 607-613.

Pommerening, A., and Stoyan, D. 2006. Edge-correction needs in estimating indices of spatial forest structure. Can. J. For. Res. **36**(7): 1723-1739.

Porté, A., and Bartelink, H.H. 2002. Modelling mixed forest growth: a review of models for forest management. Ecological Modelling **150**: 141-188.

Preston, C. 1977. Spatial birth-and-death processes. Bulletin of the International Statistical Institute **46**(2): 371-391.

Pretzsch, H. 1997. Analysis and modeling of spatial structures. Methodological considerations based on mixed beech-larch stands in Lower Saxony. Forest Ecology and Management **97**: 237-253.

Propp, J.G., and Wilson, D.B. 1996. Exact sampling with coupled Markov chains and applications to statistical mechanics. Random Structures and Algorithms **9**: 223-252.

Radtke, P.J., and Burkhart, H.E. 1998. A comparison of methods for edge-bias compensation. Can. J. For. Res. **28**(6): 942-945.

Rathbun, S.L., and Cressie, N.A.C. 1994. A space-time survival point process for a longleaf pine forest in southern Georgia. Journal of the American Statistical Association **89**(428): 1164-1174.

Reed, D.D., Liechty, H.O., and Burton, A.J. 1989. A simple procedure for mapping tree locations in forest stands. For Sci **35**(3): 657-662.

Rencher, A.C. 2000. Linear models in statistics. John Wiley & Sons, New York, NY. 578 pp.

Renshaw, E., and Särkkä, A. 2001. Gibbs point processes for studying the development of spatial-temporal stochastic processes. Computational Statistics and Data Analysis **36**: 81-105.

Ripley, B.D. 1976. The second-order analysis of stationary point processes. Journal of Applied Probability **13**: 255-266.

Ripley, B.D. 1977. Modelling spatial patterns. Journal of the Royal Statistical Society. Series B **39**: 172-192.

Ripley, B.D. 1988. Statistical inference for spatial processes. Cambridge Univ. Press, Cambridge, UK. 148 pp.

Robert, C.P., and Casella, G. 2004. Monte Carlo Statistical Methods. 2nd ed. Springer, New York, NY. 645 pp.

Roberts, G.O. 1996. Markov chain concepts related to sampling algorithms. *In* Markov chain Monte Carlo in practice. *Edited by* W. R. Gilks, S. Richardson, and D. J. Spiegelhalter. Chapman & Hall, New York, NY. pp. 45-57.

Ruelle, D. 1969. Statistical mechanics. W. A. Benjamin, Inc. , New York, NY. 219 pp.

SAMAB. 1996. The Southern Appalachian assessment summary report. USDA For. Serv. R8 Tech. Pub. 25. 118 pp.

Särkkä, A. 1993. Pseudo-likelihood approach for pair potential estimation of Gibbs processes. PhD diss, University of Jyväskylä, Jyväskylä, Finland.

Särkkä, A., and Renshaw, E. 2006. The analysis of marked point patterns evolving through time and space. Computational Statistics and Data Analysis **51**: 1698-1718.

Särkkä, A., and Tomppo, E. 1998. Modelling interactions between trees by means of field observations. Forest Ecology and Management **108**(1-2): 57.

Saunders, R., and Funk, G.M. 1977. Poisson limits for a clustering model of Strauss. Journal of Applied Probability **14**: 776-784.

Schabenberger, O., and Gotway, C.A. 2005. Statistical methods for spatial data analysis. Chapman & Hall / CRC, Washington, DC. 488 pp.

Scott, D.W. 1992. Multivariate density estimation: theory, practice, and visualization. John Wiley & Sons, New York, NY. 317 pp.

Shi, H., and Zhang, L. 2003. Local analysis of tree competition and growth. Forest Science **49**(6): 938-955.

Smith, B.J. 2005. Bayesian output analysis program (BOA) version 1.1 user's manual [online]. Available from http://www.public-health.uiowa.edu/boa].

Sterner, R.W., Ribic, C.A., and Schatz, G.E. 1986. Testing for life historical changes in spatial patterns of four tropical tree species. . Journal of Ecology **74**: 621-633.

Stoyan, D. 2000. Basic ideas of spatial statistics. *In* Statistical physics and spatial statistics: the art of analyzing and modeling spatial structures and pattern formation. *Edited by* K. Mecke and D. Stoyan. Springer-Verlag, Berlin. pp. 3-21.

Stoyan, D. 2006. On estimators of the nearest neighbour distance distribution function for stationary point processes. Metrika **64**(2): 139-150.

Stoyan, D., Bertram, U., and Wendrock, H. 1993. Estimation variances for estimators of product densities and pair correlation functions of planar point processes. Annals of the Institute of Statistical Mathematics **45**(2): 211-221.

Stoyan, D., and Grabarnik, P. 1991. Statistics for the stationary Strauss model by the cusp point method. Statistics **22**: 283-289.

Stoyan, D., Kendall, W.S., and Mecke, J. 1995. Stochastic geometry and its applications. 2nd ed. John Wiley & Sons, New York, NY. 436 pp.

Stoyan, D., and Penttinen, A. 2000. Recent applications of point process methods in forestry statistics. Statistical Science **15**(1): 61-78.

Stoyan, D., and Stoyan, H. 1994. Fractals, random shapes, and point fields. John Wiley & Sons, New York, NY. 389 pp.

Stoyan, D., and Stoyan, H. 1996. Estimating pair correlation functions of planar cluster processes. Biometrical Journal **38**(3): 259-271.

Stoyan, D., and Stoyan, H. 1998. Non-homogeneous Gibbs process models for forestry - a case study. Biometrical Journal **40**(5): 521-531.

Strauss, D.J. 1975. A model for clustering. Biometrika **63**: 467-475.

Stroock, D.W. 1994. A concise introduction to the theory of integration. 2nd ed. Birkhäuser, Boston, MA. 184 pp.

Takacs, R. 1983. Estimator for the pair-potential of a Gibbsian point process. Institut für Mathematik, Johannes Kepler Universität Linz, Austria. Institutsbericht 238. pp.

Takacs, R. 1986. Estimator for the pair-potential of a Gibbsian point process. Statistics **17**(3): 429-433.

Tierney, L. 1994. Markov chains for exploring posterior distributions. The Annals of Statistics **22**(4): 1701-1762.

Tomppo, E. 1986. Models and methods for analysing spatial patterns of trees. Communicationes Instituti Forestalis Fenniae **138**: 65.

Van Lieshout, M.N.M. 2000. Markov point processes and their applications. Imperial College Press, London, UK. 175 pp.

Van Lieshout, M.N.M., and Stoica, R.S. 2006. Perfect simulation for marked point processes. Computational Statistics and Data Analysis **51**(2): 679-698.

Ward, J.S., Parker, G.R., and Ferrandino, F.J. 1996. Long-term spatial dynamics in an old-growth deciduous forest. Forest Ecology and Management **83**: 189-202.

Warren, W.G. 1972. Point processes in forestry. *In* Stochastic point processes: Statistical analysis, theory, and applications. *Edited by* P. A. Lewis. John Wiley & Sons, New York, NY. pp. 801-816.

Weisstein, E.W. 2007. Serpentine curve [online]. Available from http://mathworld.wolfram.com/SerpentineCurve.html [cited Aug. 30, 2007].

Wendel, G.W. 1975. Stump sprout growth and quality of several Appalachian hardwood species after clearcutting. USDA For. Serv. Res. Pap. NE-329.

Wendel, G.W., and Trimble, G.R. 1968. Reproduction after seed-tree harvest cuttings in Appalachian hardwoods. USDA For. Serv. Res. Pap. NE-99. 16 pp.

Williams, M.S., Williams, M.T., and Mowrer, H.T. 2001. A boundary reconstruction method for circular fixed-area plots in environmental survey. Journal of Agricultural, Biological, and Environmental Statistics **6**(4): 479-494.

Zhao, D. 2003. Modeling mixed species forest stands. Ph.D. diss., University of Georgia, Athens, GA.

Zhao, D., Borders, B., Wilson, M., and Rathbun, S.L. 2006. Modeling neighborhood effects on the growth and survival of individual trees in a natural temperate species-rich forest. Ecological Modelling **196**: 90-102.

Appendix A

The following figures present *L*-function plots of Coweeta data where the data is screened by tree diameter. As the data are screened to larger DBH limits (from 10cm to 30cm by 4cm increments), the number of trees on the Coweeta plots (*n*) which meet these limits decreases (Table A1). Consequently, the 95% Monte Carlo bands of CSR will increase, as decreasing *n* will allow samples from CSR to show more variable spatial structure.

Table A1. Number of trees by DBH limit on Coweeta gradient plots.

DBH Limit (cm)	*Plot 118*	*Plot 218*	*Plot 318*	*Plot 427*	*Plot 527*
10	353	264	349	384	268
14	239	207	282	294	207
18	171	169	227	204	162
22	116	137	173	146	137
26	80	114	128	111	110
30	59	90	101	88	89

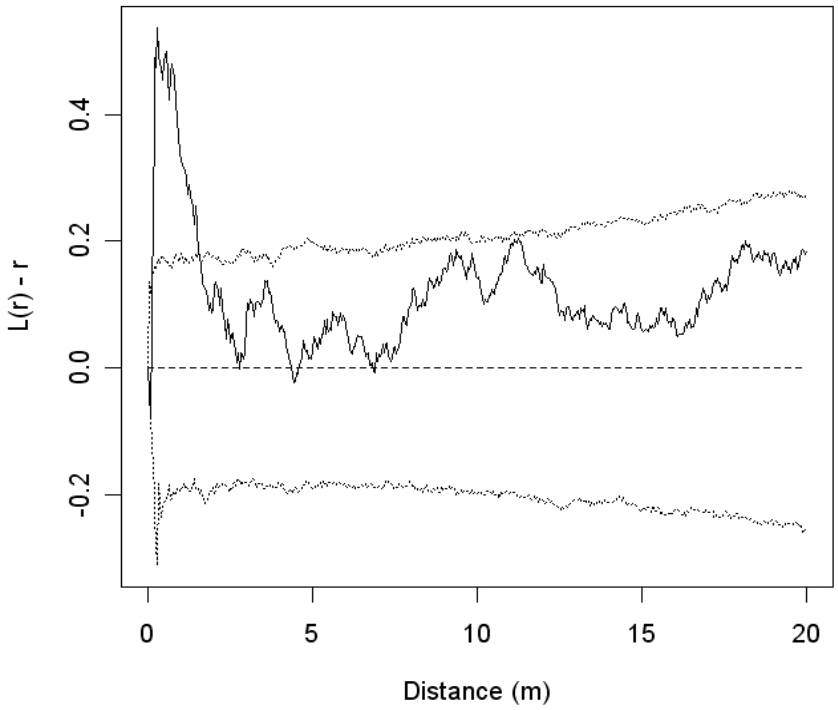

Figure A1. Observed $L(r) - r$ (solid line) and 95% Monte Carlo simulation envelope for CSR (dashed lines) for Plot 118, DBH limit = 10cm. Note: This figure reproduces Fig. 18.

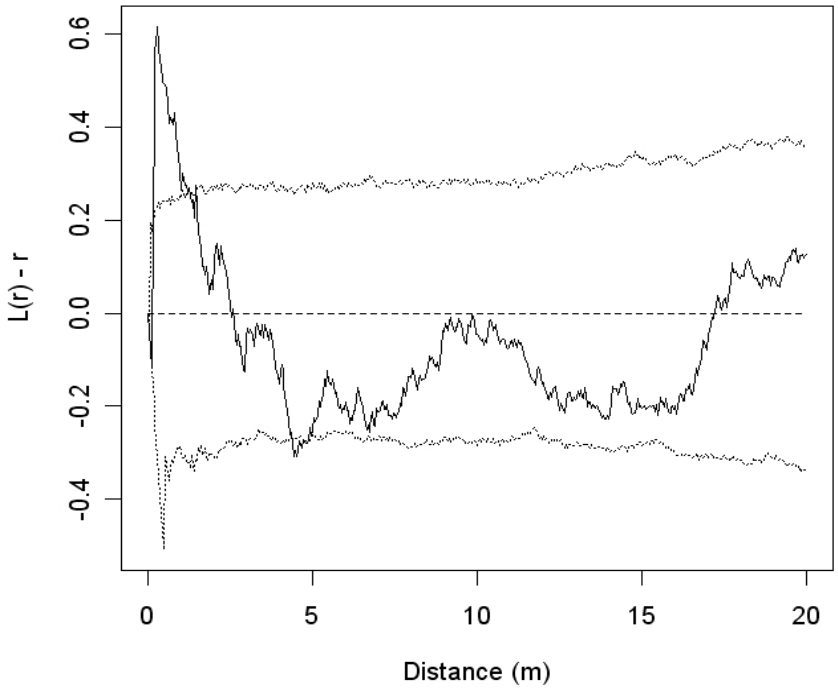

Figure A2. Observed $L(r) - r$ (solid line) and 95% Monte Carlo simulation envelope for CSR (dashed lines) for Plot 118, DBH limit = 14cm.

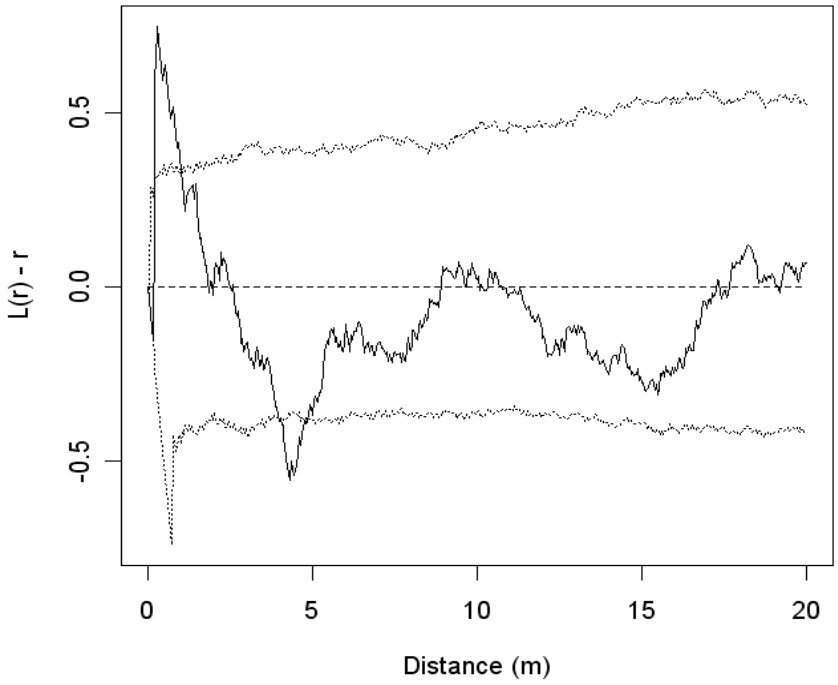

Figure A3. Observed $L(r) - r$ (solid line) and 95% Monte Carlo simulation envelope for CSR (dashed lines) for Plot 118, DBH limit = 18cm.

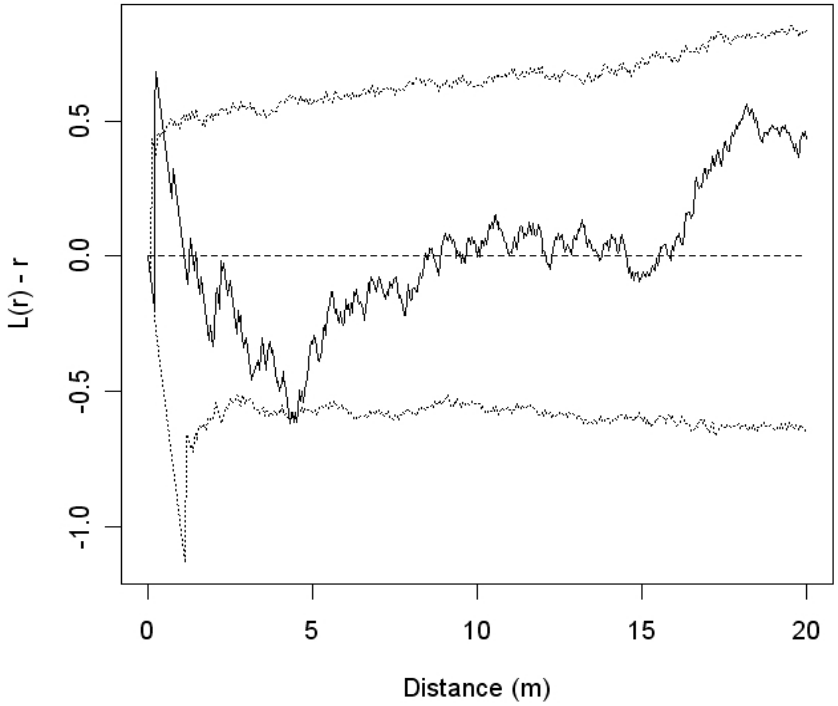

Figure A4. Observed $L(r) - r$ (solid line) and 95% Monte Carlo simulation envelope for CSR (dashed lines) for Plot 118, DBH limit = 22cm.

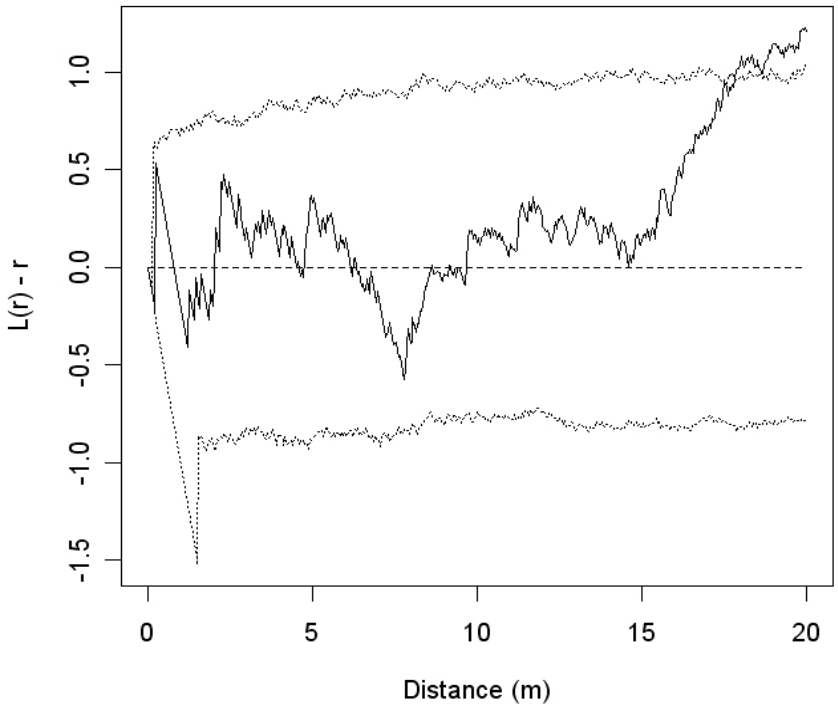

Figure A5. Observed $L(r)$ - r (solid line) and 95% Monte Carlo simulation envelope for CSR (dashed lines) for Plot 118, DBH limit = 26cm.

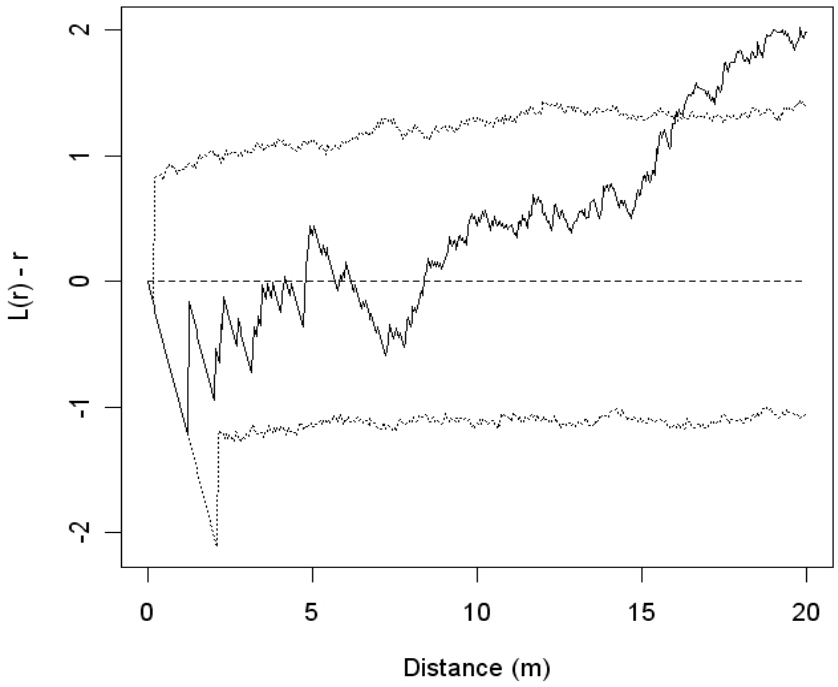

Figure A6. Observed $L(r)$ - r (solid line) and 95% Monte Carlo simulation envelope for CSR (dashed lines) for Plot 118, DBH limit = 30cm.

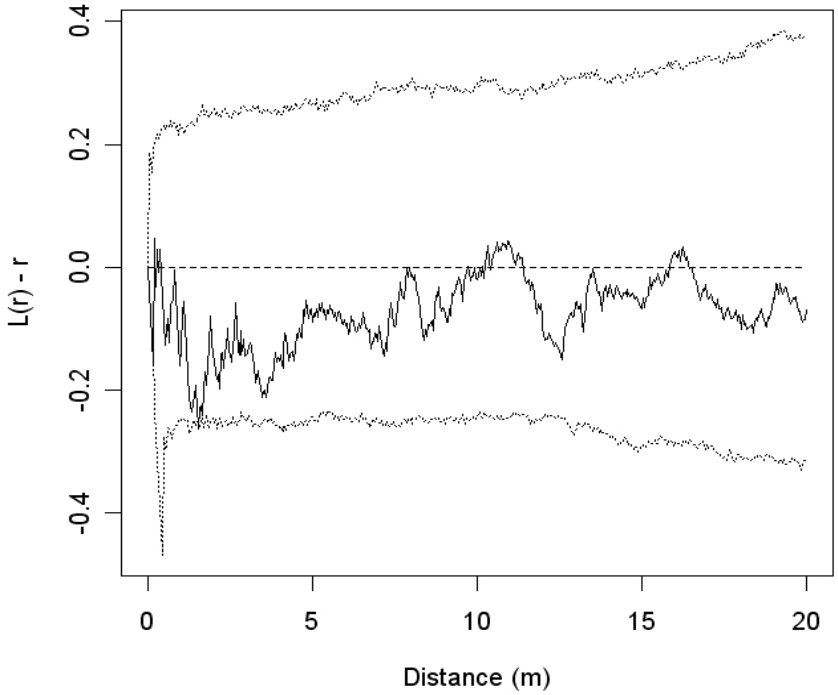

Figure A7. Observed $L(r)$ - r (solid line) and 95% Monte Carlo simulation envelope for CSR (dashed lines) for Plot 218, DBH limit = 10cm. Note: This figure reproduces Fig. 19.

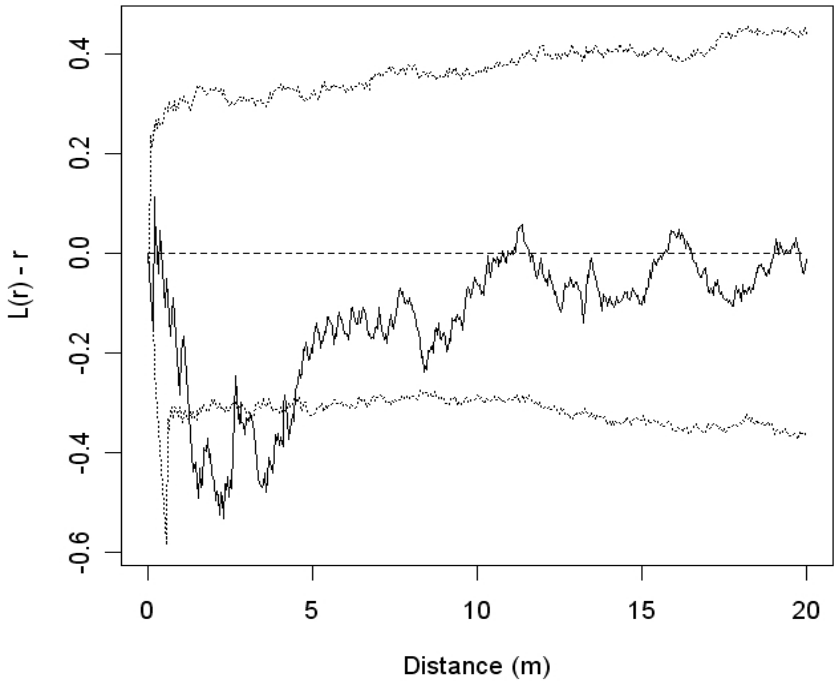

Figure A8. Observed $L(r)$ - r (solid line) and 95% Monte Carlo simulation envelope for CSR (dashed lines) for Plot 218, DBH limit = 14cm.

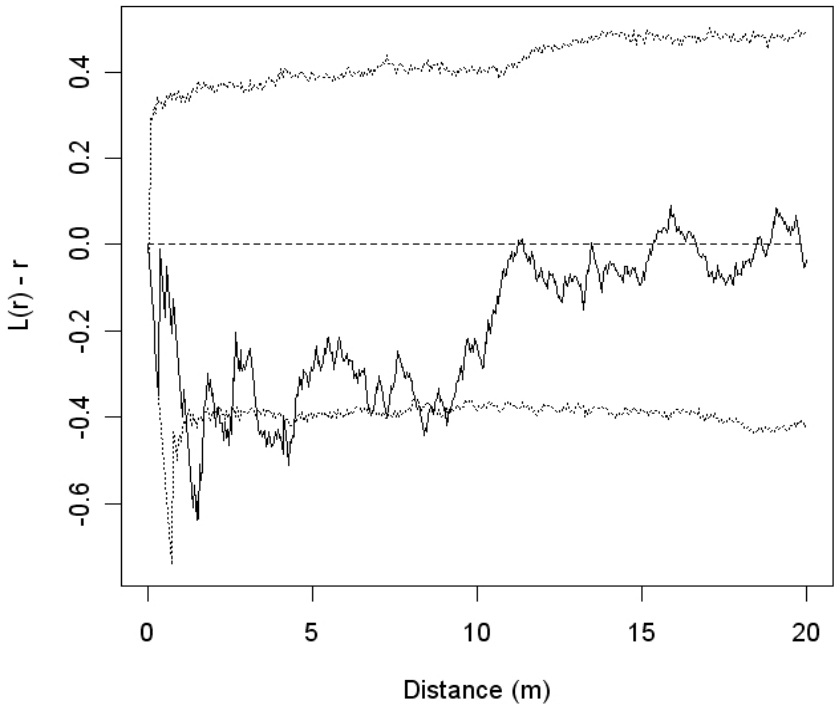

Figure A9. Observed $L(r) - r$ (solid line) and 95% Monte Carlo simulation envelope for CSR (dashed lines) for Plot 218, DBH limit = 18cm.

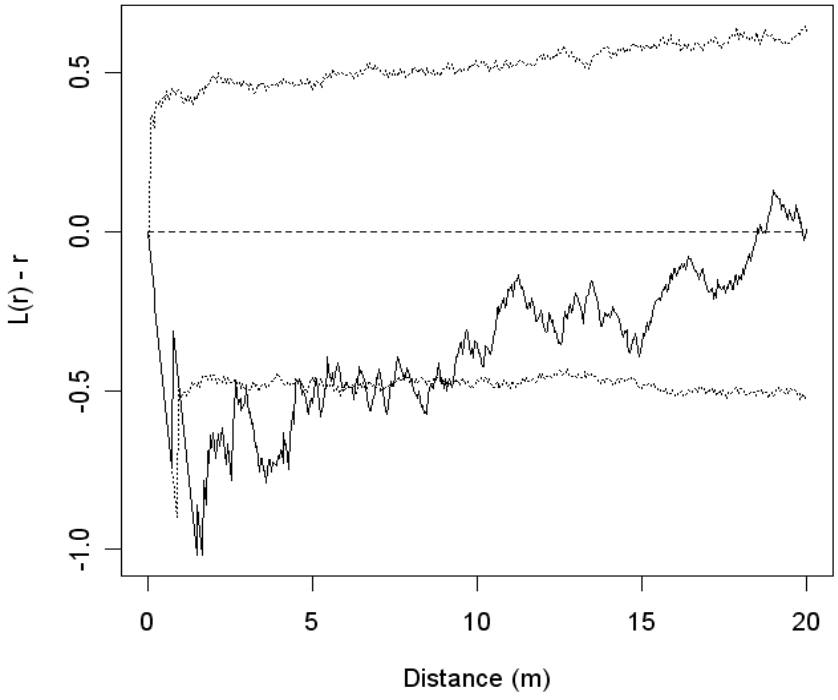

Figure A10. Observed $L(r) - r$ (solid line) and 95% Monte Carlo simulation envelope for CSR (dashed lines) for Plot 218, DBH limit = 22cm.

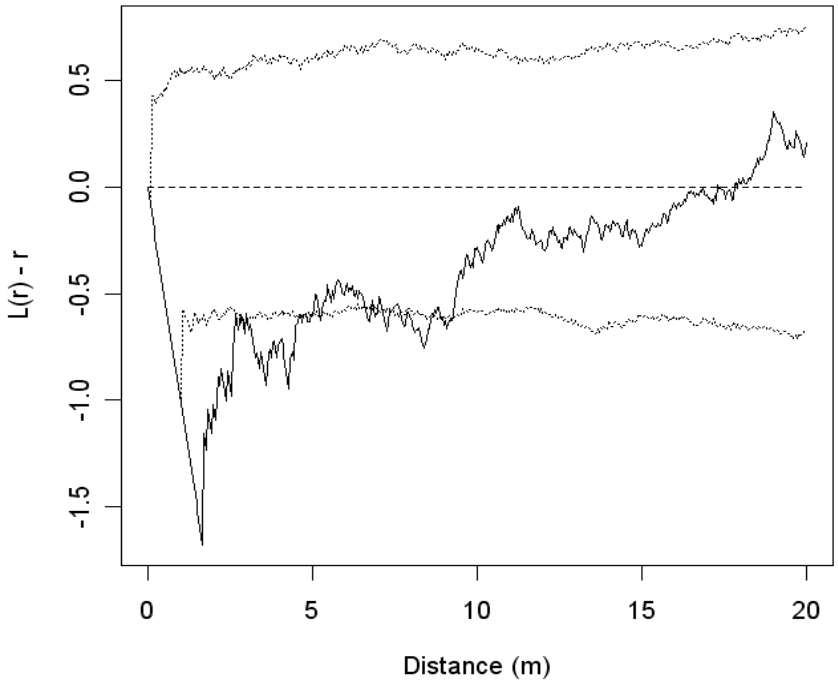

Figure A11. Observed $L(r) - r$ (solid line) and 95% Monte Carlo simulation envelope for CSR (dashed lines) for Plot 218, DBH limit = 26cm.

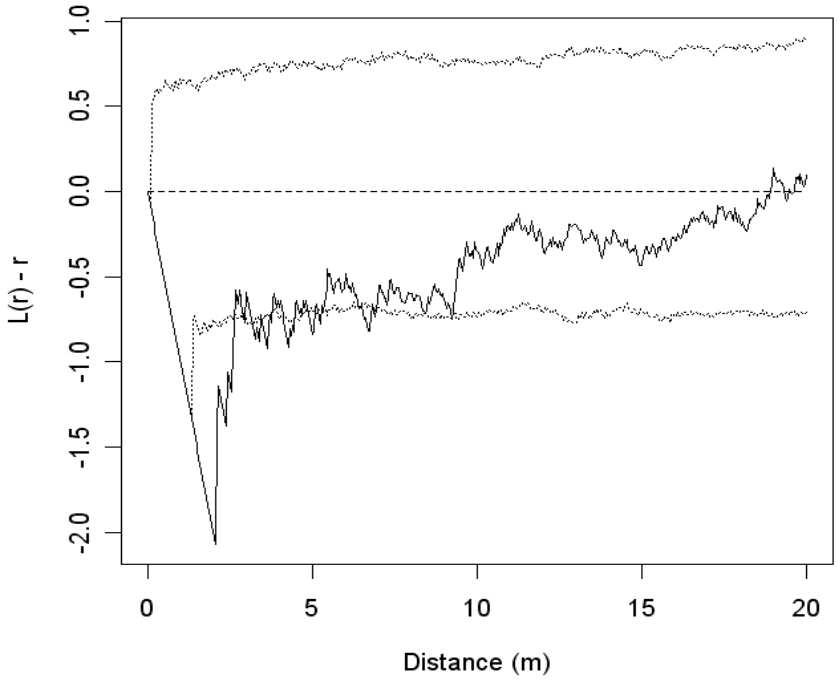

Figure A12. Observed $L(r) - r$ (solid line) and 95% Monte Carlo simulation envelope for CSR (dashed lines) for Plot 218, DBH limit = 30cm.

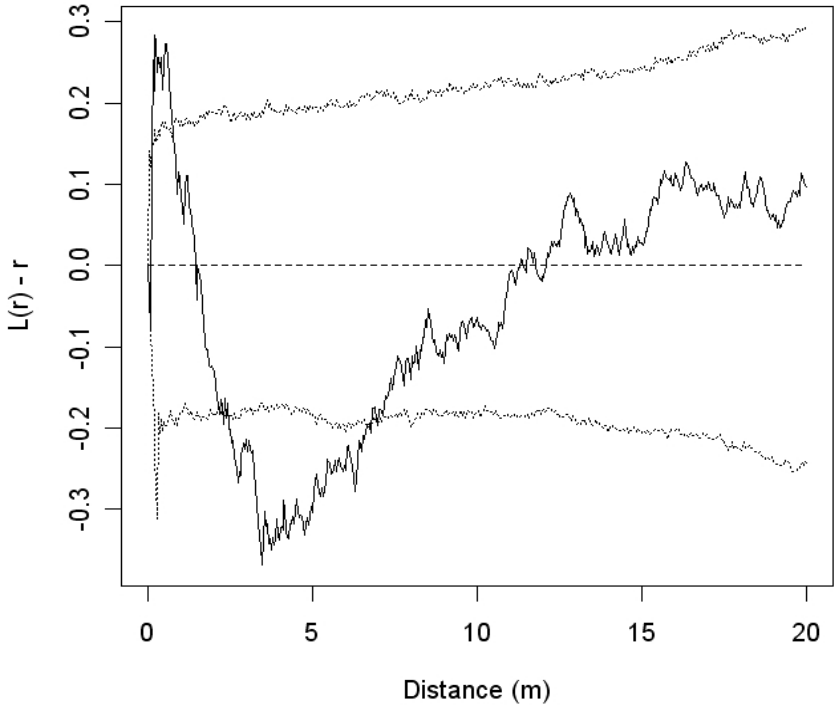

Figure A13. Observed $L(r) - r$ (solid line) and 95% Monte Carlo simulation envelope for CSR (dashed lines) for Plot 318, DBH limit = 10cm. Note: This figure reproduces Fig. 20.

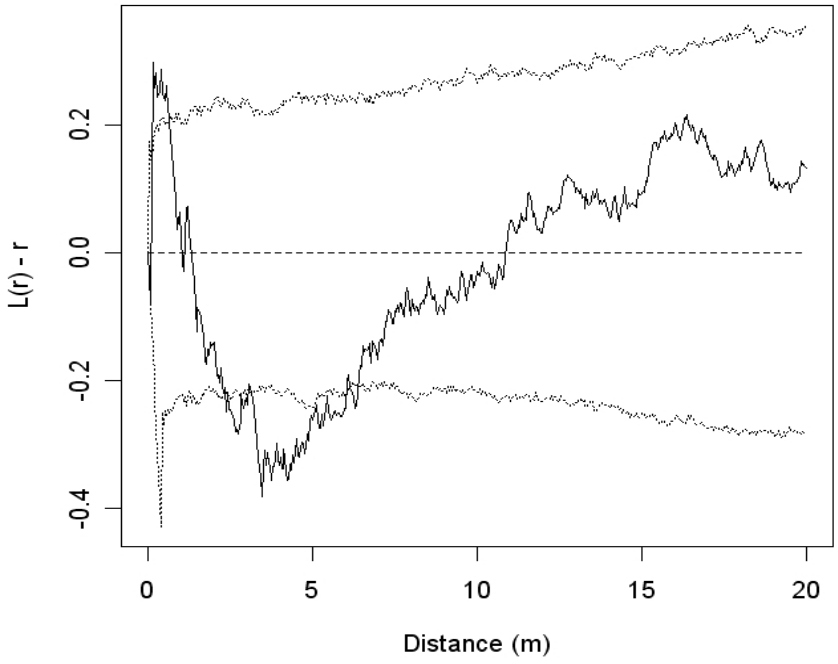

Figure A14. Observed $L(r) - r$ (solid line) and 95% Monte Carlo simulation envelope for CSR (dashed lines) for Plot 318, DBH limit = 14cm.

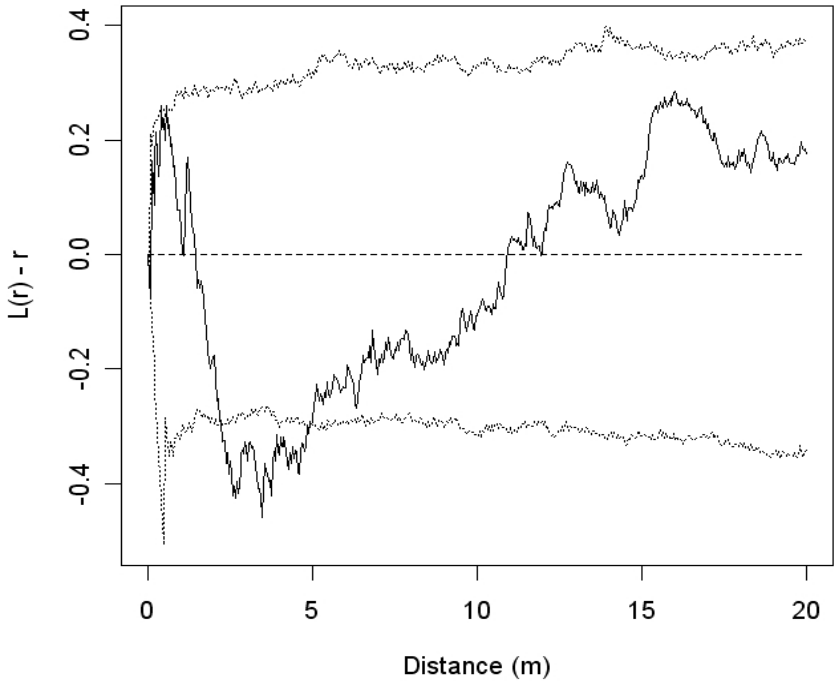

Figure A15. Observed $L(r) - r$ (solid line) and 95% Monte Carlo simulation envelope for CSR (dashed lines) for Plot 318, DBH limit = 18cm.

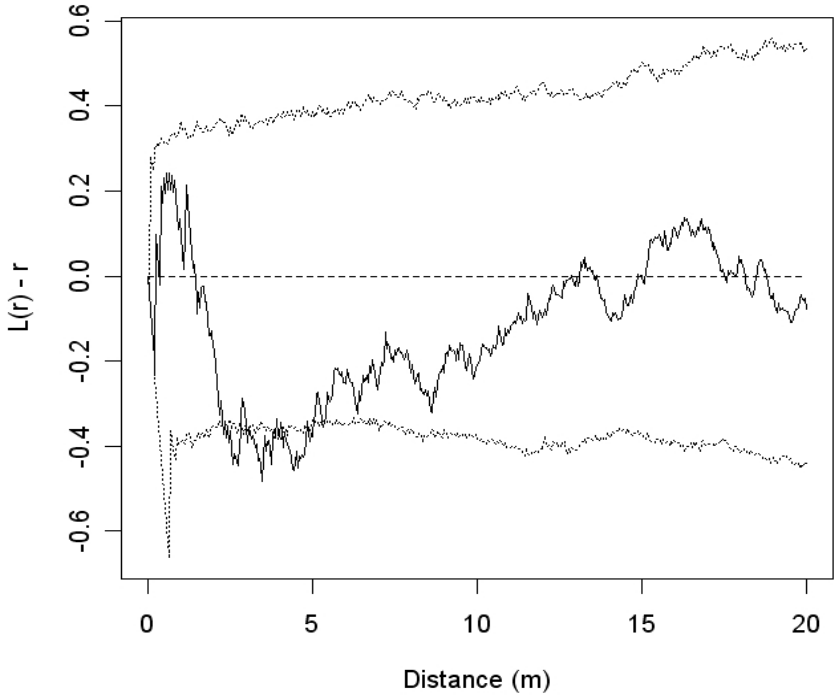

Figure A16. Observed $L(r) - r$ (solid line) and 95% Monte Carlo simulation envelope for CSR (dashed lines) for Plot 318, DBH limit = 22cm.

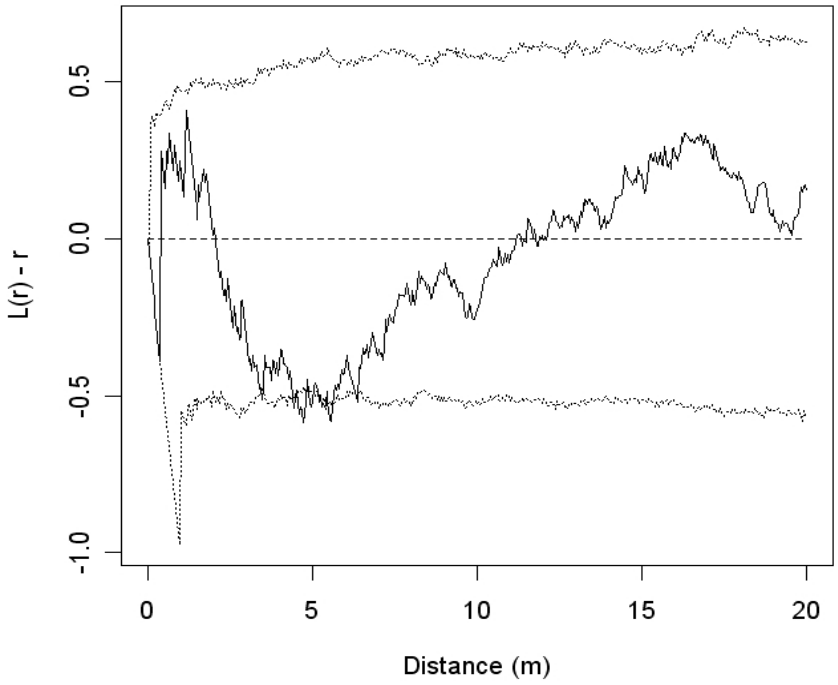

Figure A17. Observed $L(r) - r$ (solid line) and 95% Monte Carlo simulation envelope for CSR (dashed lines) for Plot 318, DBH limit = 26cm.

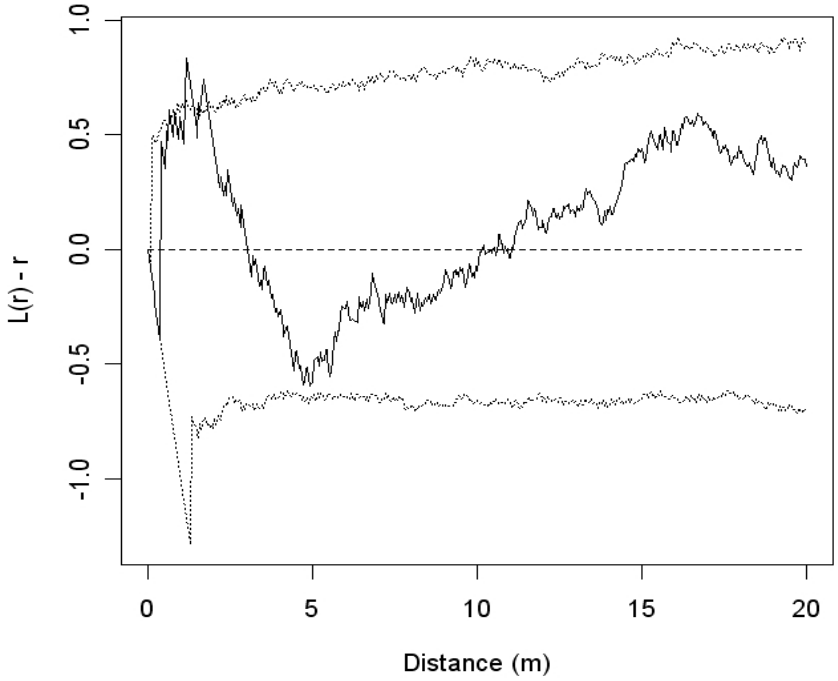

Figure A18. Observed $L(r) - r$ (solid line) and 95% Monte Carlo simulation envelope for CSR (dashed lines) for Plot 318, DBH limit = 30cm.

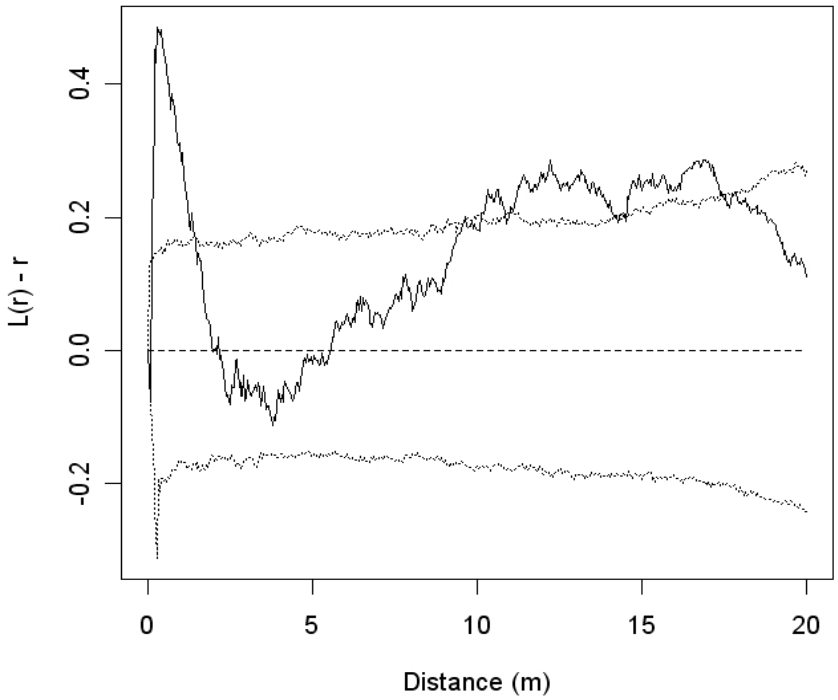

Figure A19. Observed $L(r) - r$ (solid line) and 95% Monte Carlo simulation envelope for CSR (dashed lines) for Plot 427, DBH limit = 10cm. Note: This figure reproduces Fig. 21.

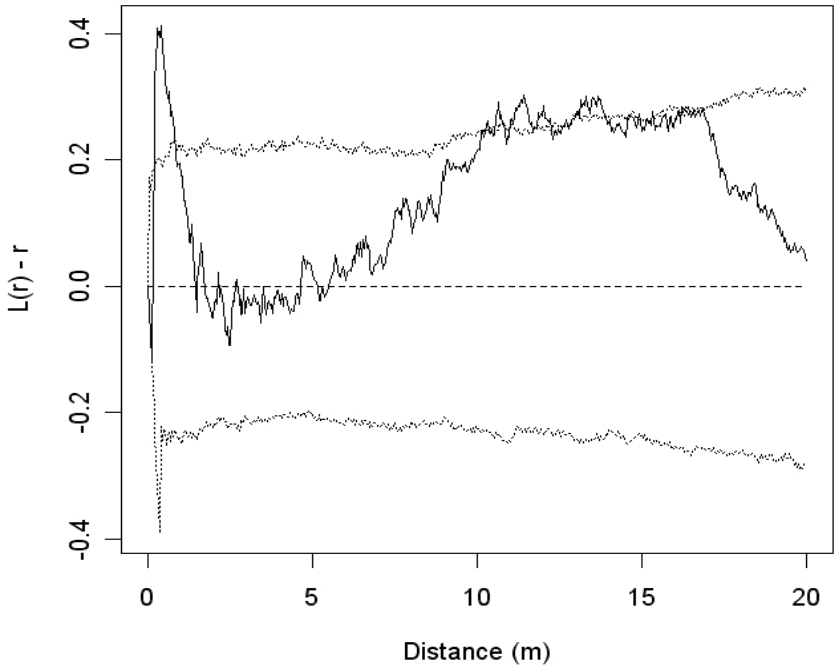

Figure A20. Observed $L(r) - r$ (solid line) and 95% Monte Carlo simulation envelope for CSR (dashed lines) for Plot 427, DBH limit = 14cm.

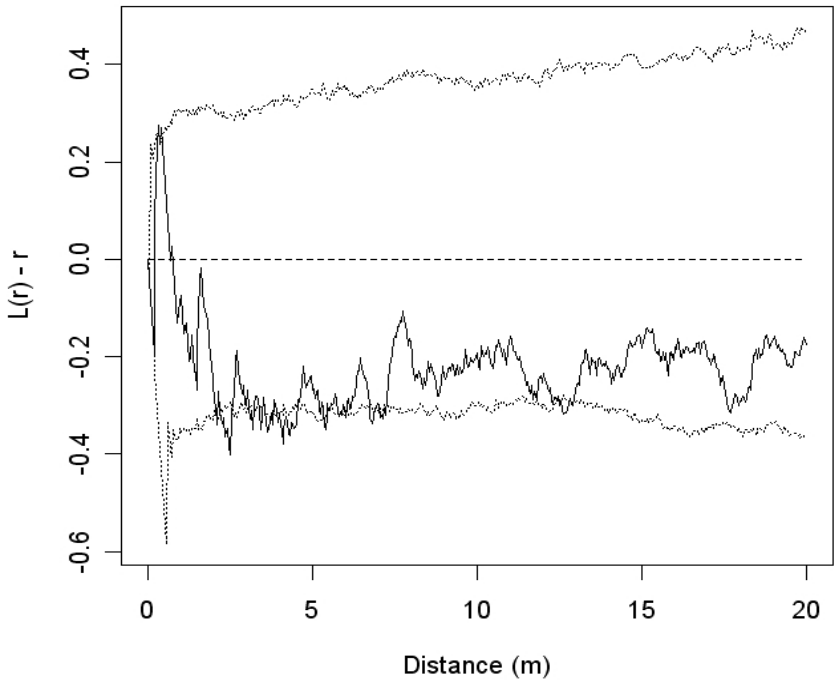

Figure A21. Observed $L(r) - r$ (solid line) and 95% Monte Carlo simulation envelope for CSR (dashed lines) for Plot 427, DBH limit = 18cm.

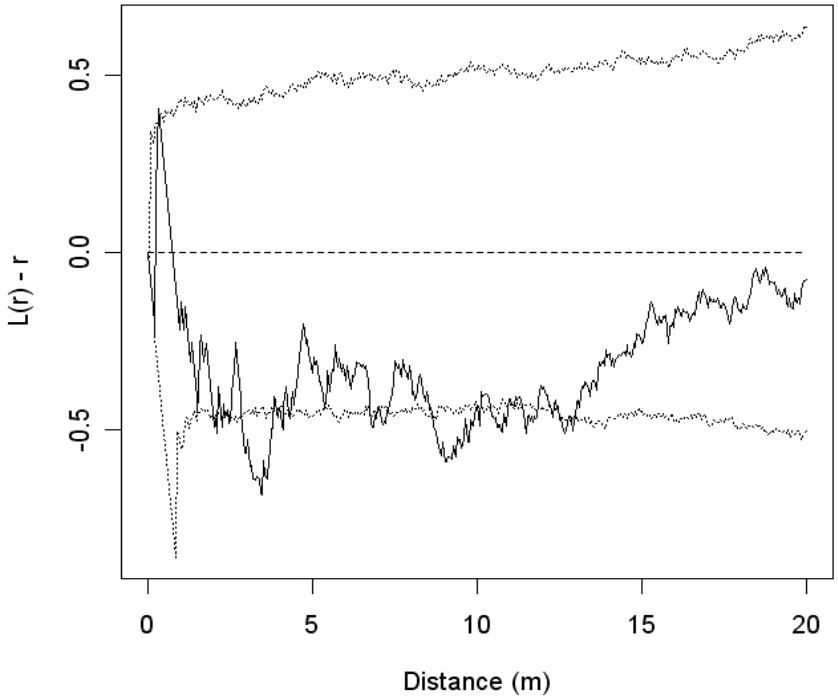

Figure A22. Observed $L(r) - r$ (solid line) and 95% Monte Carlo simulation envelope for CSR (dashed lines) for Plot 427, DBH limit = 22cm.

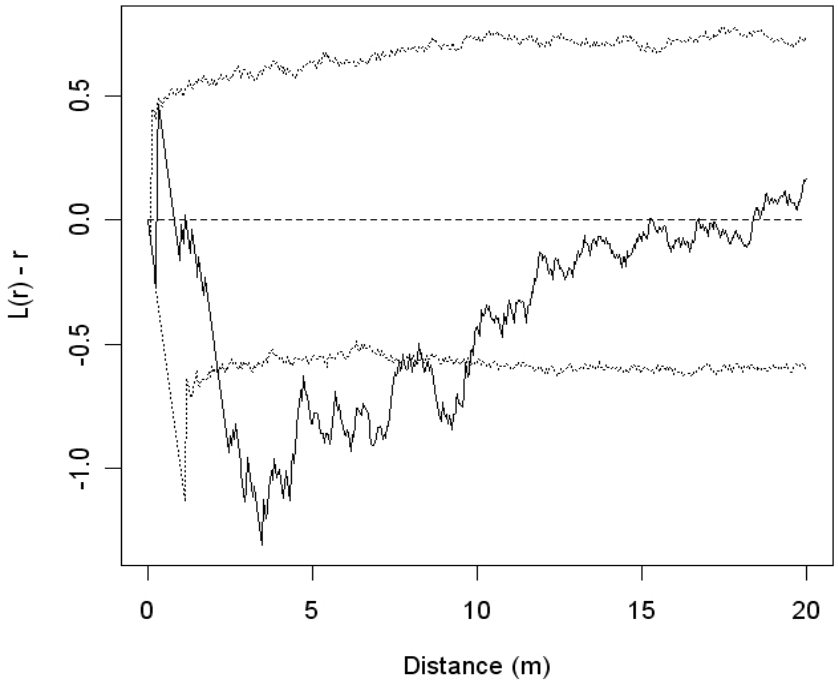

Figure A23. Observed $L(r) - r$ (solid line) and 95% Monte Carlo simulation envelope for CSR (dashed lines) for Plot 427, DBH limit = 26cm.

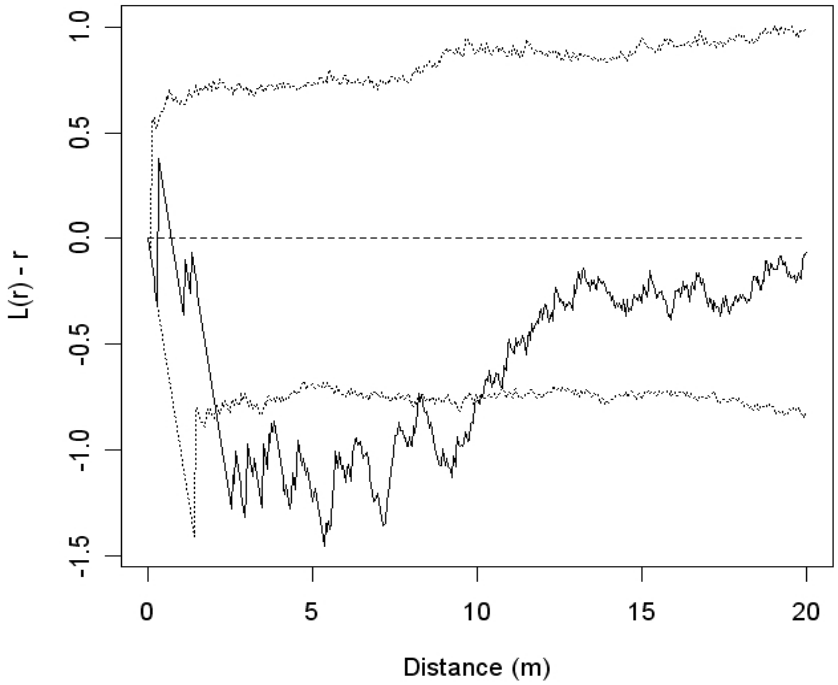

Figure A24. Observed $L(r) - r$ (solid line) and 95% Monte Carlo simulation envelope for CSR (dashed lines) for Plot 427, DBH limit = 30cm.

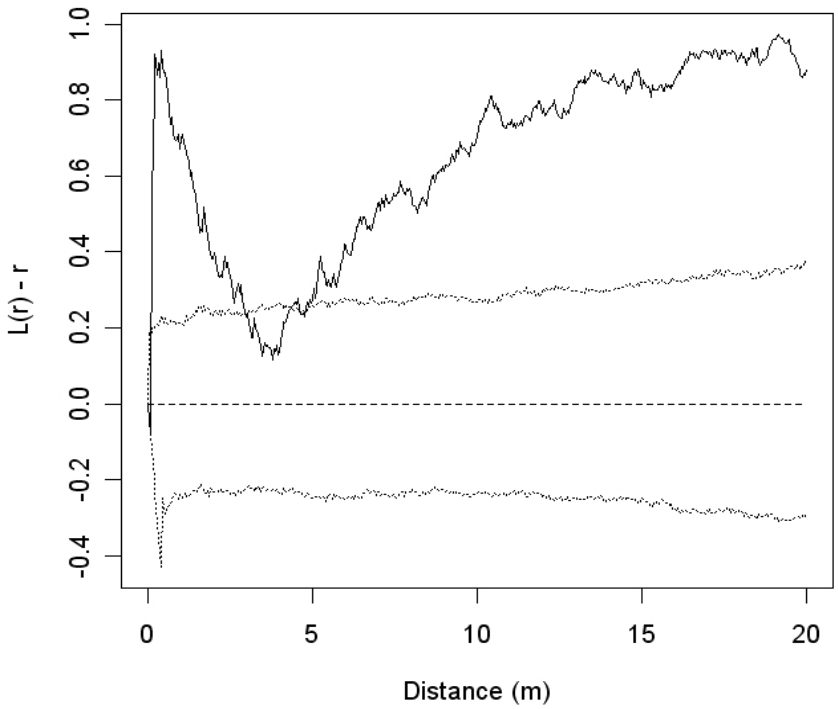

Figure A25. Observed $L(r) - r$ (solid line) and 95% Monte Carlo simulation envelope for CSR (dashed lines) for Plot 527, DBH limit = 10cm. Note: This figure reproduces Fig. 22.

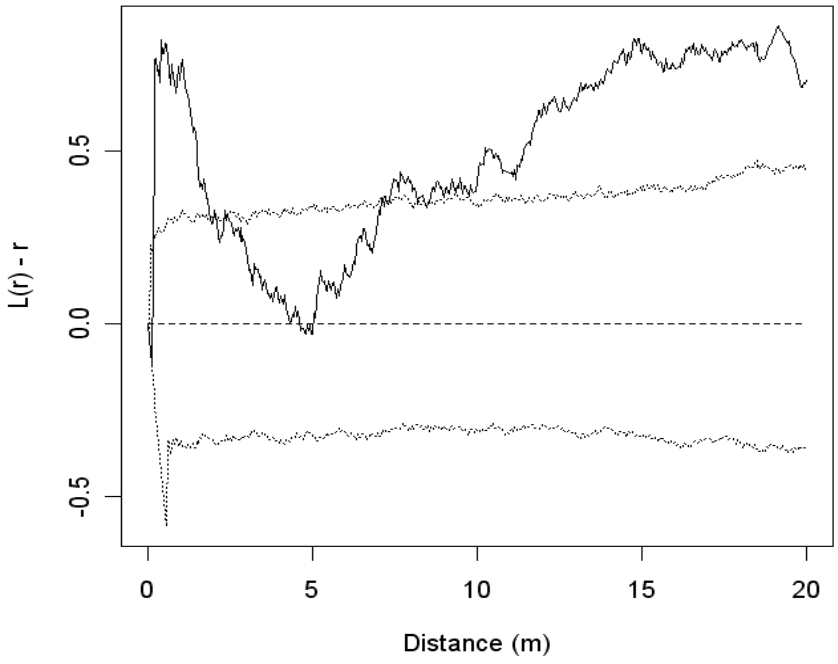

Figure A26. Observed $L(r) - r$ (solid line) and 95% Monte Carlo simulation envelope for CSR (dashed lines) for Plot 527, DBH limit = 14cm.

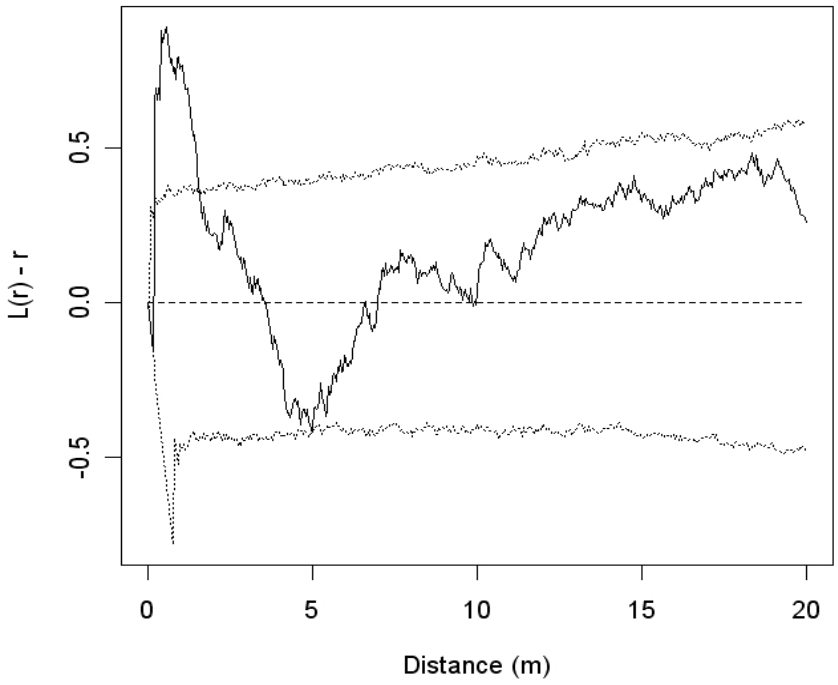

Figure A27. Observed $L(r) - r$ (solid line) and 95% Monte Carlo simulation envelope for CSR (dashed lines) for Plot 527, DBH limit = 18cm.

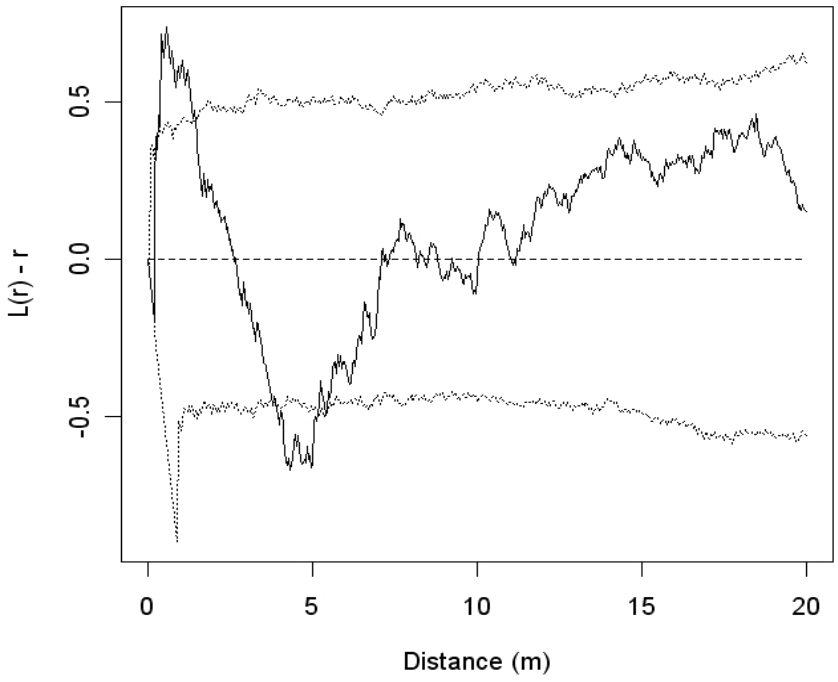

Figure A28. Observed $L(r) - r$ (solid line) and 95% Monte Carlo simulation envelope for CSR (dashed lines) for Plot 527, DBH limit = 22cm.

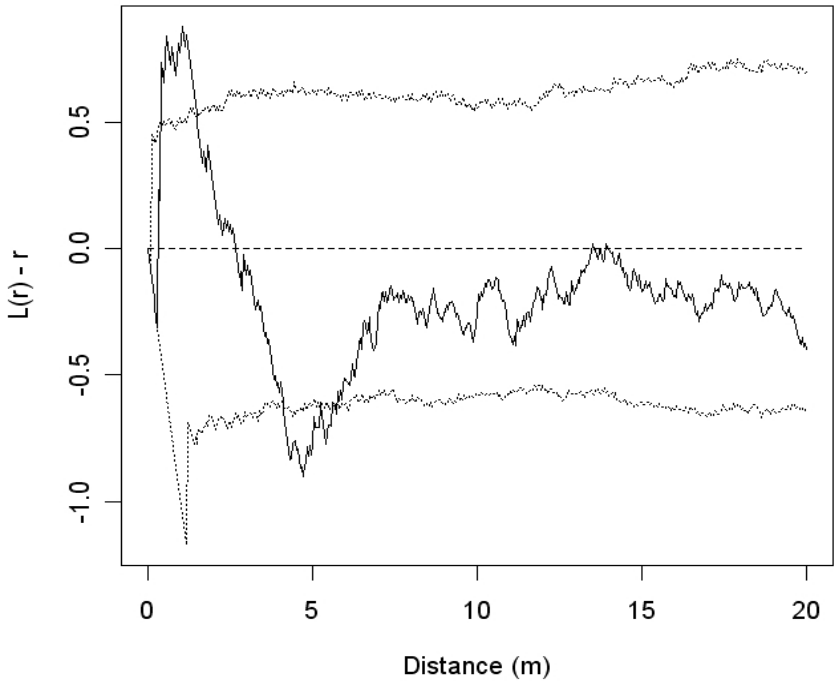

Figure A29. Observed $L(r) - r$ (solid line) and 95% Monte Carlo simulation envelope for CSR (dashed lines) for Plot 527, DBH limit = 26cm.

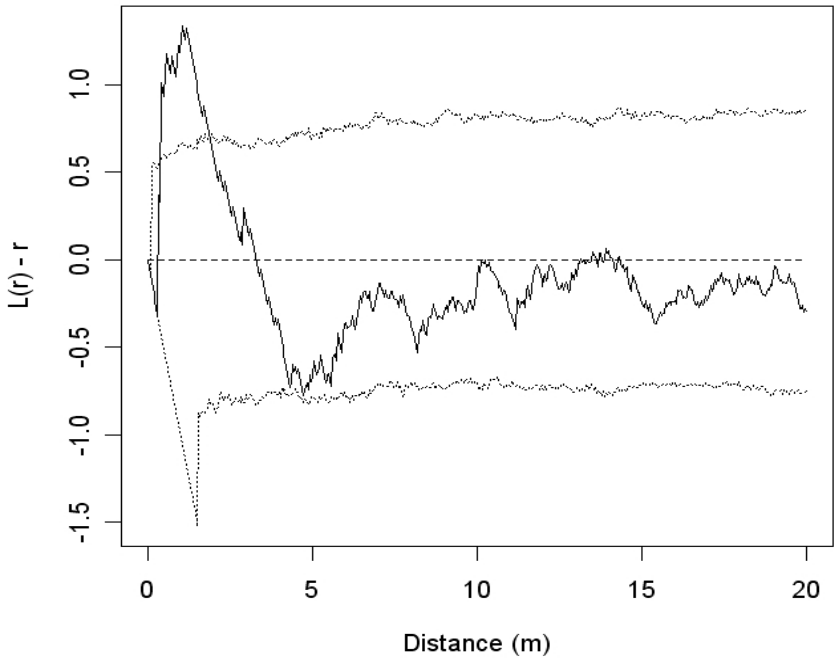

Figure A30. Observed $L(r) - r$ (solid line) and 95% Monte Carlo simulation envelope for CSR (dashed lines) for Plot 527, DBH limit = 30cm.

Appendix B

The following figures present trace plots of the "pseudo-canonical" statistics from the simulation of the Gibbs pairwise interaction models (F, HCME, and HCSP). For reference, the pseudo-canonical statistics for each model is provided (Table B1).

Table B1. Canonical parameters and "pseudo-canonical" statistics for the proposed Gibbs point process models with 2nd-order interaction.

Model	*Canonical Parameters*	*Pseudo-canonical statistic*
Fiksel	$\log(\beta)$	$n(\mathbf{x})$
	θ_1	$PCS_1(\mathbf{x}) = \sum_{i<j} -e_{ij} \exp\left(-\theta_2 \left\|x_i - x_j\right\|\right) I_{(r_0,R]}\left(\left\|x_i - x_j\right\|\right)$
Hard-Core Modified Exponential	$\log(\beta)$	$n(\mathbf{x})$
	θ_1	$PCS_1(\mathbf{x}) = \sum_{i<j} e_{ij} \left(\frac{-1}{\left\|x_i - x_j\right\|^2 + 1}\right) I_{(r_0,R]}\left(\left\|x_i - x_j\right\|\right)$
	θ_2	$PCS_2(\mathbf{x}) = \sum_{i<j} -e_{ij} \exp\left(-\theta_3 \left\|x_i - x_j\right\|\right) I_{(r_0,R]}\left(\left\|x_i - x_j\right\|\right)$
Hard-Core Serpentine	$\log(\beta)$	$n(\mathbf{x})$
	θ_1	$PCS_1(\mathbf{x}) = \sum_{i<j} -e_{ij} \left(\frac{\left\|x_i - x_j\right\| + \theta_2}{\left\|x_i - x_j\right\|^3 + 1}\right) I_{(r_0,R]}\left(\left\|x_i - x_j\right\|\right)$

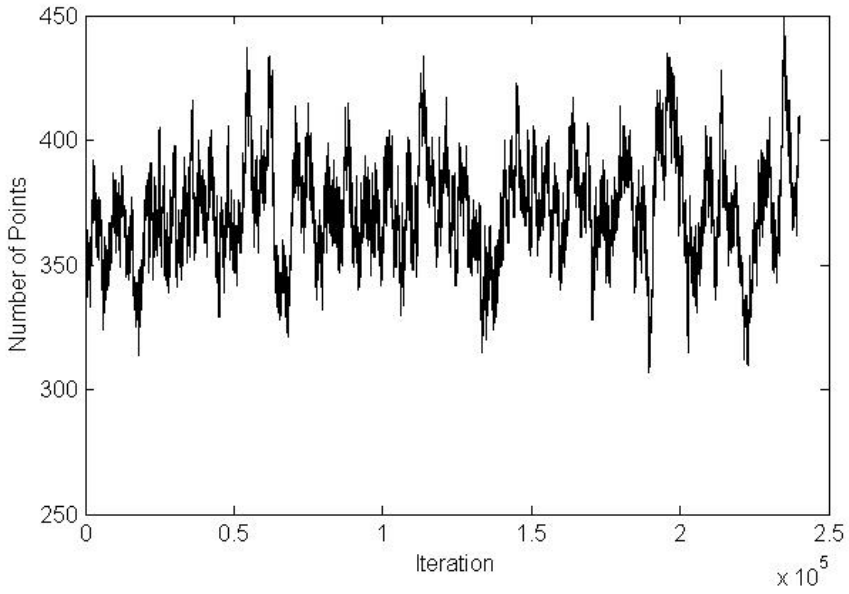

Figure B1. Trace plot of the number of points for Ex. 1 of the Fiksel model with $(\beta = 0.0545,\ \theta_1 = -1,\ \theta_2 = 2.5,\ R = 5)$.

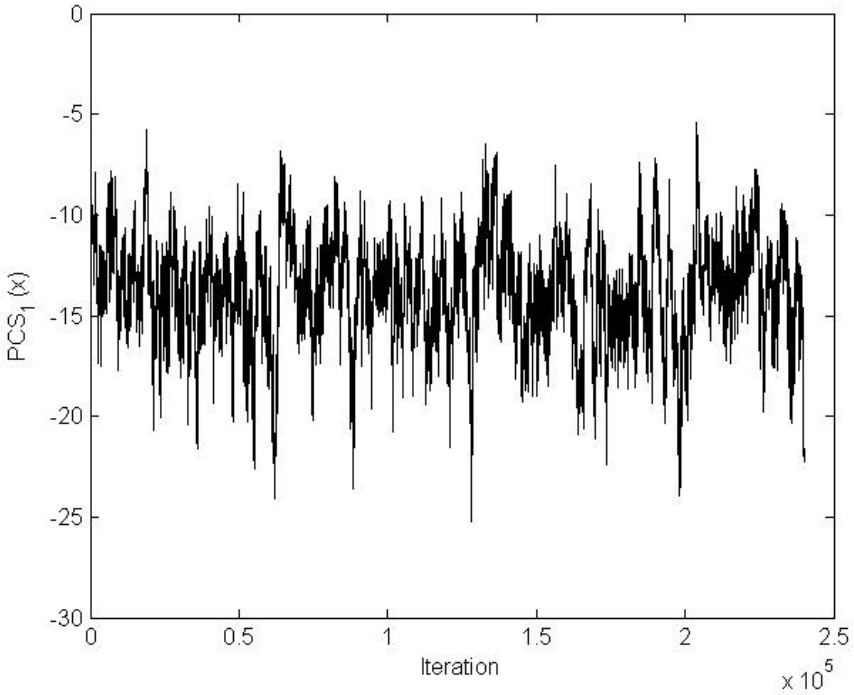

Figure B2. Trace plot of the pseudo-canonical statistic PCS_1 (**x**) for Ex. 1 of the Fiksel model with $(\beta = 0.0545,\ \theta_1 = -1,\ \theta_2 = 2.5,\ R = 5)$.

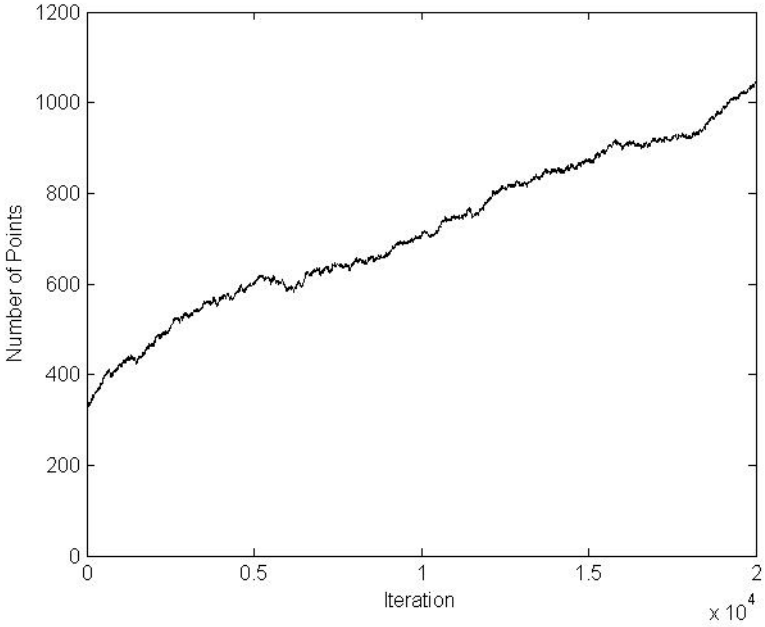

Figure B3. Trace plot of the number of points for Ex. 2 of the Fiksel model with $\left(\beta = 0.0545,\ \theta_1 = -1,\ \theta_2 = 1,\ R = 5\right)$ after 20,000 iterations.

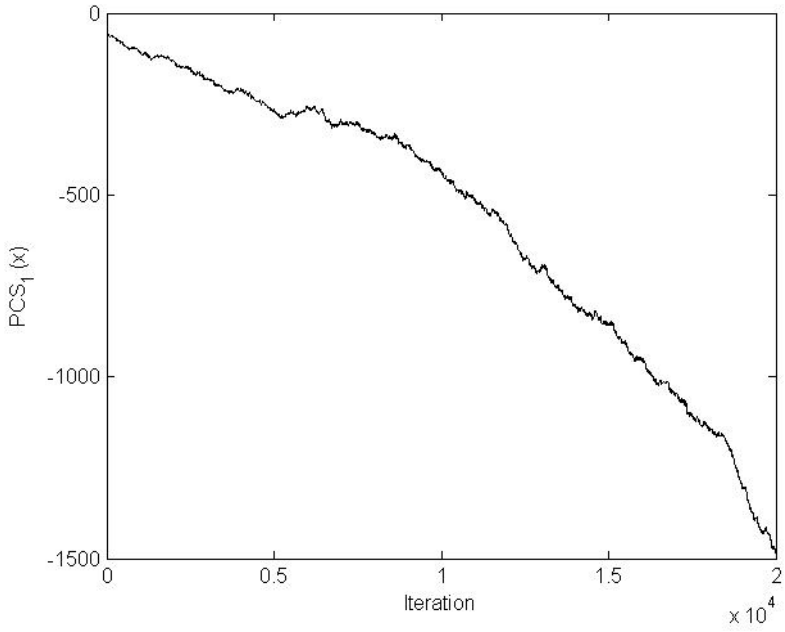

Figure B4. Trace plot of the pseudo-canonical statistic PCS_1 (**x**) for Ex. 2 of the Fiksel model $\left(\beta = 0.0545,\ \theta_1 = -1,\ \theta_2 = 1,\ R = 5\right)$.

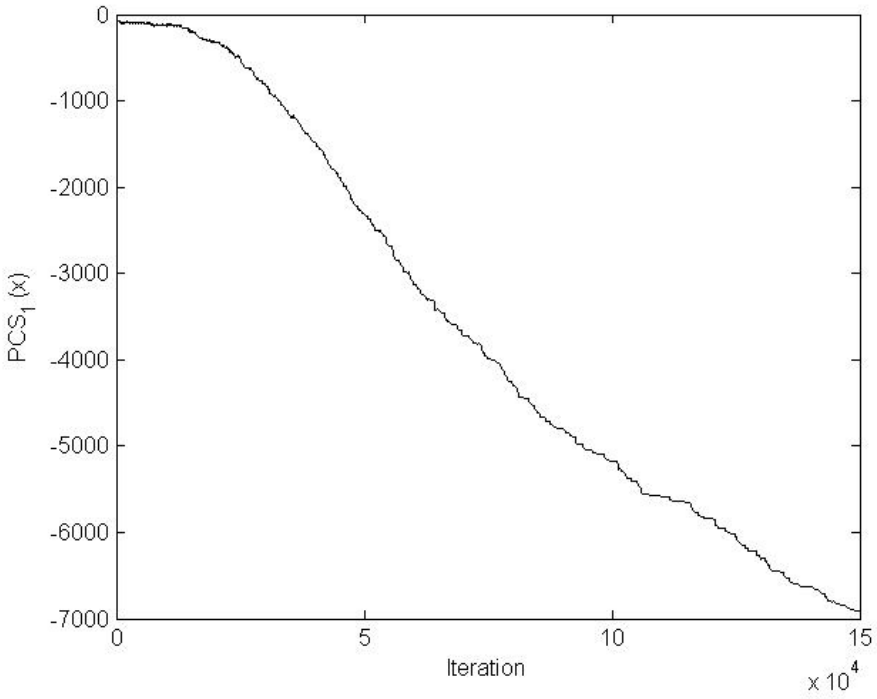

Figure B5. Trace plot of the pseudo-canonical statistic PCS_1 (**x**) for Ex. 2 of the Fiksel model with $(\beta = 0.0545,\ \theta_1 = -1,\ \theta_2 = 1,\ R = 5)$ and fixed $n = 349$.

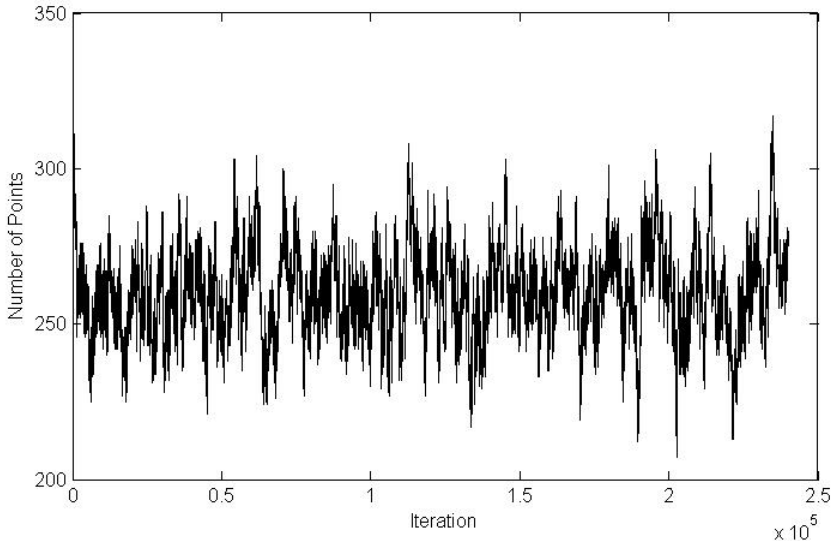

Figure B6. Trace plot of the number of points for Ex. 1 of the HCME model with $(\beta = 0.0545,\ \theta_1 = 1,\ \theta_2 = -2.5,\ \theta_3 = 2.5,\ R = 5)$ after 20,000 iterations.

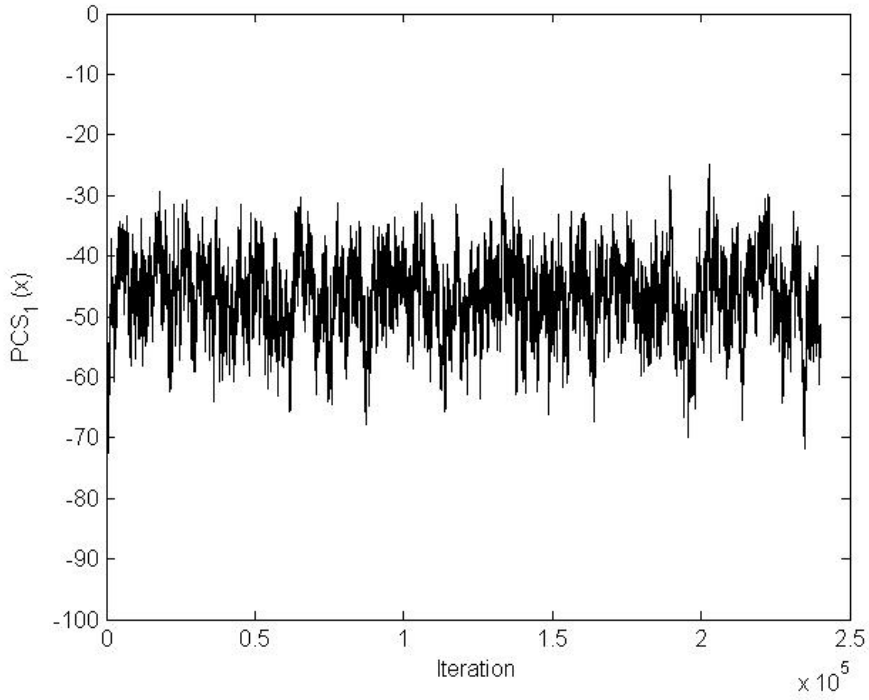

Figure B7. Trace plot of the pseudo-canonical statistic PCS_1 (**x**) for Ex. 1 of the HCME model with $(\beta = 0.0545,\ \theta_1 = 1,\ \theta_2 = -2.5,\ \theta_3 = 2.5,\ R = 5)$.

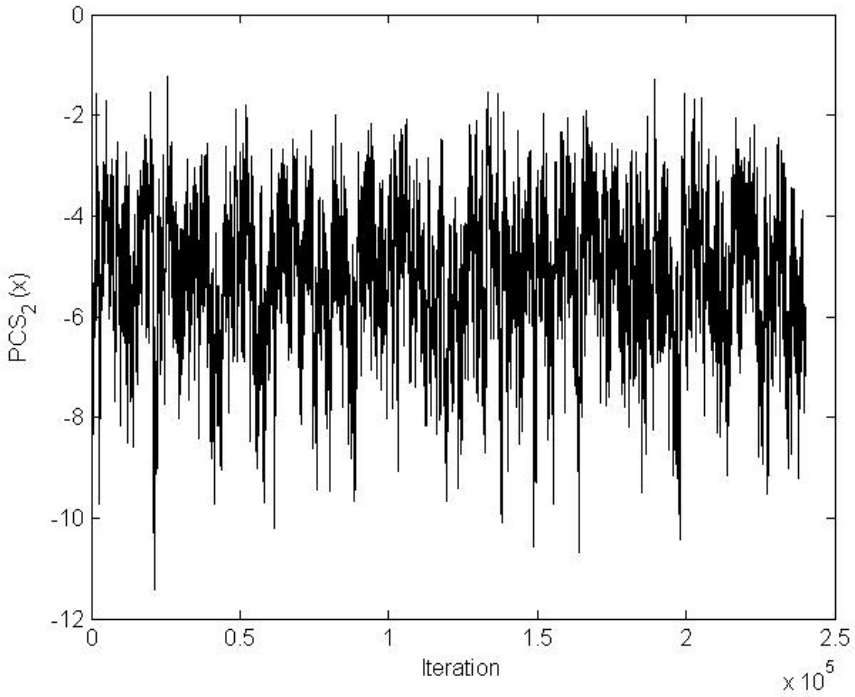

Figure B8. Trace plot of the pseudo-canonical statistic PCS_2 (**x**) for Ex. 1 of the HCME model with $(\beta = 0.0545,\ \theta_1 = 1,\ \theta_2 = -2.5,\ \theta_3 = 2.5,\ R = 5)$.

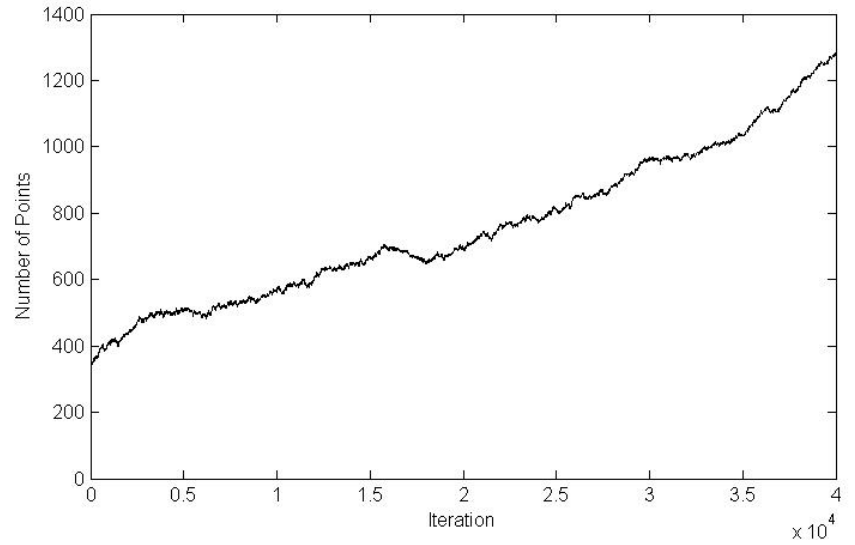

Figure B9. Trace plot of the number of points for Ex. 2 of the HCME model with $(\beta = 0.0545,\ \theta_1 = 1,\ \theta_2 = -2.5,\ \theta_3 = 1,\ R = 5)$ after 40,000 iterations.

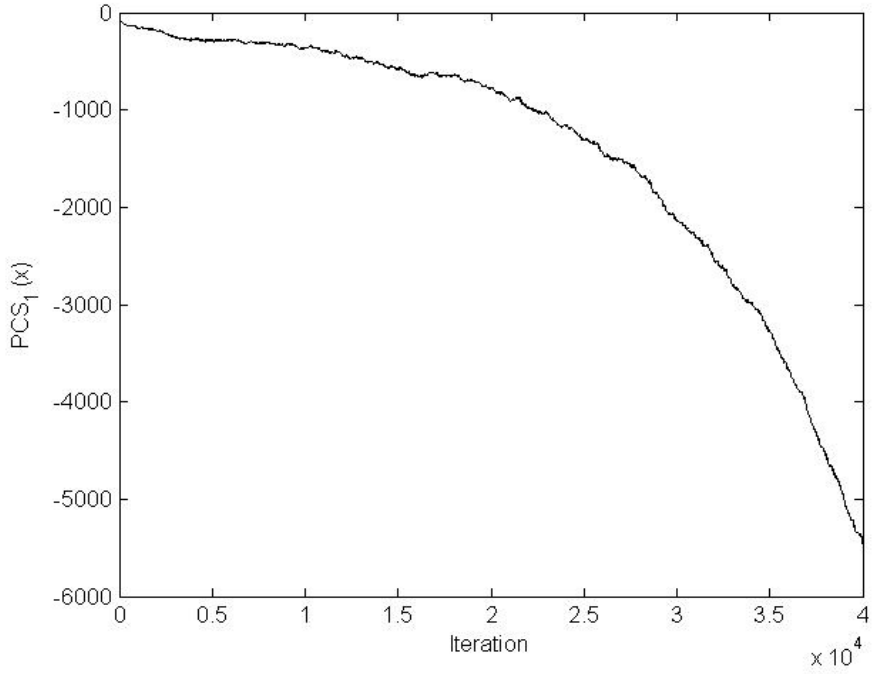

Figure B10. Trace plot of the pseudo-canonical statistic PCS_1 (**x**) for Ex. 2 of the HCME model with $(\beta = 0.0545,\ \theta_1 = 1,\ \theta_2 = -2.5,\ \theta_3 = 1.0,\ R = 5)$.

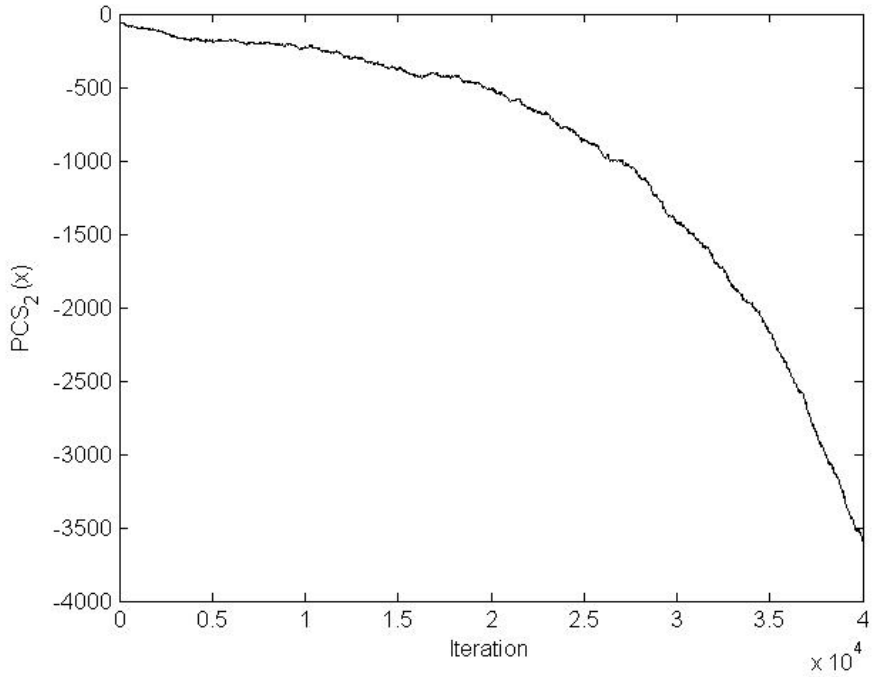

Figure B11. Trace plot of the pseudo-canonical statistic PCS_2 (**x**) for Ex. 2 of the HCME model with $(\beta = 0.0545,\ \theta_1 = 1,\ \theta_2 = -2.5,\ \theta_3 = 1,\ R = 5)$.

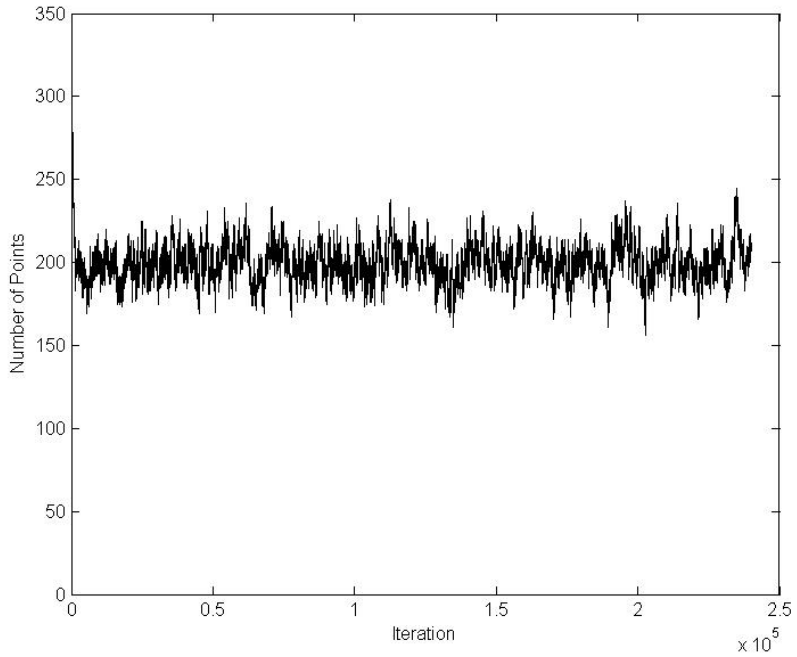

Figure B12. Trace plot of the number of points for Ex. 1 of the HCSP model with $\left(\beta = 0.0545,\ \theta_1 = 2.5,\ \theta_2 = -0.25,\ R = 5\right)$.

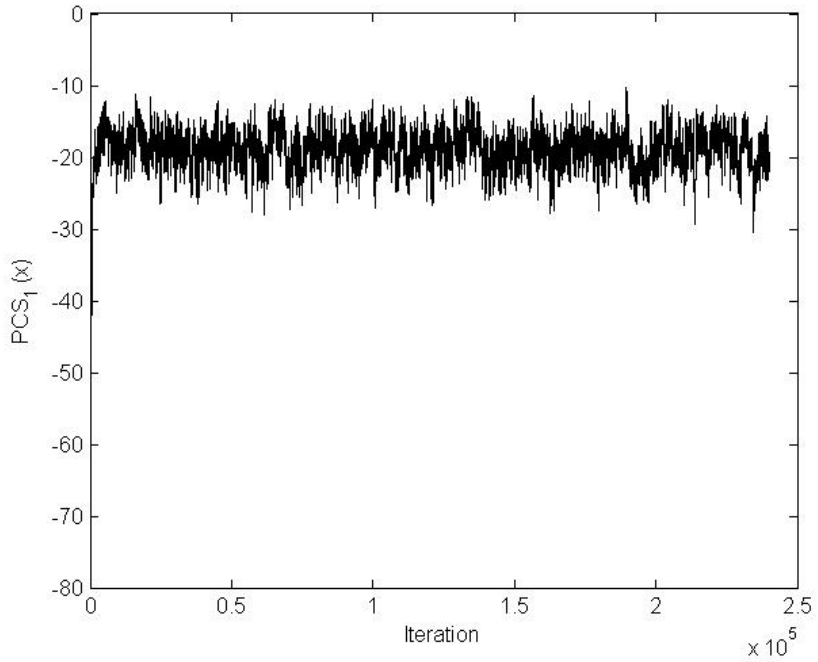

Figure B13. Trace plot of the pseudo-canonical statistic PCS_1 (**x**) for Ex. 1 of the HCSP model with $\left(\beta = 0.0545,\ \theta_1 = 2.5,\ \theta_2 = -0.25,\ R = 5\right)$.

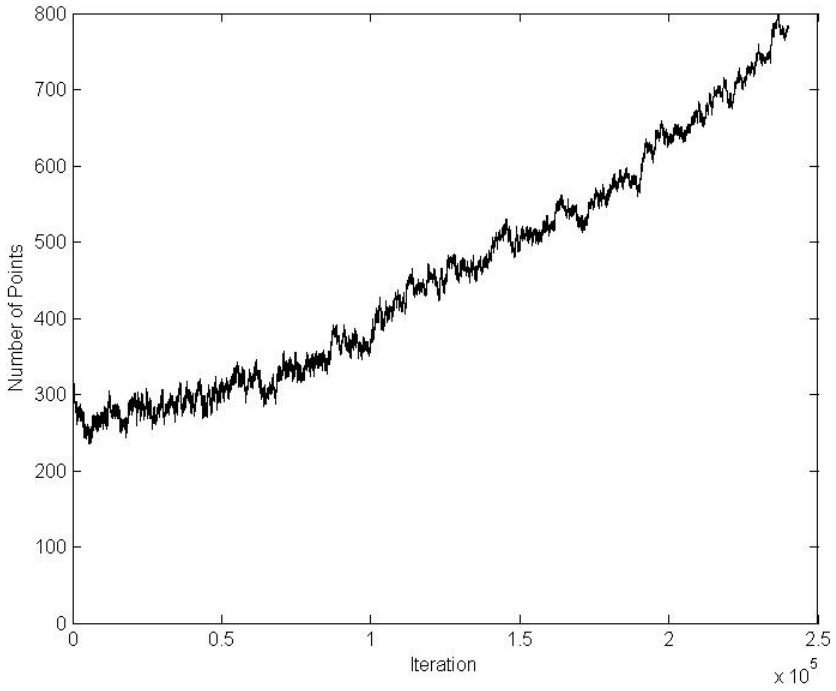

Figure B14. Trace plot of the number of points for Ex. 2 of the HCSP model with $\left(\beta = 0.0545,\ \theta_1 = 2.5,\ \theta_2 = -1,\ R = 5\right)$.

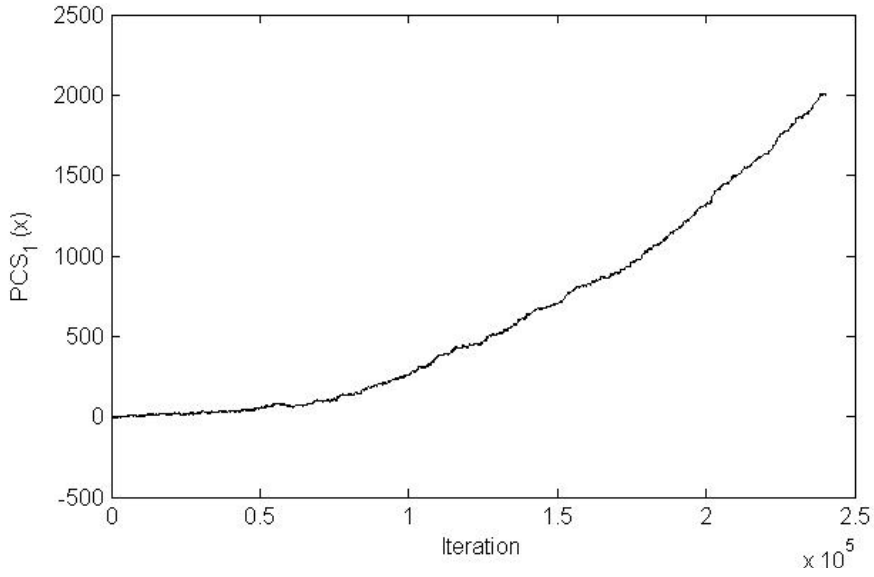

Figure B15. Trace plot of the pseudo-canonical statistic PCS_1 (**x**) for Ex. 2 of the HCSP model with $\left(\beta = 0.0545,\ \theta_1 = 2.5,\ \theta_2 = -1,\ R = 5\right)$.

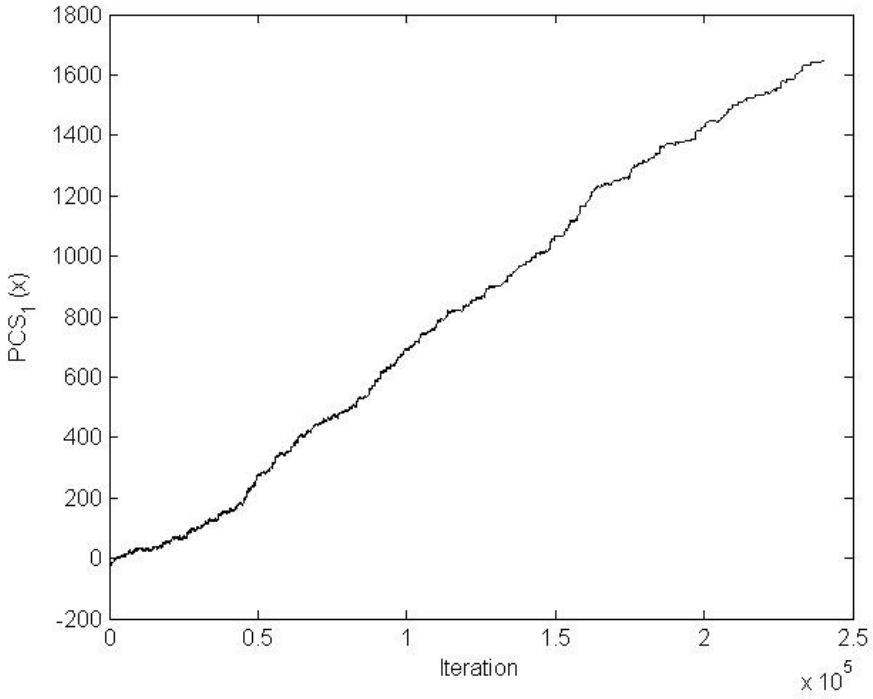

Figure B16. Trace plot of the pseudo-canonical statistic PCS_1 (**x**) for Ex. 2 of the HCSP model with $(\beta = 0.0545,\ \theta_1 = 2.5,\ \theta_2 = -1,\ R = 5)$ and fixed $n = 349$.

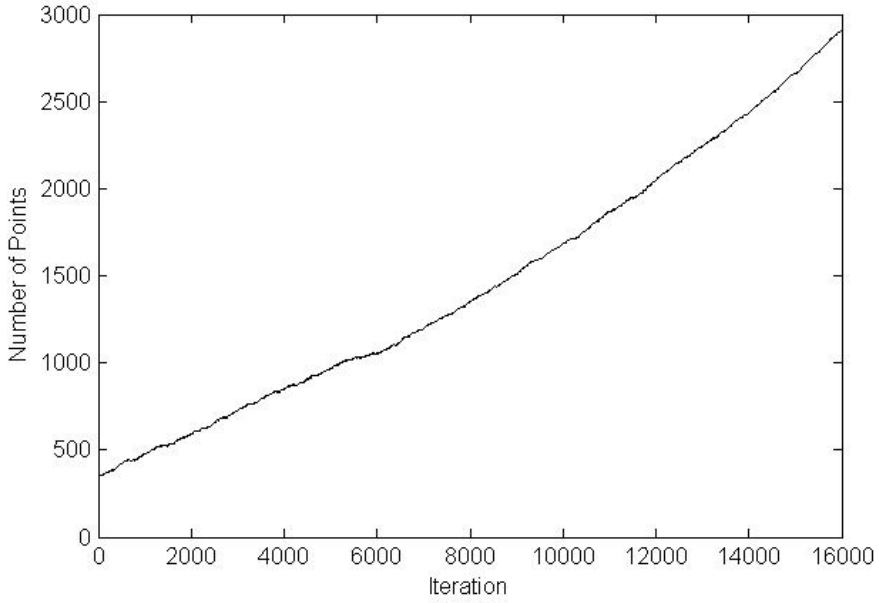

Figure B17. Trace plot of the number of points for Ex. 3 of the HCSP model with $(\beta = 0.0545,\ \theta_1 = -2.5,\ \theta_2 = -1,\ R = 5)$ after 16,000 iterations.

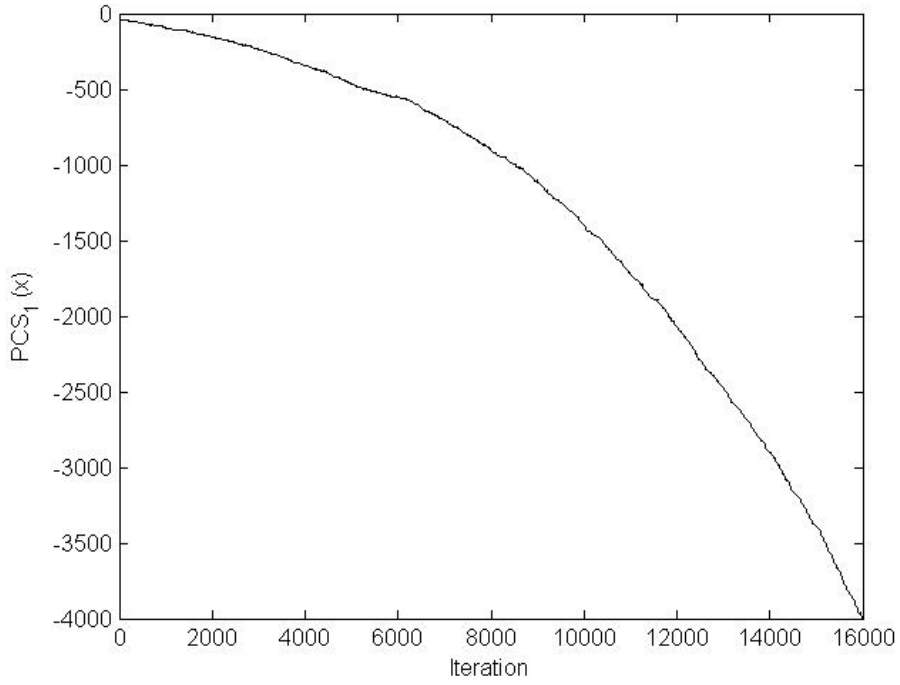

Figure B18. Trace plot of the pseudo-canonical statistic PCS_1 (**x**) for Ex. 3 of the HCSP model with $(\beta = 0.0545,\ \theta_1 = -2.5,\ \theta_2 = -1,\ R = 5)$ after 16,000 iterations.

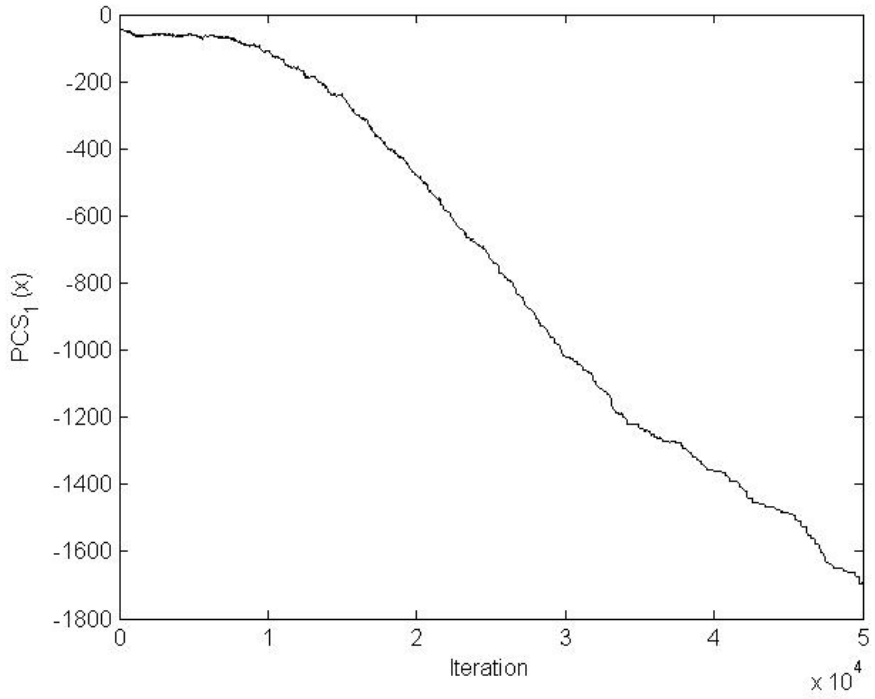

Figure B19. Trace plot of the pseudo-canonical statistic PCS_1 (**x**) for Ex. 3 of the HCSP model with $(\beta = 0.0545,\ \theta_1 = -2.5,\ \theta_2 = -1,\ R = 5)$ with fixed n = 349 after 50,000 iterations.

Appendix C

The following figures present trace plots of the "pseudo-canonical" statistics from the simulation of the Gibbs triplets models (EXPTP, GTP, and SPTP). For reference, the pseudo-canonical statistics for each model is provided (Table C1).

Table C1. Canonical parameters and "pseudo-canonical" statistics for the proposed Gibbs point process models with 2nd- and 3rd-order interaction.

Model	*Canonical Parameters*	*Pseudo-canonical statistic*
Exponential Triplets	$\log(\beta)$	$n(\mathbf{x})$
	θ_1	$PCS_1(\mathbf{x}) = \sum_{i<j} -e_{ij} \exp\left(-\left\|x_i - x_j\right\|\right) I_{(r_0,R]}\left(\left\|x_i - x_j\right\|\right)$
	θ_2	$PCS_2(\mathbf{x}) = \sum_{i<j<k} -\left(\frac{e_{ij}+e_{ik}+e_{jk}}{3}\right) \exp\left(-\left(\frac{r_{ij}+r_{ik}+r_{jk}}{3}\right)\right) I_{(r_0,R]}(r_{ij}) I_{(r_0,R]}(r_{ik}) I_{(r_0,R]}(r_{jk})$
Geyer Triplets	$\log(\beta)$	$n(\mathbf{x})$
	θ_1	$PCS_1(\mathbf{x}) = \sum_{i<j} e_{ij} \left(\frac{-1}{\left\|x_i - x_j\right\|^2 + 1}\right) I_{(r_0,R]}\left(\left\|x_i - x_j\right\|\right)$
	θ_2	$PCS_2(\mathbf{x}) = \sum_{i<j<k} -\left(\frac{e_{ij}+e_{ik}+e_{jk}}{3}\right) I_{(r_0,R]}(r_{ij}) I_{(r_0,R]}(r_{ik}) I_{(r_0,R]}(r_{jk})$
Serpentine Triplets	$\log(\beta)$	$n(\mathbf{x})$
	θ_1	$PCS_1(\mathbf{x}) = \sum_{i<j} -e_{ij} \left(\frac{\left\|x_i - x_j\right\| + \theta_2}{\left\|x_i - x_j\right\|^3 + 1}\right) I_{(r_0,R]}\left(\left\|x_i - x_j\right\|\right)$
	θ_3	$PCS_2(\mathbf{x}) = \sum_{i<j<k} -\left(\frac{e_{ij}+e_{ik}+e_{jk}}{3}\right) \exp\left(-\left(\frac{r_{ij}+r_{ik}+r_{jk}}{3}\right)\right) I_{(r_0,R]}(r_{ij}) I_{(r_0,R]}(r_{ik}) I_{(r_0,R]}(r_{jk})$

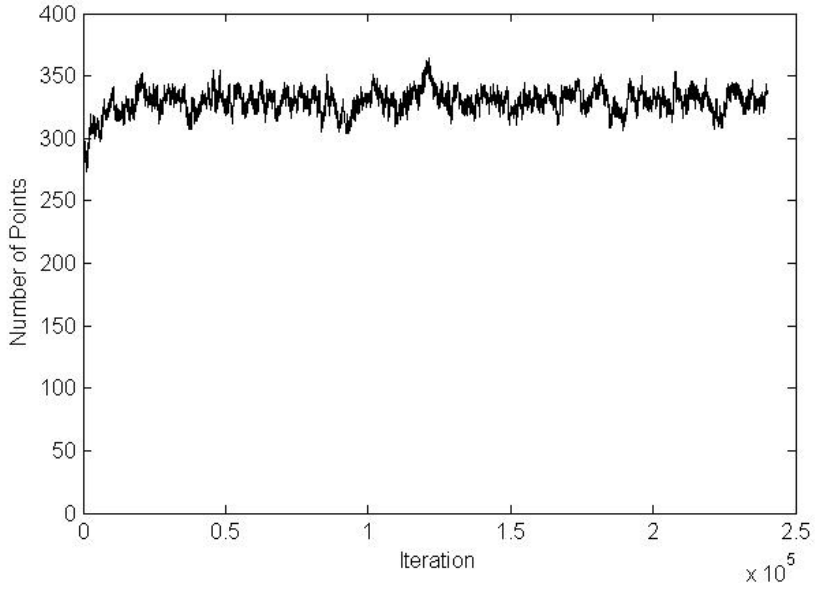

Figure C1. Trace plot of the number of points for Ex. 1 of the Geyer triplets model with $(\beta = 0.0545,\ \theta_1 = -1,\ \theta_2 = 2.5,\ R = 5)$.

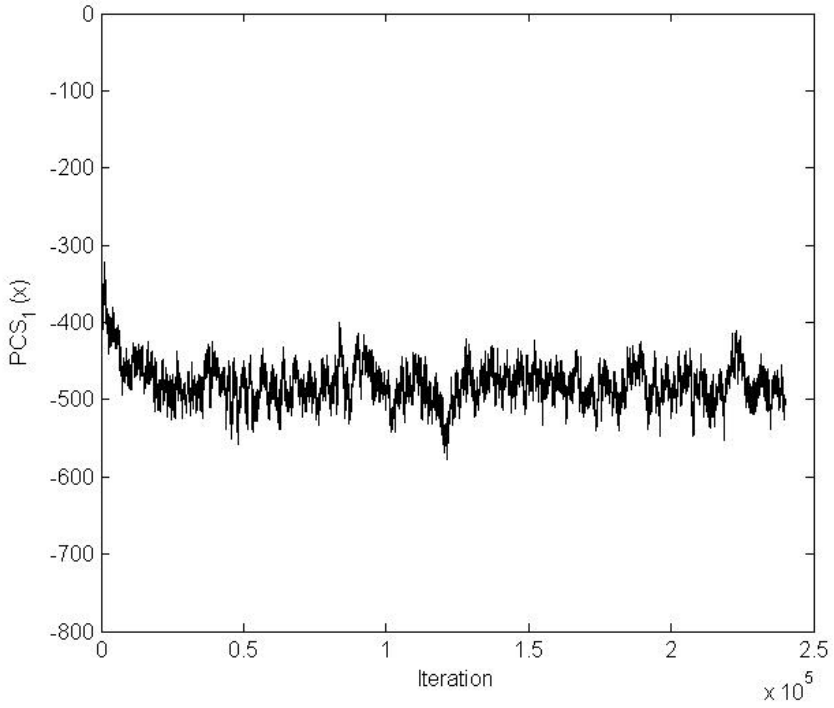

Figure C2. Trace plot of the pseudo-canonical statistic PCS_1 (**x**) for Ex. 1 of the Geyer triplets model with $(\beta = 0.0545,\ \theta_1 = -1,\ \theta_2 = 2.5,\ R = 5)$.

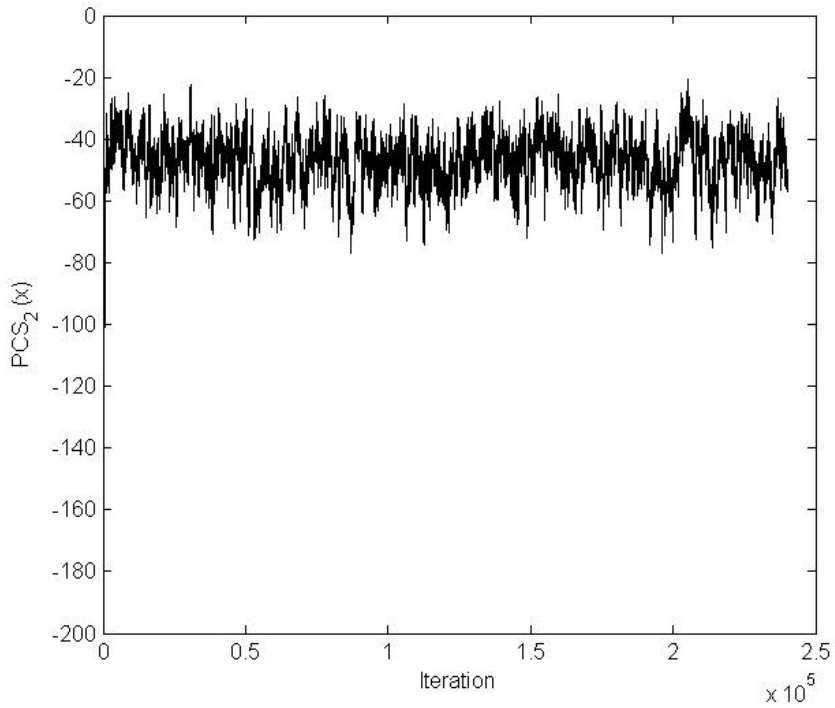

Figure C3. Trace plot of the pseudo-canonical statistic PCS_2 (**x**) for Ex. 1 of the Geyer triplets model with $(\beta = 0.0545,\ \theta_1 = -1,\ \theta_2 = 2.5,\ R = 5)$.

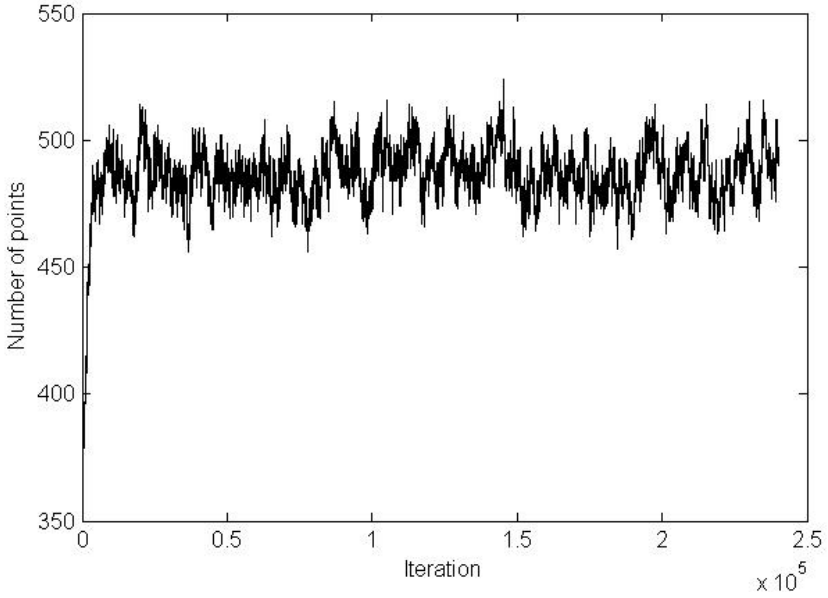

Figure C4. Trace plot of the number of points for Ex. 2 of the Geyer triplets model with $(\beta = 0.0545,\ \theta_1 = -1,\ \theta_2 = 1,\ R = 5)$.

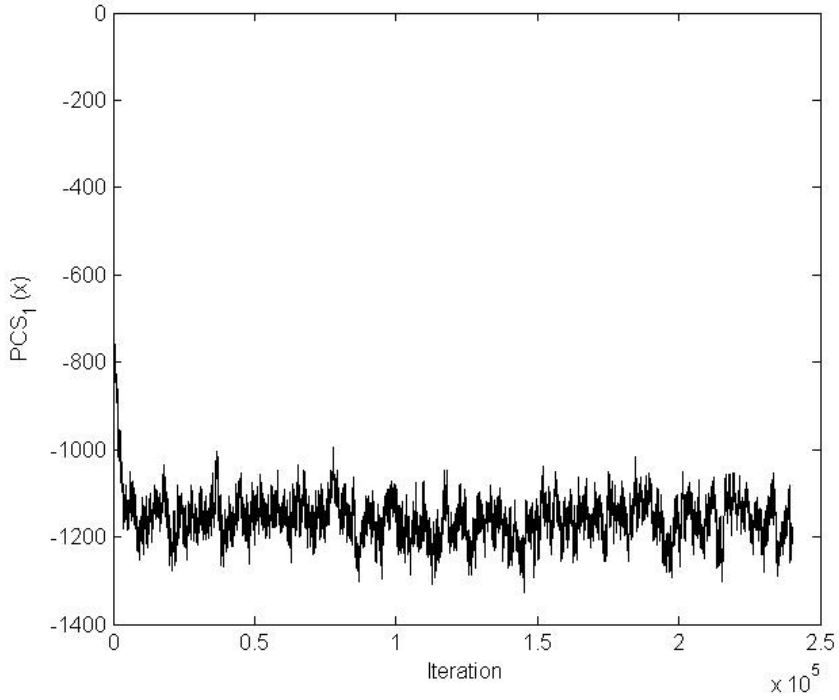

Figure C5. Trace plot of the pseudo-canonical statistic PCS_1 (**x**) for Ex. 2 of the Geyer triplets model with $\left(\beta = 0.0545,\ \theta_1 = -1,\ \theta_2 = 1,\ R = 5\right)$.

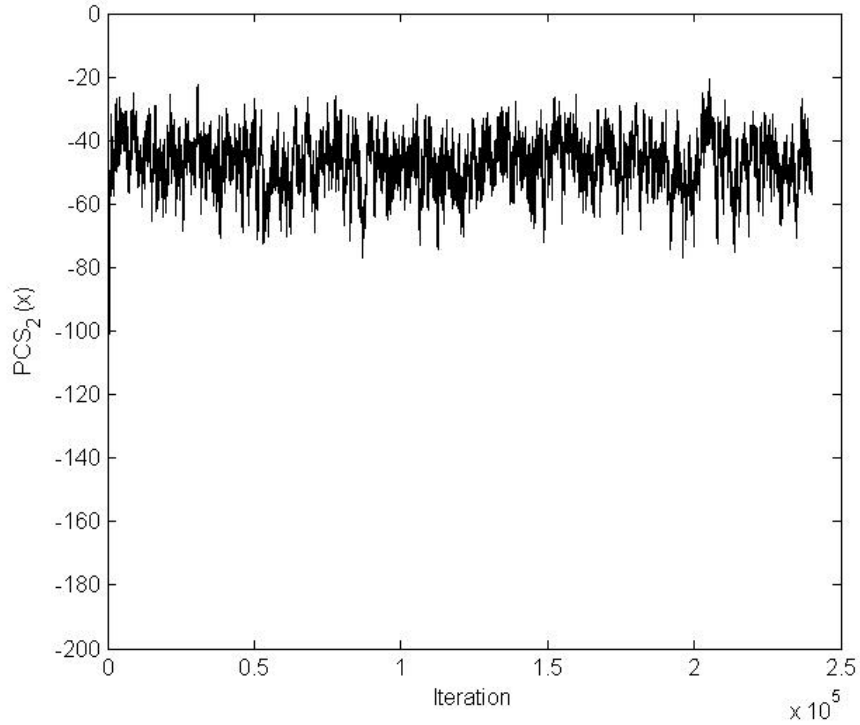

Figure C6. Trace plot of the pseudo-canonical statistic PCS_2 (**x**) for Ex. 2 of the Geyer triplets model with $\left(\beta = 0.0545,\ \theta_1 = -1,\ \theta_2 = 1,\ R = 5\right)$.

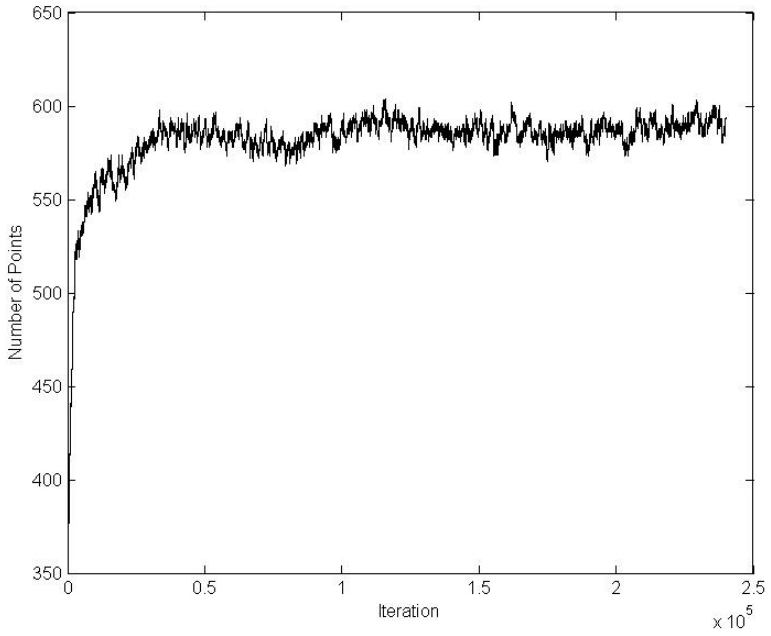

Figure C7. Trace plot of the number of points for Ex. 3 of the Geyer triplets model with $(\beta = 0.0545,\ \theta_1 = -2.5,\ \theta_2 = 2.5,\ R = 5)$.

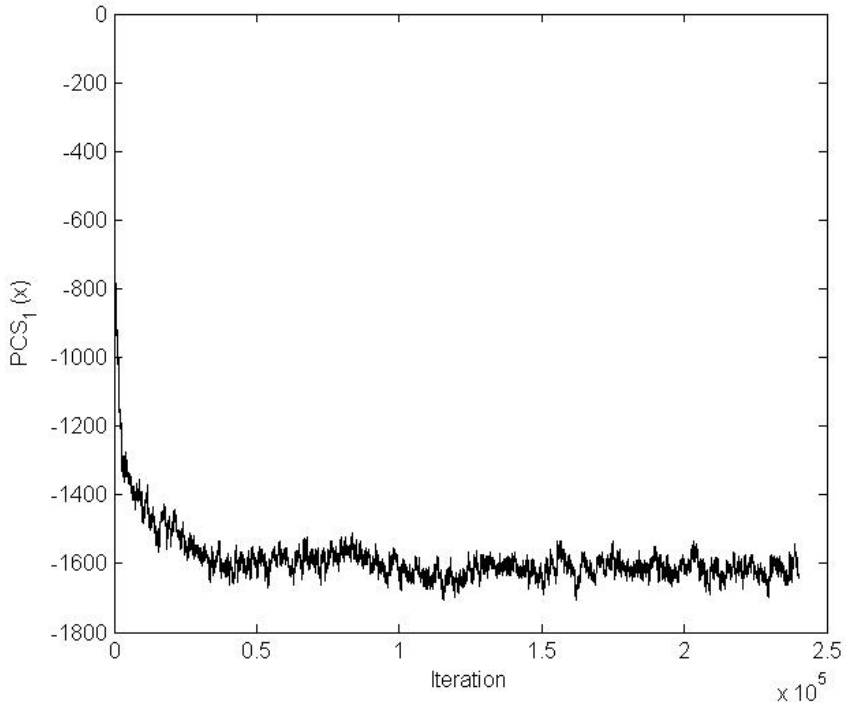

Figure C8. Trace plot of the pseudo-canonical statistic PCS_1 (**x**) for Ex. 3 of the Geyer triplets model with $(\beta = 0.0545,\ \theta_1 = -2.5,\ \theta_2 = 2.5,\ R = 5)$.

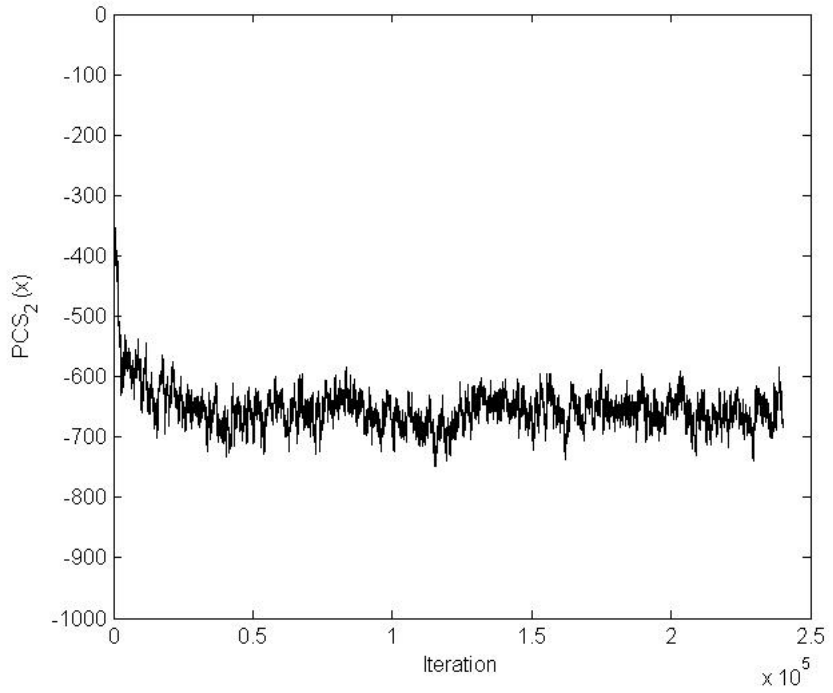

Figure C9. Trace plot of the pseudo-canonical statistic PCS_2 (**x**) for Ex. 3 of the Geyer triplets model with $(\beta = 0.0545,\ \theta_1 = -2.5,\ \theta_2 = 2.5,\ R = 5)$.

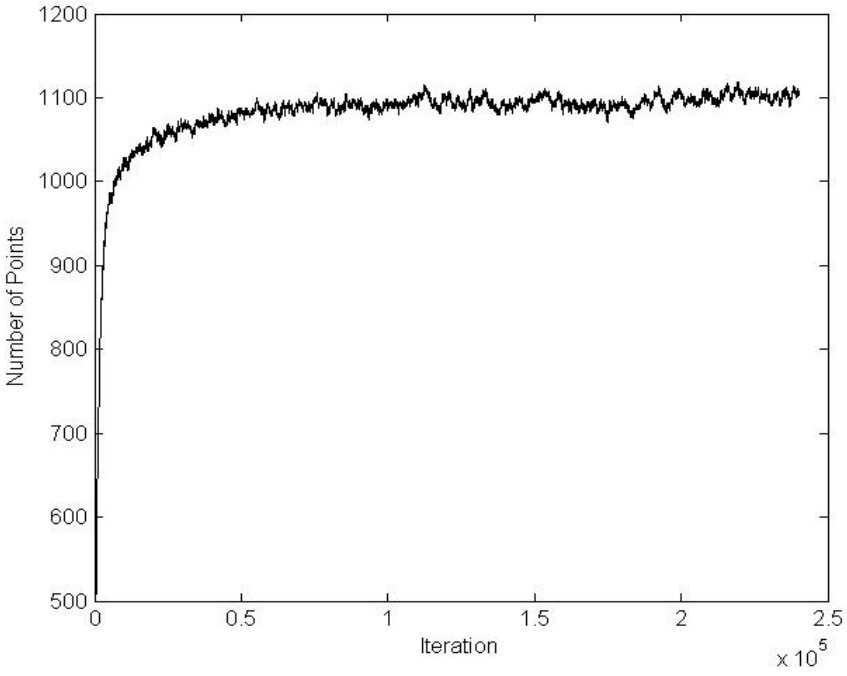

Figure C10. Trace plot of the number of points for Ex. 4 of the Geyer triplets model with $(\beta = 0.0545,\ \theta_1 = -2.5,\ \theta_2 = 1,\ R = 5)$.

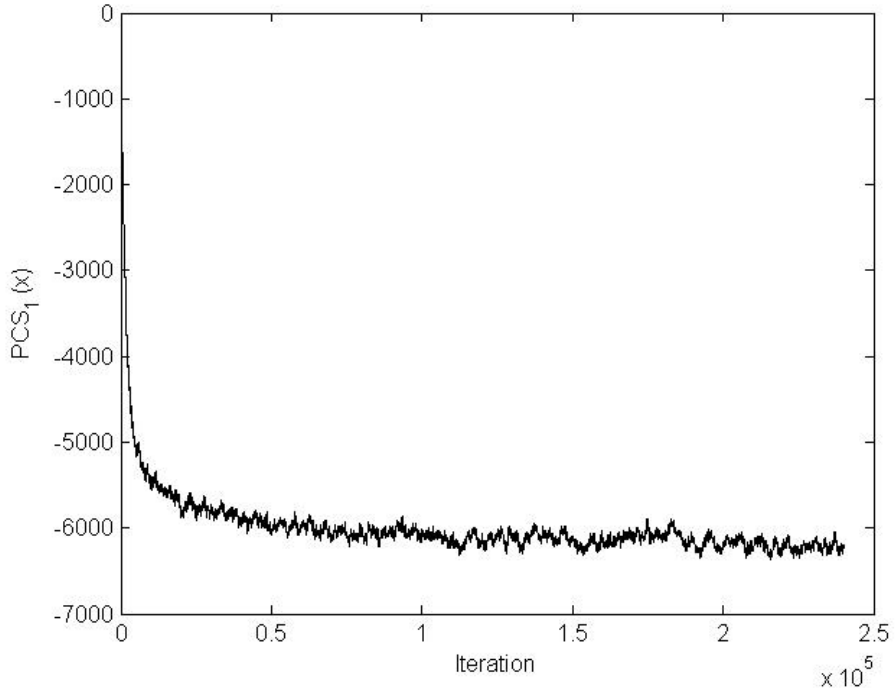

Figure C11. Trace plot of the pseudo-canonical statistic PCS_1 (**x**) for Ex. 4 of the Geyer triplets model with $(\beta = 0.0545,\ \theta_1 = -2.5,\ \theta_2 = 1,\ R = 5)$..

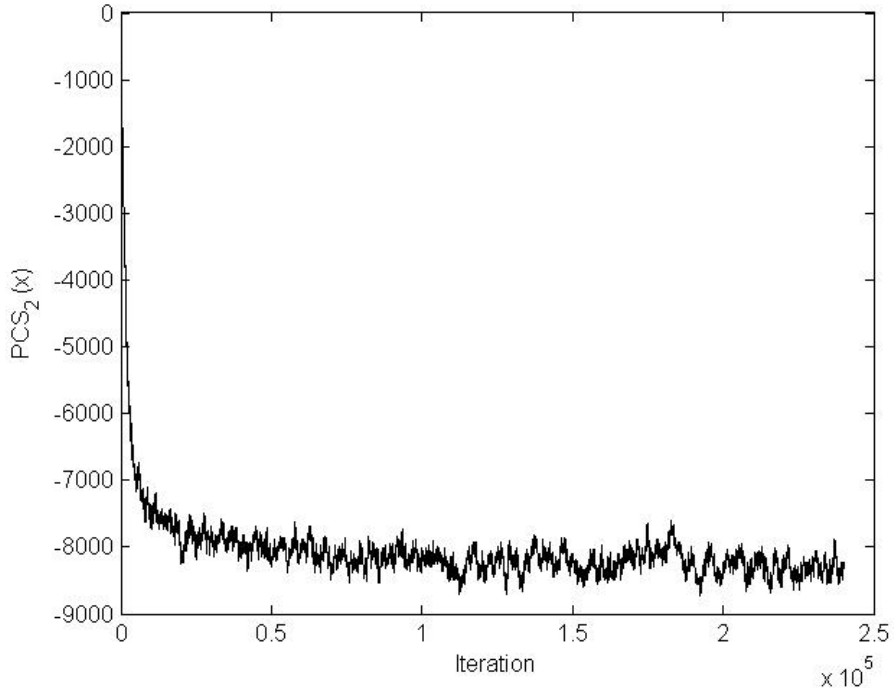

Figure C12. Trace plot of the pseudo-canonical statistic PCS_2 (**x**) for Ex. 4 of the Geyer triplets model with $(\beta = 0.0545,\ \theta_1 = -2.5,\ \theta_2 = 1,\ R = 5)$.

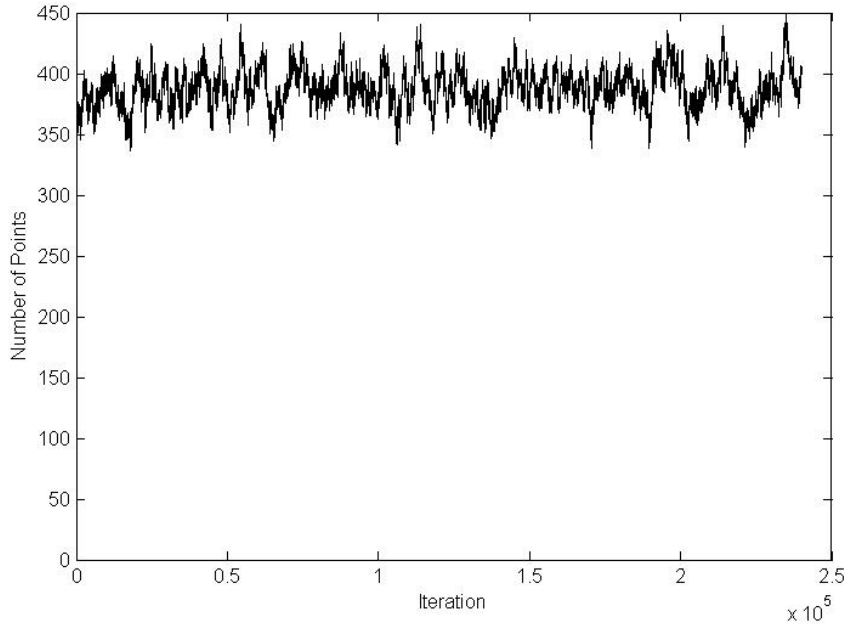

Figure C13. Trace plot of the number of points for Ex. 1 of the EXPTP model with $\left(\beta = 0.0545,\ \theta_1 = -2.5,\ \theta_2 = 2.5,\ R = 5\right)$.

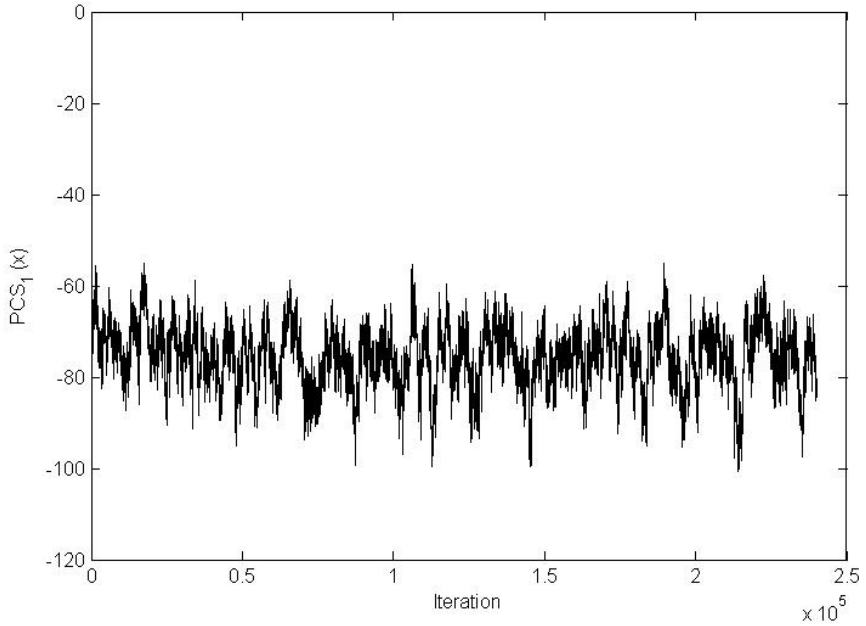

Figure C14. Trace plot of the pseudo-canonical statistic PCS_1 (**x**) for Ex. 1 of the EXPTP model with $\left(\beta = 0.0545,\ \theta_1 = -2.5,\ \theta_2 = 2.5,\ R = 5\right)$.

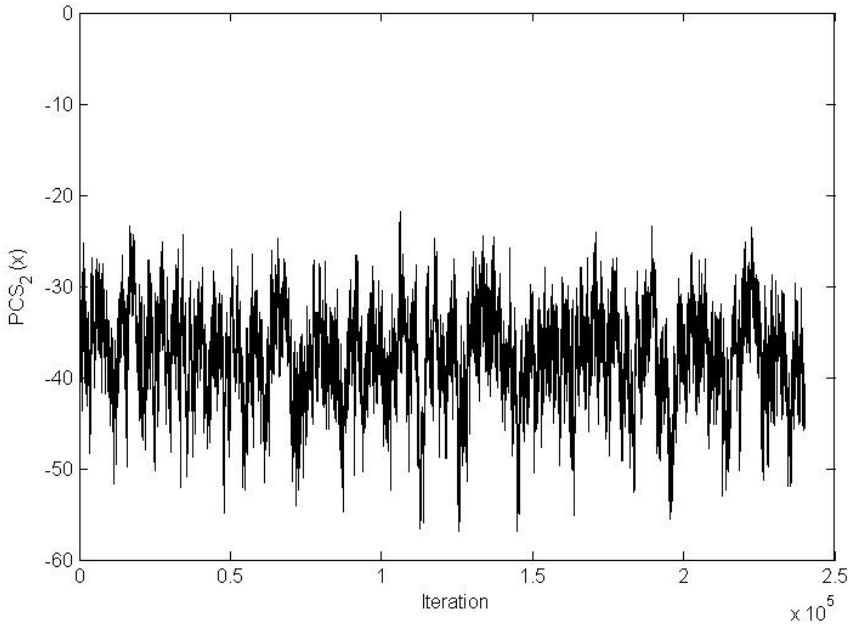

Figure C15. Trace plot of the pseudo-canonical statistic PCS_2 (**x**) for Ex. 1 of the EXPTP model with $(\beta = 0.0545,\ \theta_1 = -2.5,\ \theta_2 = 2.5,\ R = 5)$.

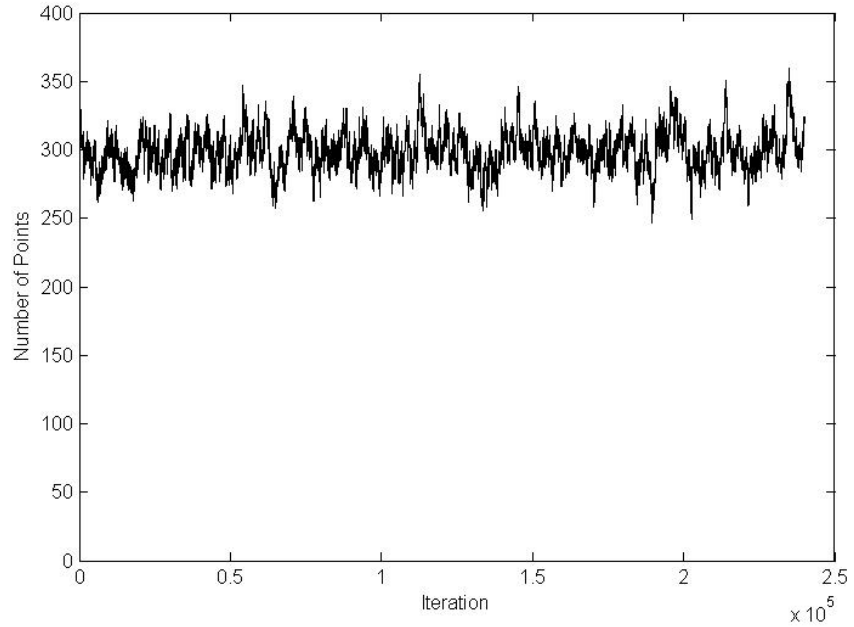

Figure C16. Trace plot of the number of points for Ex. 2 of the EXPTP model with $(\beta = 0.0545,\ \theta_1 = -1,\ \theta_2 = 2.5,\ R = 5)$.

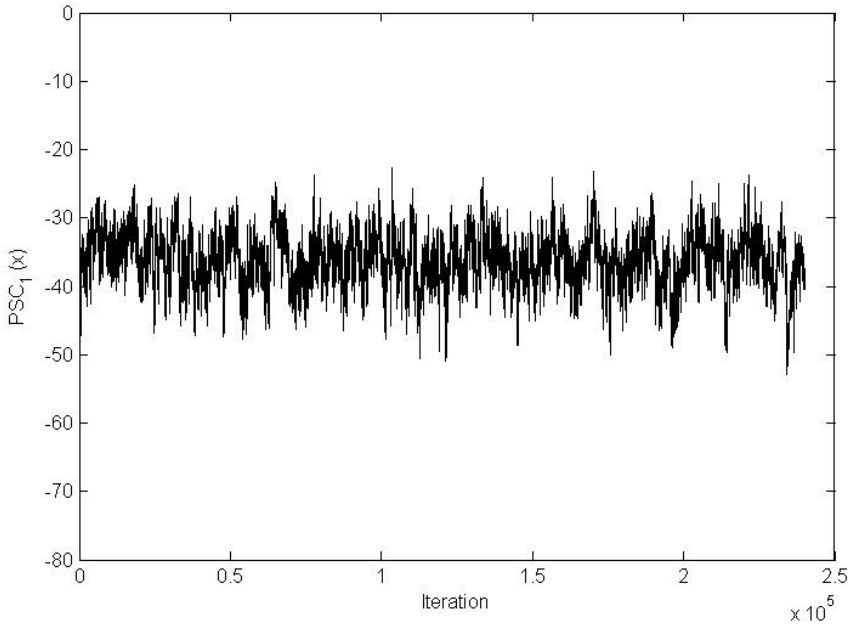

Figure C17. Trace plot of the pseudo-canonical statistic PCS_1 (**x**) for Ex. 2 of the EXPTP model with $(\beta = 0.0545,\ \theta_1 = -1,\ \theta_2 = 2.5,\ R = 5)$.

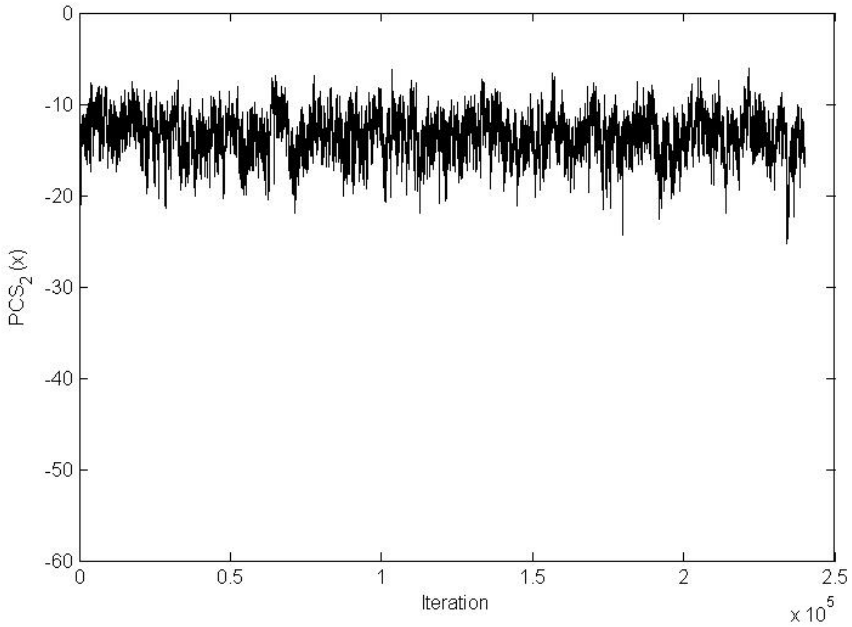

Figure C18. Trace plot of the pseudo-canonical statistic PCS_2 (**x**) for Ex. 2 of the EXPTP model with $(\beta = 0.0545,\ \theta_1 = -1,\ \theta_2 = 2.5,\ R = 5)$.

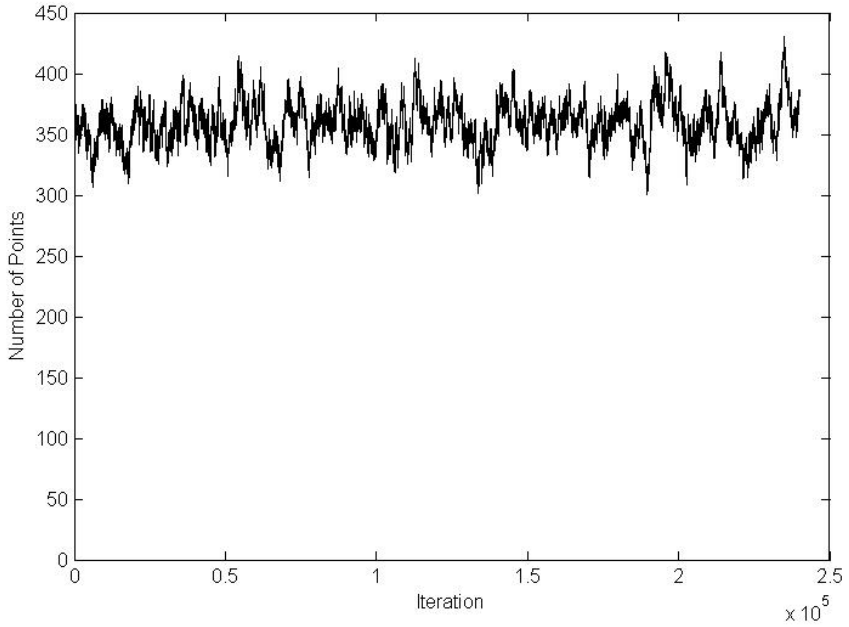

Figure C19. Trace plot of the number of points for Ex. 3 of the EXPTP model with $(\beta = 0.0545,\ \theta_1 = -1,\ \theta_2 = 1,\ R = 5)$.

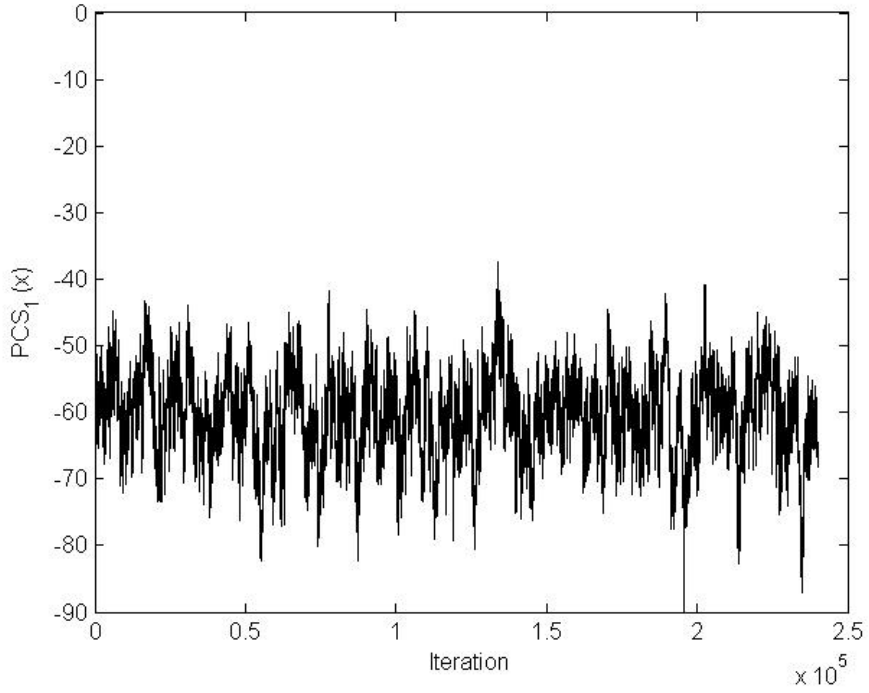

Figure C20. Trace plot of the pseudo-canonical statistic PCS_1 (**x**) for Ex. 3 of the EXPTP model with $(\beta = 0.0545,\ \theta_1 = -1,\ \theta_2 = 1,\ R = 5)$.

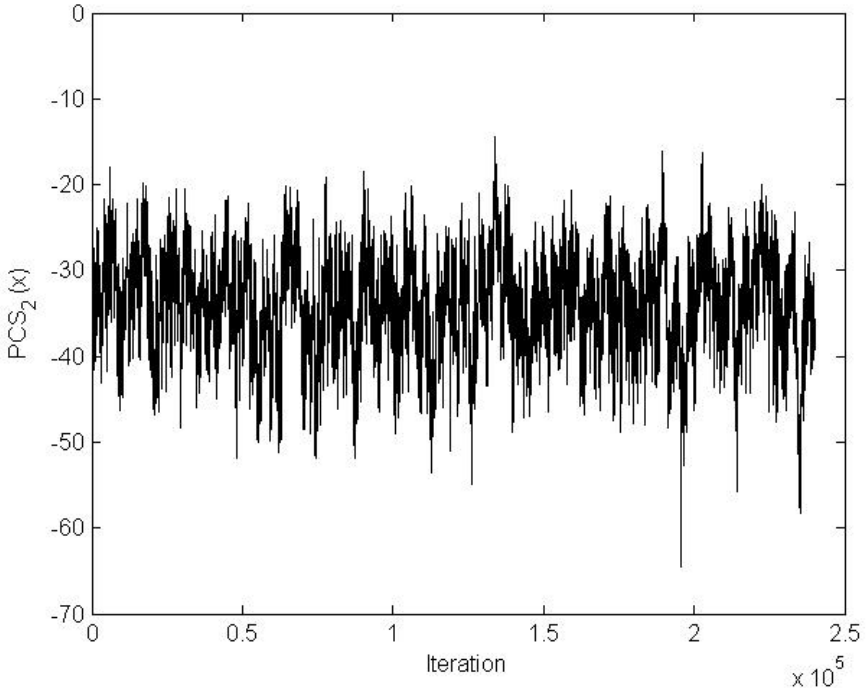

Figure C21. Trace plot of the pseudo-canonical statistic PCS_2 (**x**) for Ex. 3 of the EXPTP model with $(\beta = 0.0545,\ \theta_1 = -1,\ \theta_2 = 1,\ R = 5)$.

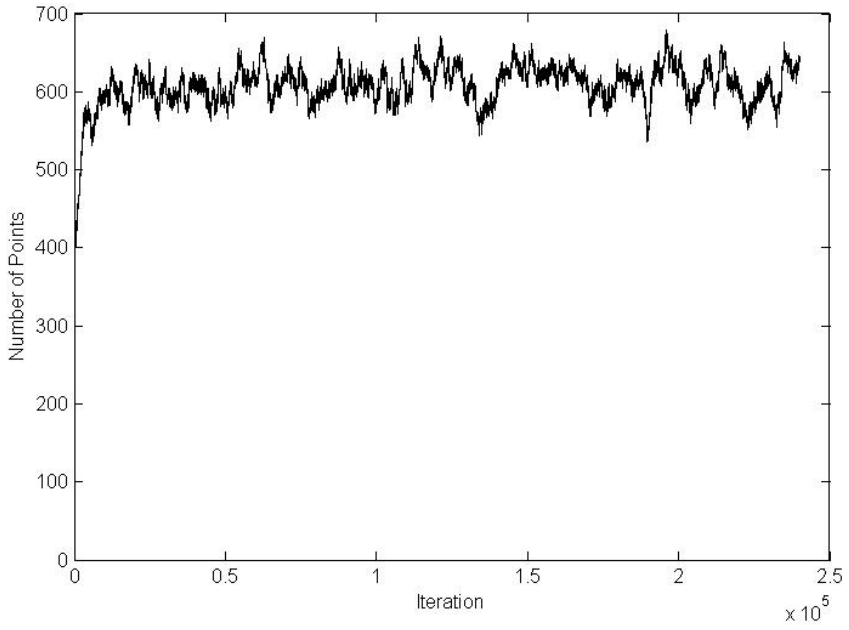

Figure C22. Trace plot of the number of points for Ex. 4 of the EXPTP model with $(\beta = 0.0545,\ \theta_1 = -2.5,\ \theta_2 = 1,\ R = 5)$.

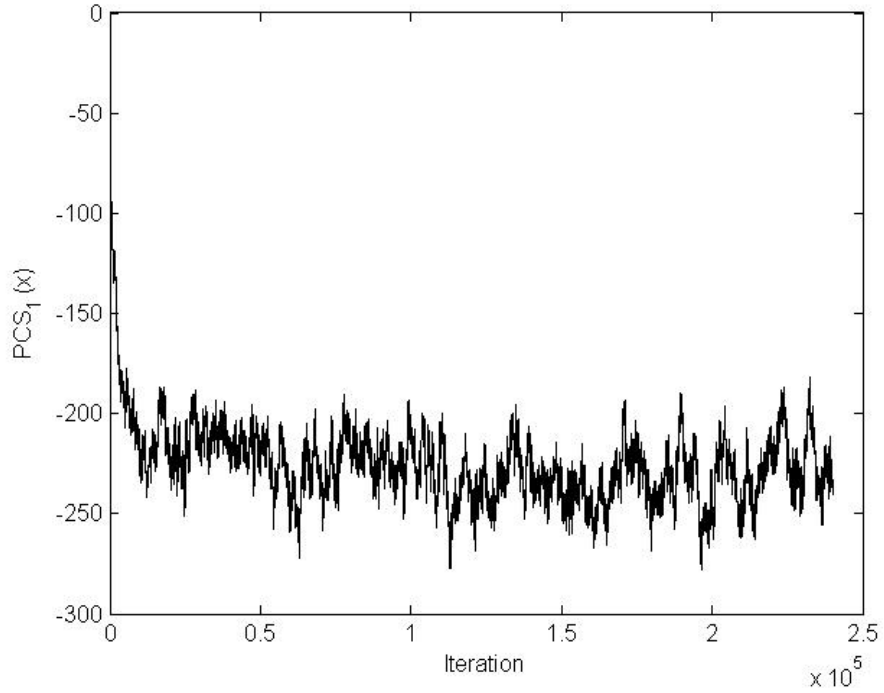

Figure C23. Trace plot of the pseudo-canonical statistic PCS_1 (**x**) for Ex. 4 of the EXPTP model with $(\beta = 0.0545,\ \theta_1 = -2.5,\ \theta_2 = 1,\ R = 5)$.

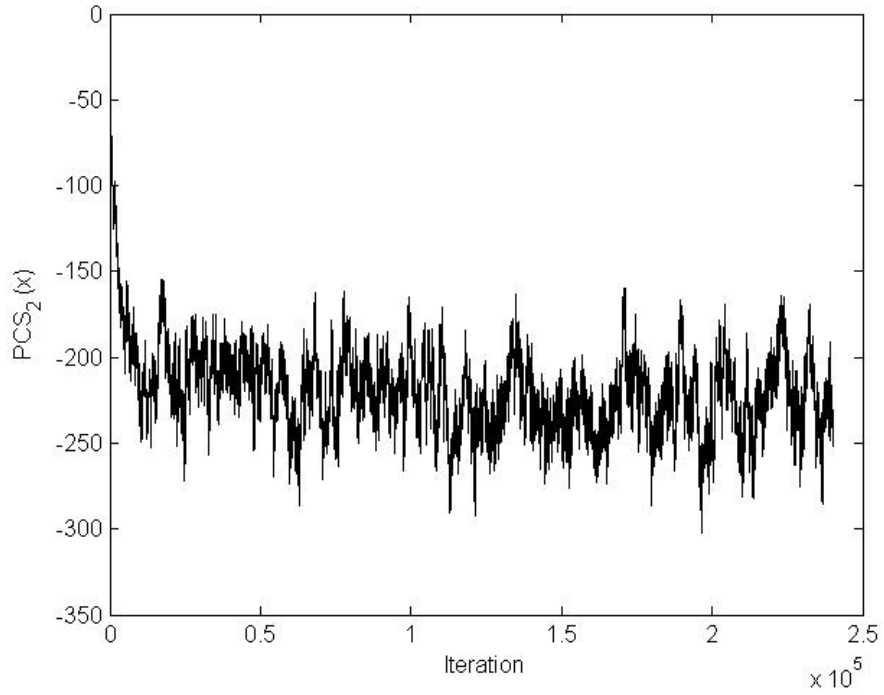

Figure C24. Trace plot of the pseudo-canonical statistic PCS_2 (**x**) for Ex. 4 of the EXPTP model with $(\beta = 0.0545,\ \theta_1 = -2.5,\ \theta_2 = 1,\ R = 5)$.

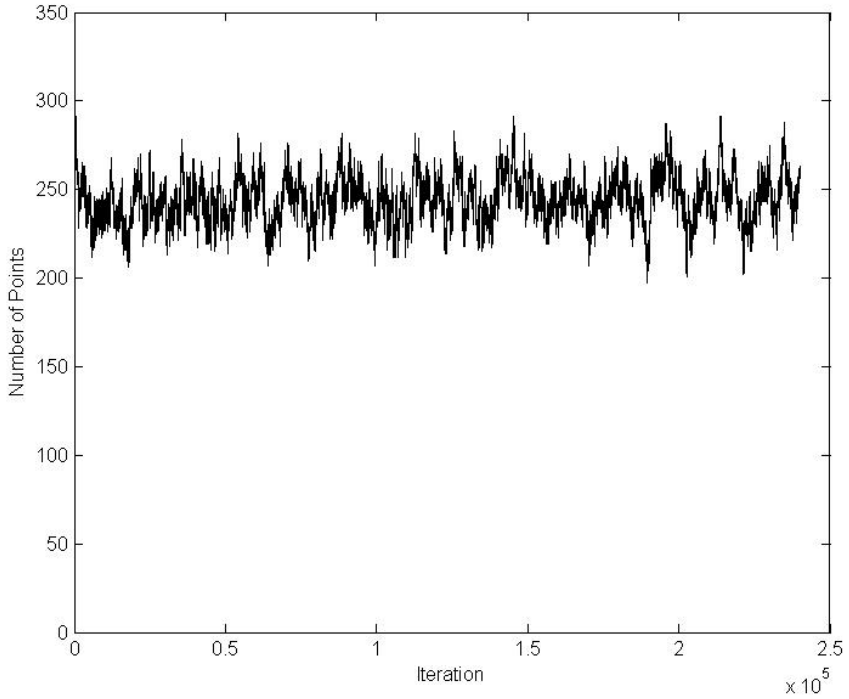

Figure C25. Trace plot of the number of points for Ex. 1 of the SPTP model with $\left(\beta = 0.0545,\ \theta_1 = 2.5,\ \theta_2 = -1,\ \theta_3 = 1,\ R = 5\right)$.

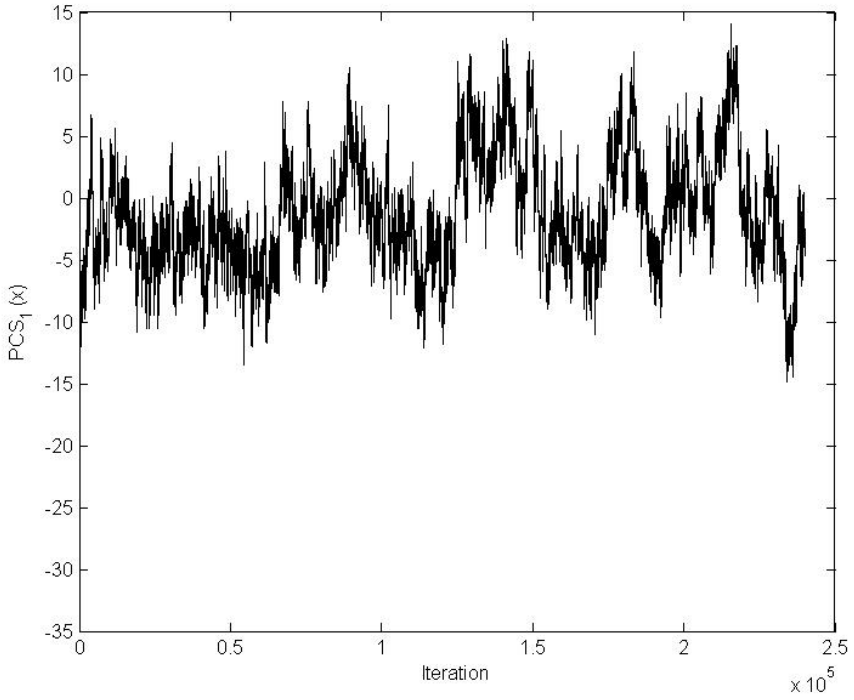

Figure C26. Trace plot of the pseudo-canonical statistic PCS_1 (**x**) for Ex. 1 of the SPTP model with $\left(\beta = 0.0545,\ \theta_1 = 2.5,\ \theta_2 = -1,\ \theta_3 = 1,\ R = 5\right)$.

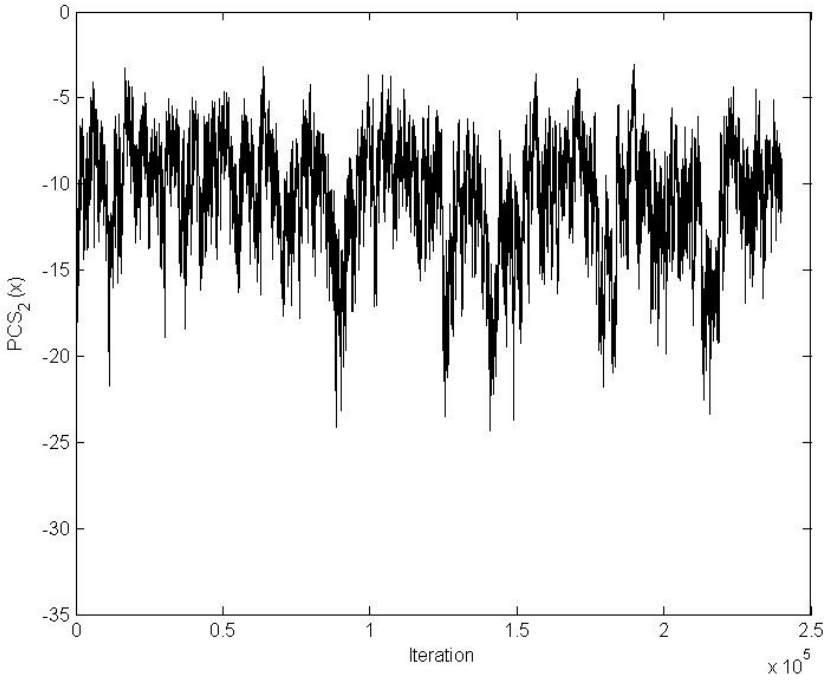

Figure C27. Trace plot of the pseudo-canonical statistic PCS_2 (**x**) for Ex. 1 of the SPTP model with $(\beta = 0.0545,\ \theta_1 = 2.5,\ \theta_2 = -1,\ \theta_3 = 1,\ R = 5)$.

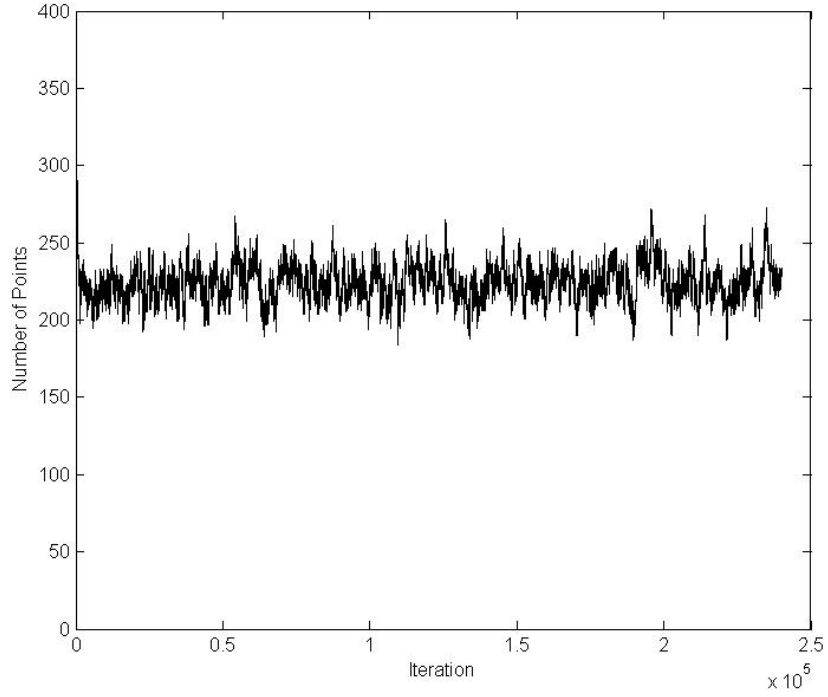

Figure C28. Trace plot of the number of points for Ex. 2 of the SPTP model with $(\beta = 0.0545,\ \theta_1 = 2.5,\ \theta_2 = -1,\ \theta_3 = 2.5,\ R = 5)$.

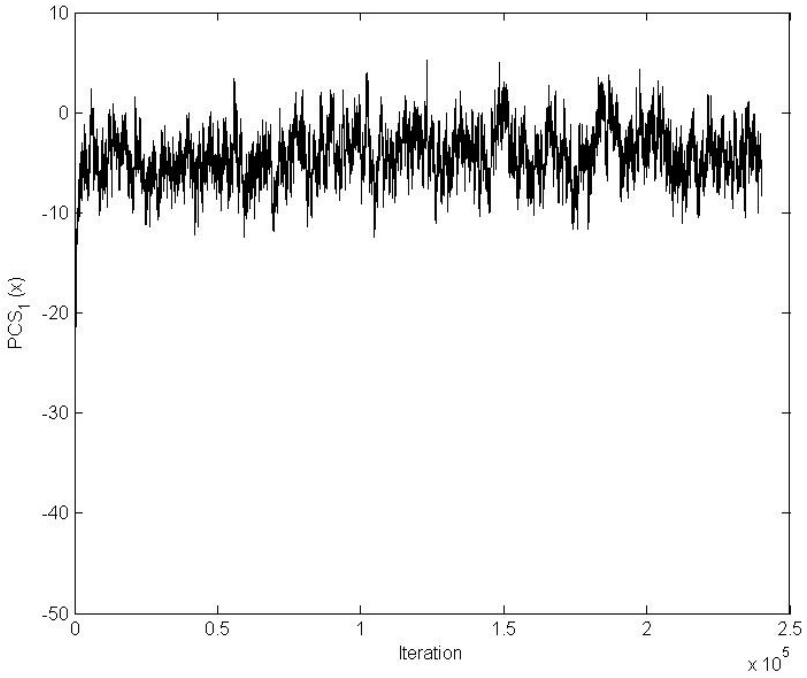

Figure C29. Trace plot of the pseudo-canonical statistic PCS_1 (**x**) for Ex. 2 of the SPTP model with $\left(\beta = 0.0545,\ \theta_1 = 2.5,\ \theta_2 = -1,\ \theta_3 = 2.5,\ R = 5\right)$.

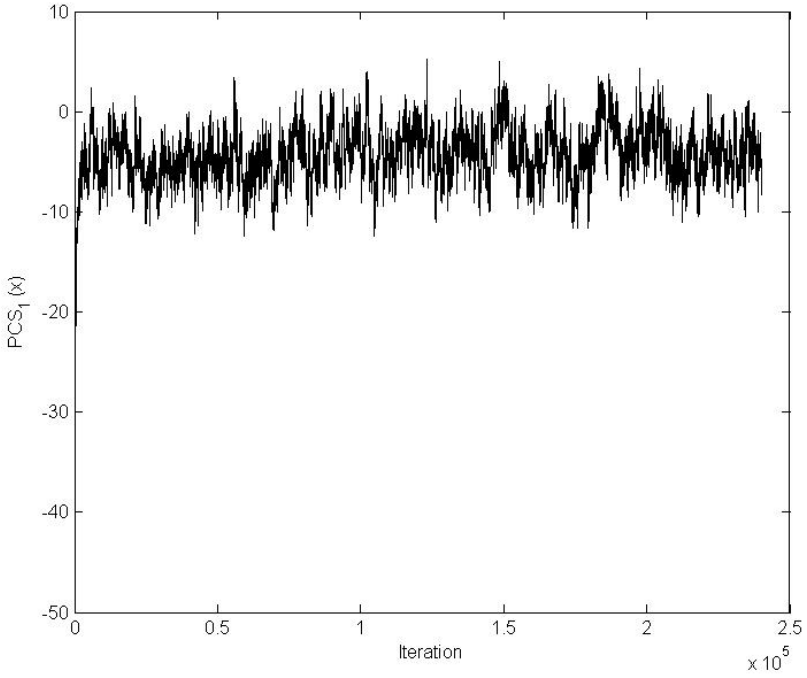

Figure C30. Trace plot of the pseudo-canonical statistic PCS_2 (**x**) for Ex. 2 of the SPTP model with $\left(\beta = 0.0545,\ \theta_1 = 2.5,\ \theta_2 = -1,\ \theta_3 = 2.5,\ R = 5\right)$.

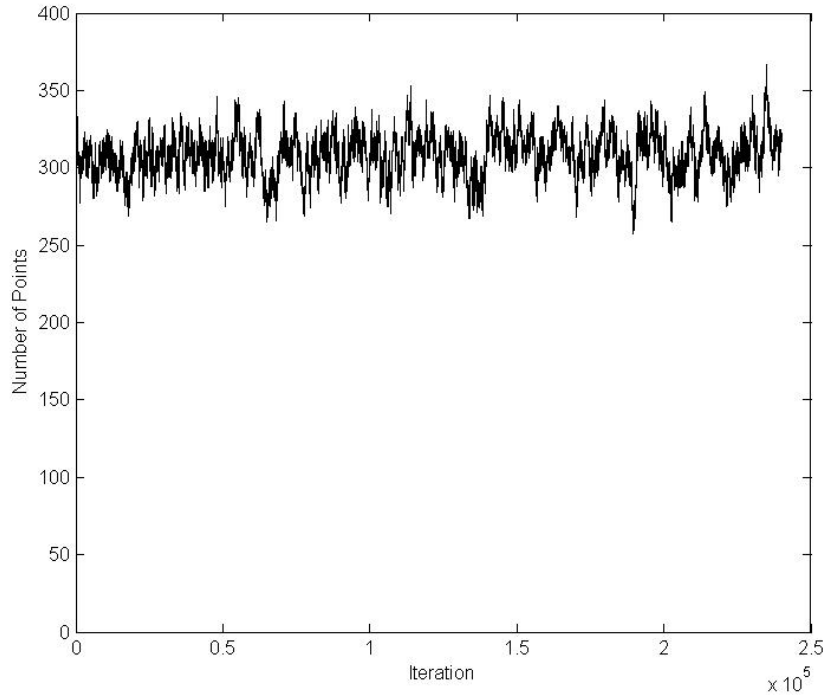

Figure C31. Trace plot of the number of points for Ex. 3 of the SPTP model with $(\beta = 0.0545,\ \theta_1 = 1,\ \theta_2 = -2.5,\ \theta_3 = 2.5,\ R = 5)$.

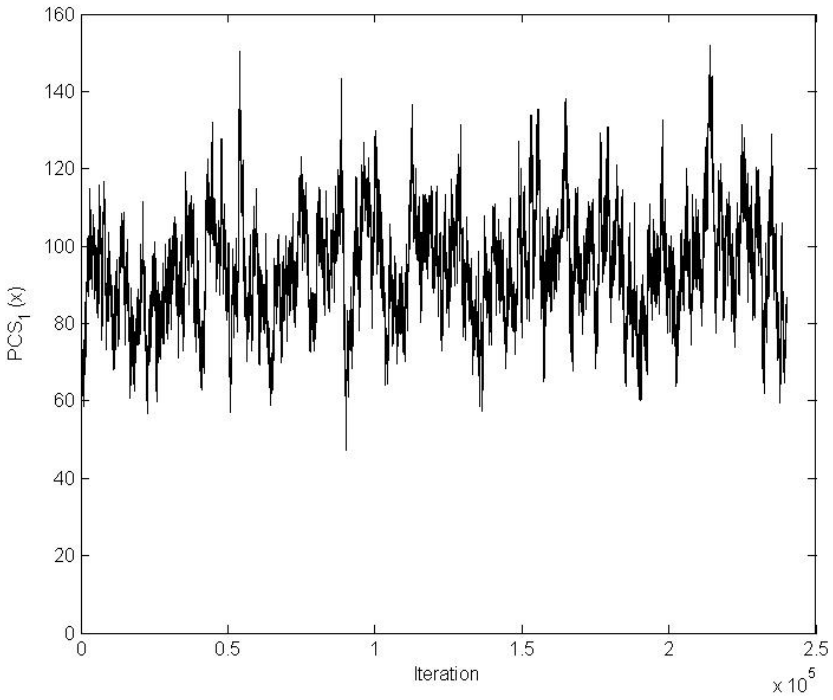

Figure C32. Trace plot of the pseudo-canonical statistic PCS_1 (**x**) for Ex. 3 of the SPTP model with $(\beta = 0.0545,\ \theta_1 = 1,\ \theta_2 = -2.5,\ \theta_3 = 2.5,\ R = 5)$.

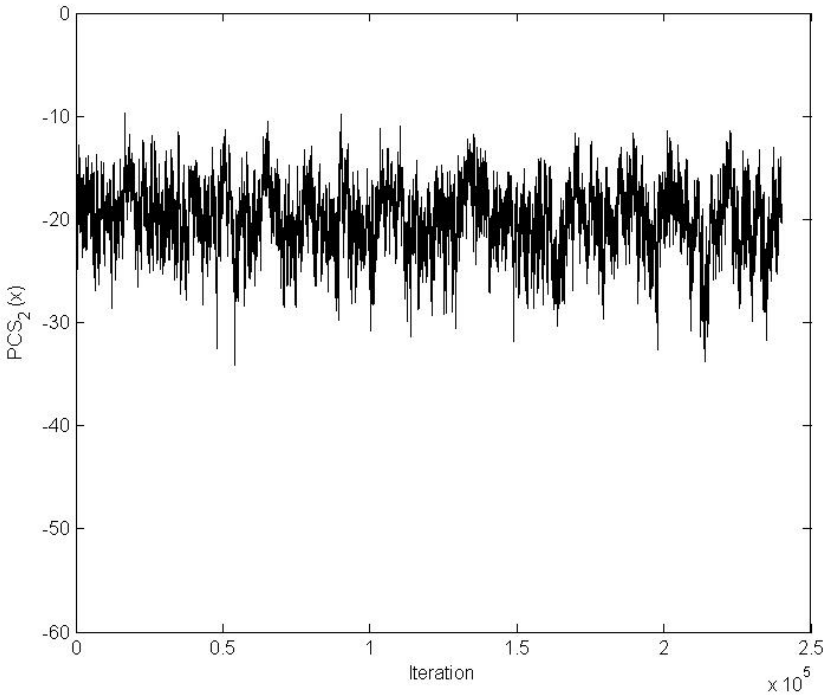

Figure C33. Trace plot of the pseudo-canonical statistic PCS_2 (**x**) for Ex. 3 of the SPTP model with $(\beta = 0.0545,\ \theta_1 = 1,\ \theta_2 = -2.5,\ \theta_3 = 2.5,\ R = 5)$.

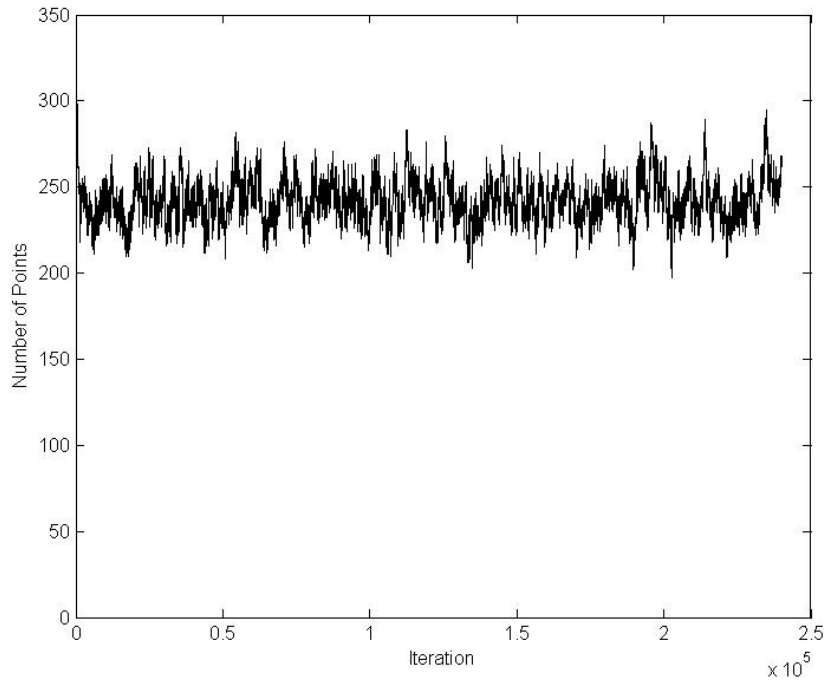

Figure C34. Trace plot of the number of points for Ex. 4 of the SPTP model with $(\beta = 0.0545,\ \theta_1 = 1,\ \theta_2 = -1,\ \theta_3 = 2.5,\ R = 5)$.

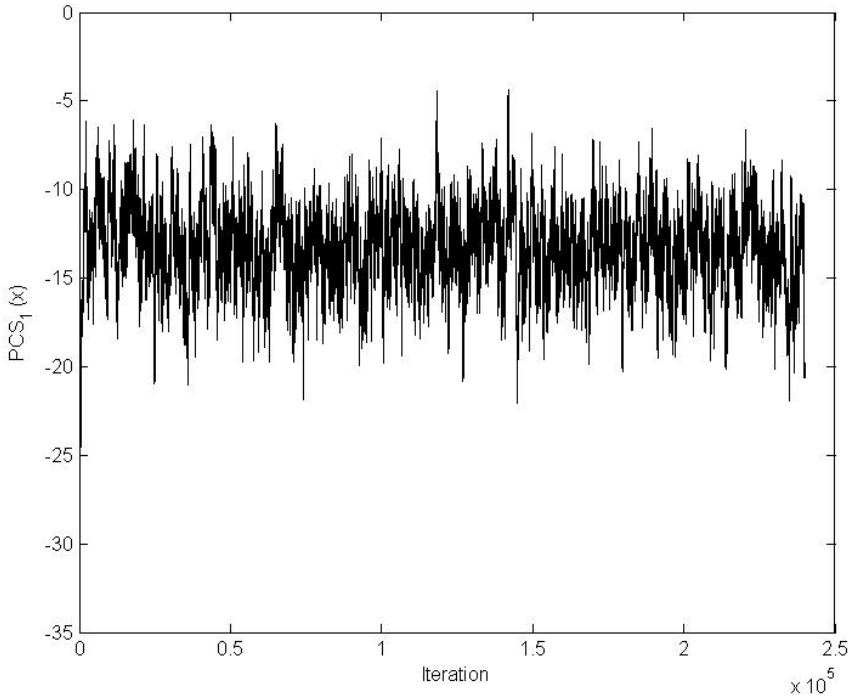

Figure C35. Trace plot of the pseudo-canonical statistic PCS_1 (**x**) for Ex. 4 of the SPTP model with $\left(\beta = 0.0545,\ \theta_1 = 1,\ \theta_2 = -1,\ \theta_3 = 2.5,\ R = 5\right)$.

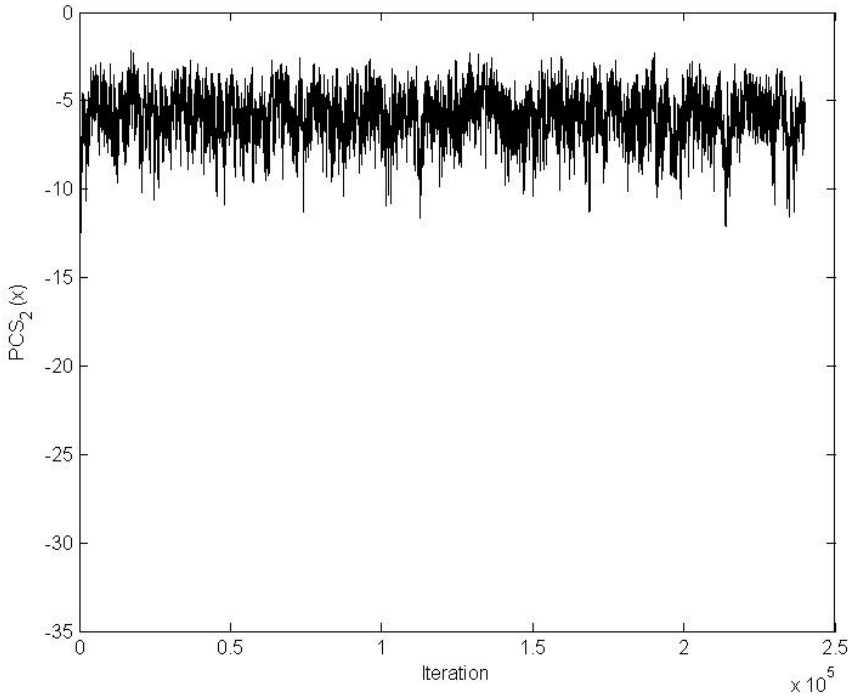

Figure C36. Trace plot of the pseudo-canonical statistic PCS_2 (**x**) for Ex. 4 of the SPTP model with $\left(\beta = 0.0545,\ \theta_1 = 1,\ \theta_2 = -1,\ \theta_3 = 2.5,\ R = 5\right)$.

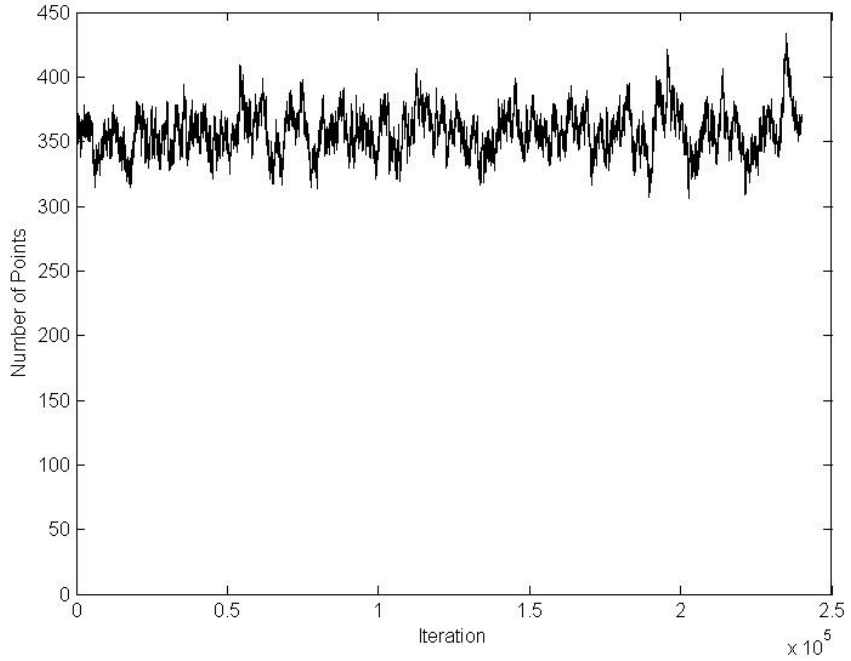

Figure C37. Trace plot of the number of points for Ex. 5 of the SPTP model with $(\beta = 0.0545,\ \theta_1 = -2.5,\ \theta_2 = -1,\ \theta_3 = 2.5,\ R = 5)$.

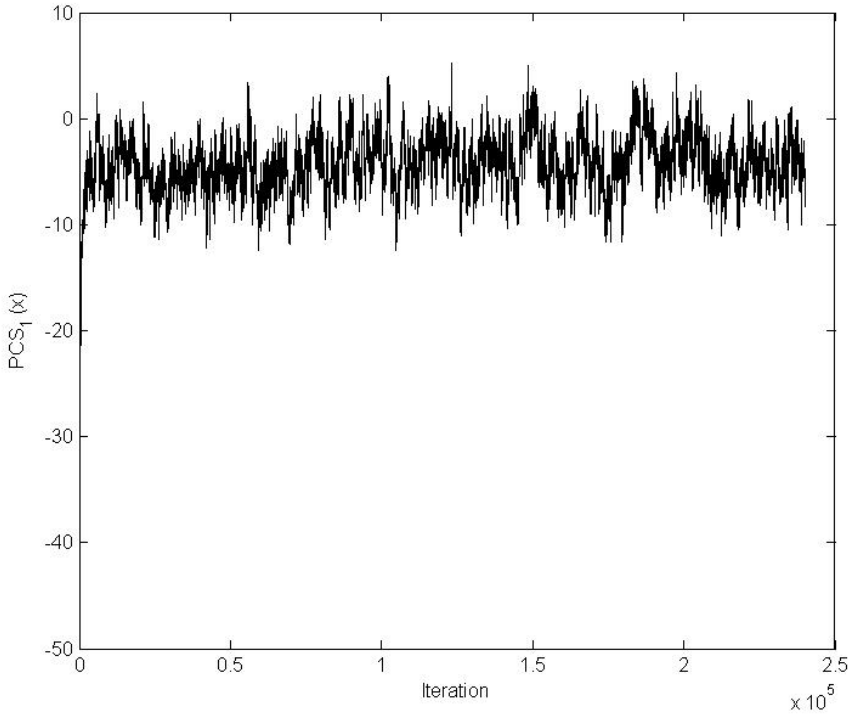

Figure C38. Trace plot of the pseudo-canonical statistic PCS_1 (**x**) for Ex. 5 of the SPTP model with $(\beta = 0.0545,\ \theta_1 = -2.5,\ \theta_2 = -1,\ \theta_3 = 2.5,\ R = 5)$.

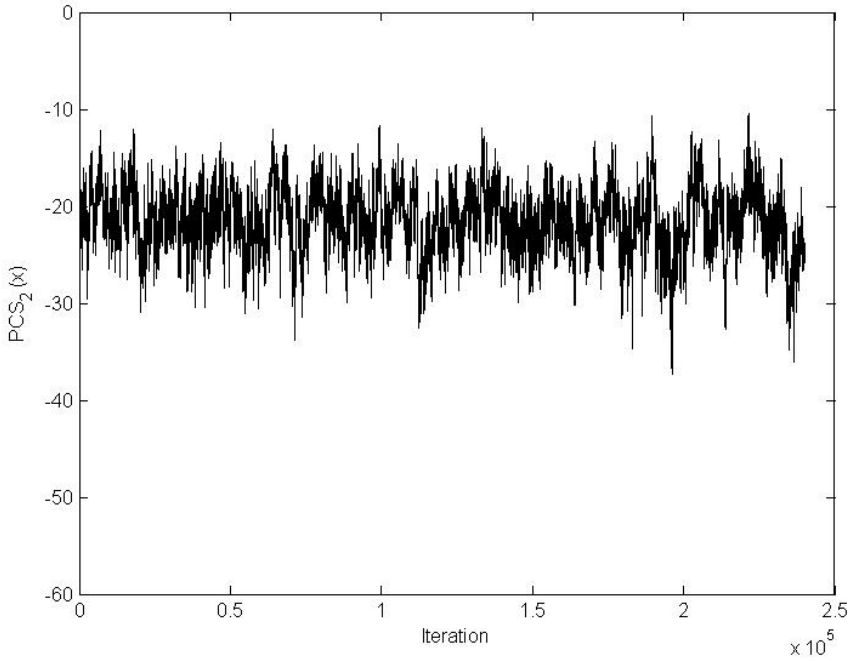

Figure C39. Trace plot of the pseudo-canonical statistic PCS_2 (**x**) for Ex. 5 of the SPTP model with $\left(\beta = 0.0545,\ \theta_1 = -2.5,\ \theta_2 = -1,\ \theta_3 = 2.5,\ R = 5\right)$.

Appendix D

In this appendix are trace plots and plots of marginal vs joint posterior densities for various simulation trials of Bognar's algorithm

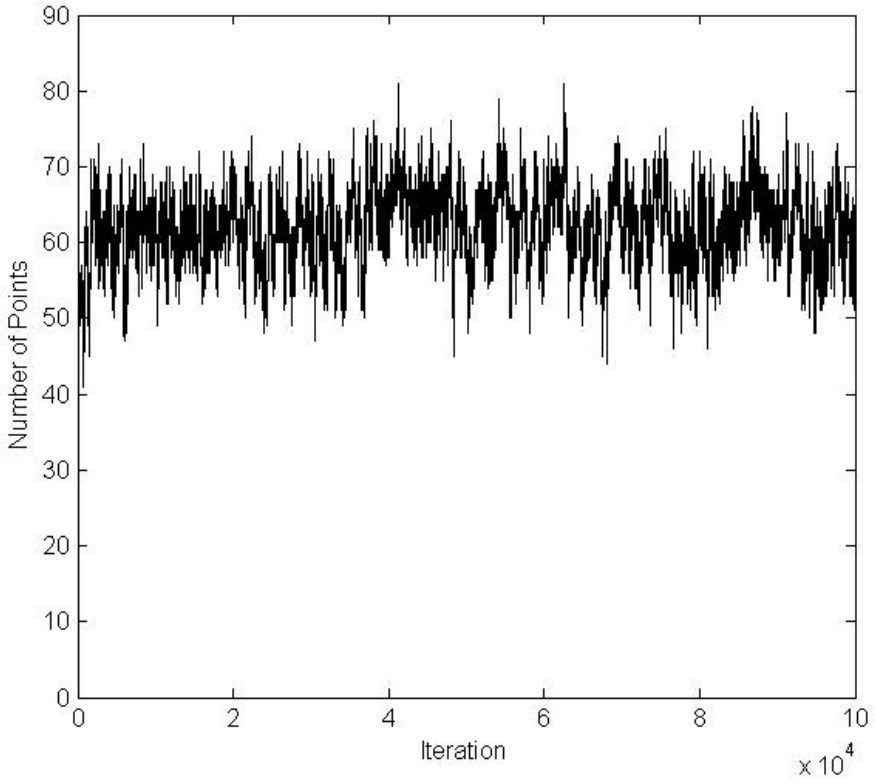

Figure D1. Trace plot of the number of points for the simulation of the EXPTP model with $\left(\beta = 0.0850,\ \theta_1 = -3.5,\ \theta_2 = 2,\ R = 2\right)$.

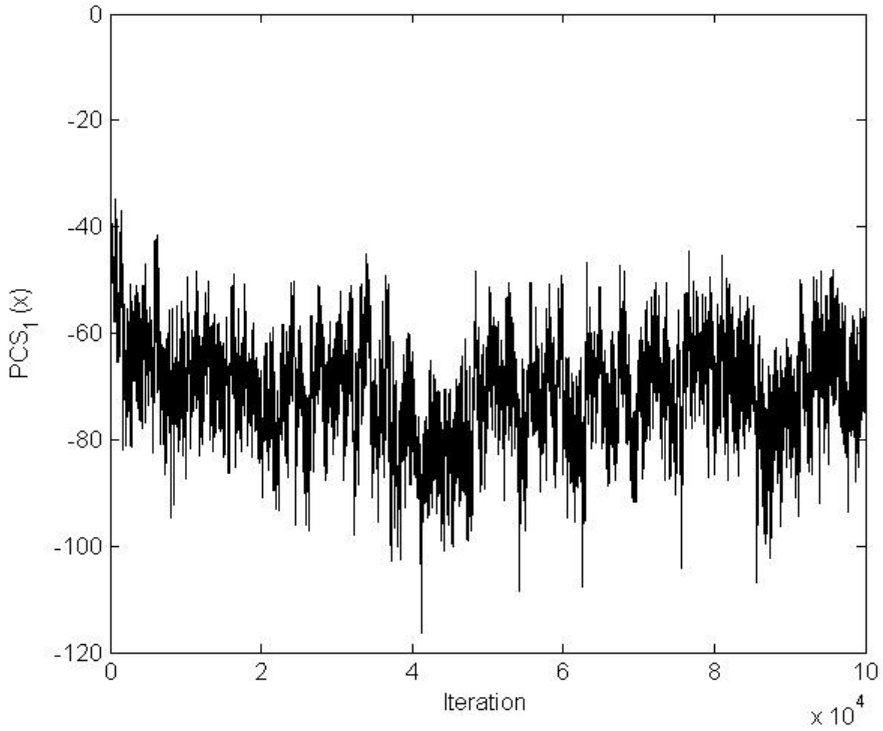

Figure D2. Trace plot of the pseudo-canonical statistic PCS_1 (**x**) for the simulation of the EXPTP model with $(\beta = 0.0850,\ \theta_1 = -3.5,\ \theta_2 = 2,\ R = 2)$.

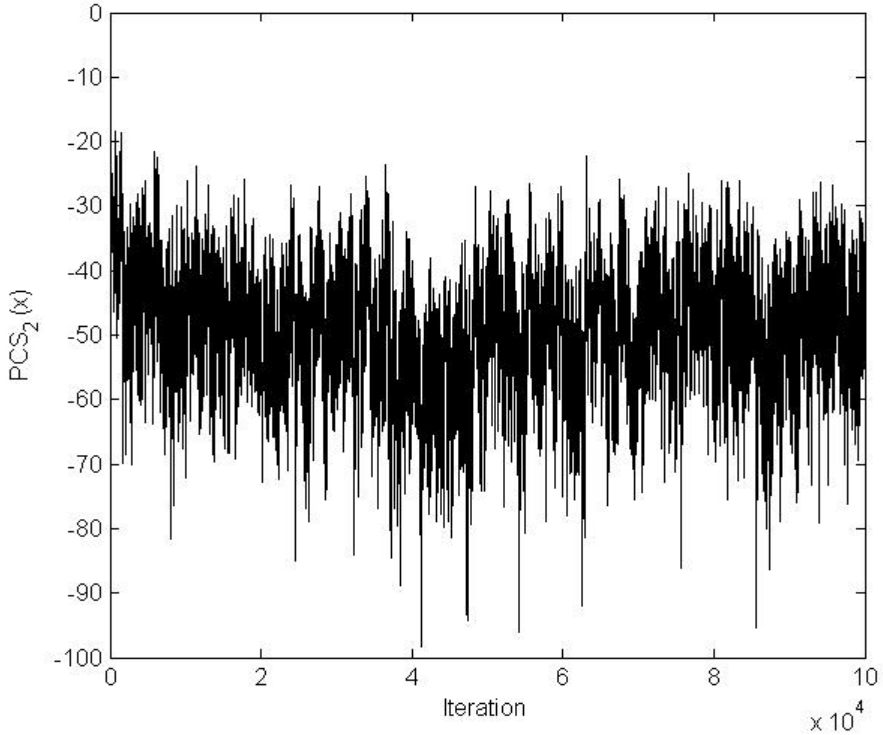

Figure D3. Trace plot of the pseudo-canonical statistic PCS_2 (**x**) for the simulation of the EXPTP model with $(\beta = 0.0850,\ \theta_1 = -3.5,\ \theta_2 = 2,\ R = 2)$.

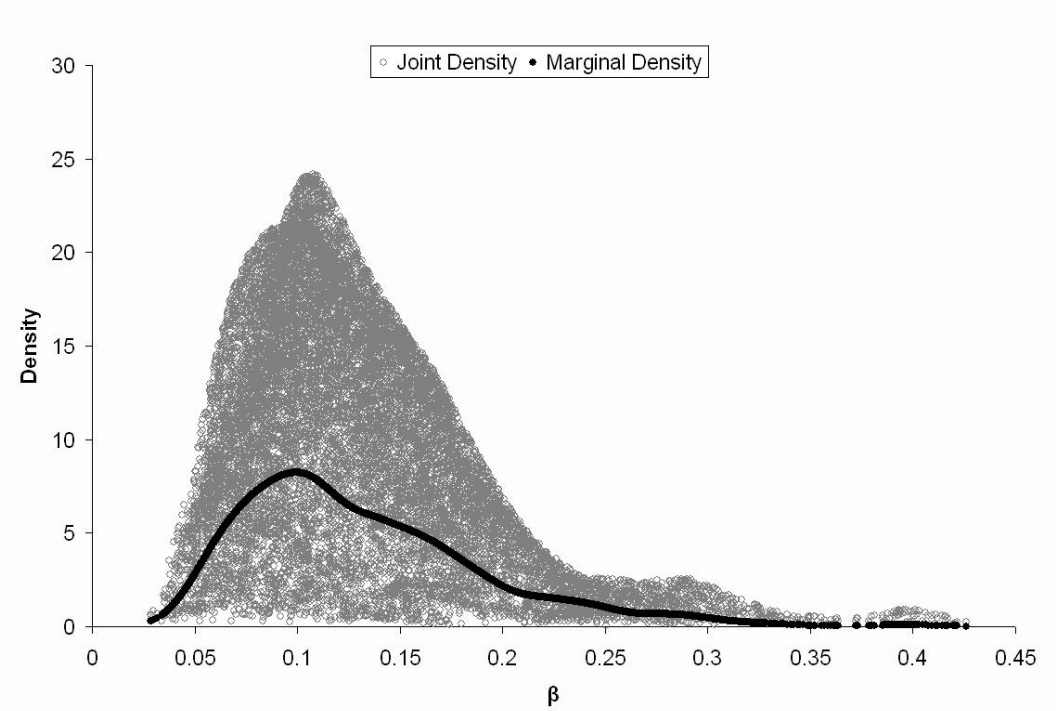

Figure D4. Joint and marginal posterior density estimation for the parameter β of the EXPTP model with true parameters $(\beta = 0.0850,\ \theta_1 = -3.5,\ \theta_2 = 2,\ R = 2)$.

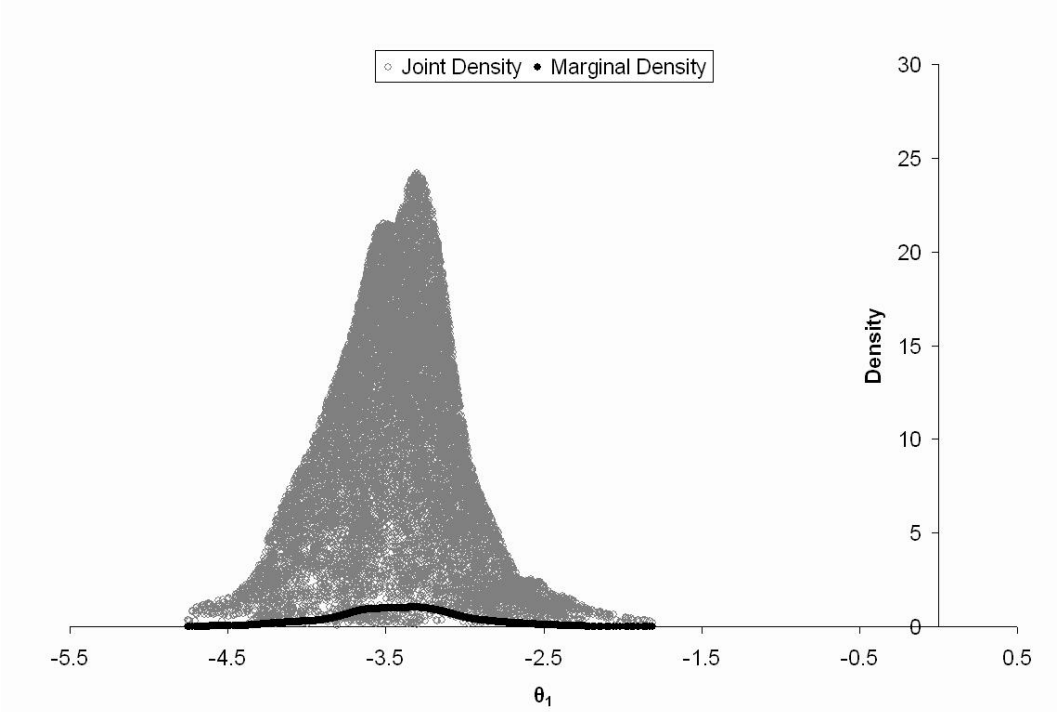

Figure D5. Joint and marginal posterior density estimation for the parameter θ_1 of the EXPTP model with true parameters $(\beta = 0.0850,\ \theta_1 = -3.5,\ \theta_2 = 2,\ R = 2)$.

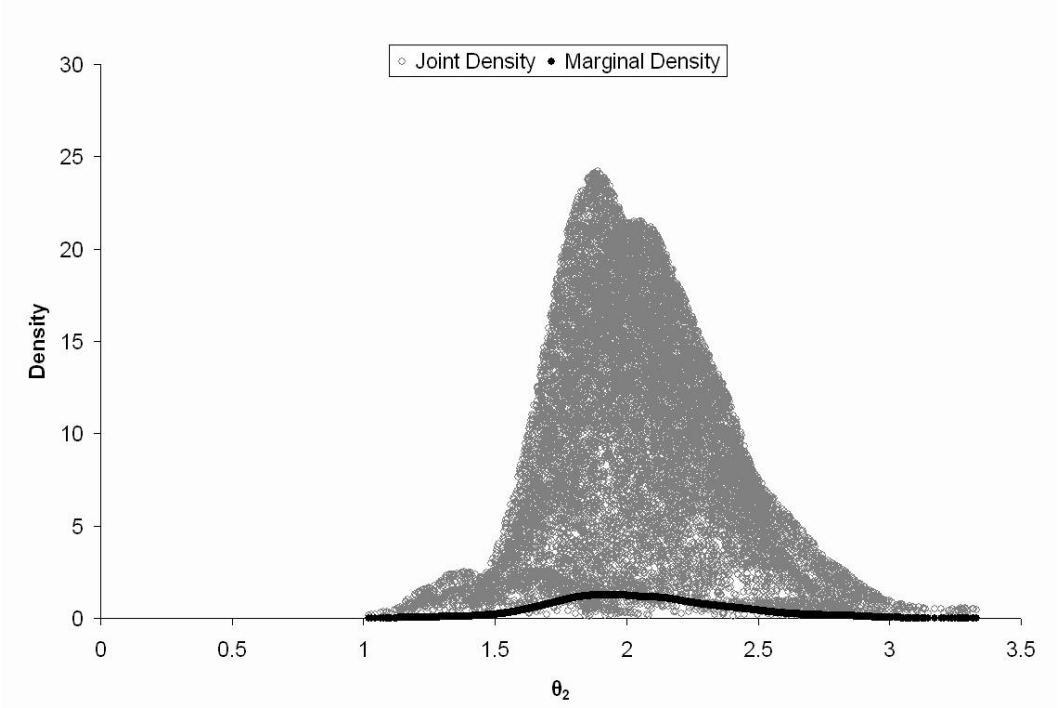

Figure D6. Joint and marginal posterior density estimation for the parameter θ_2 of the EXPTP model with true parameters $(\beta = 0.0850,\ \theta_1 = -3.5,\ \theta_2 = 2,\ R = 2)$.

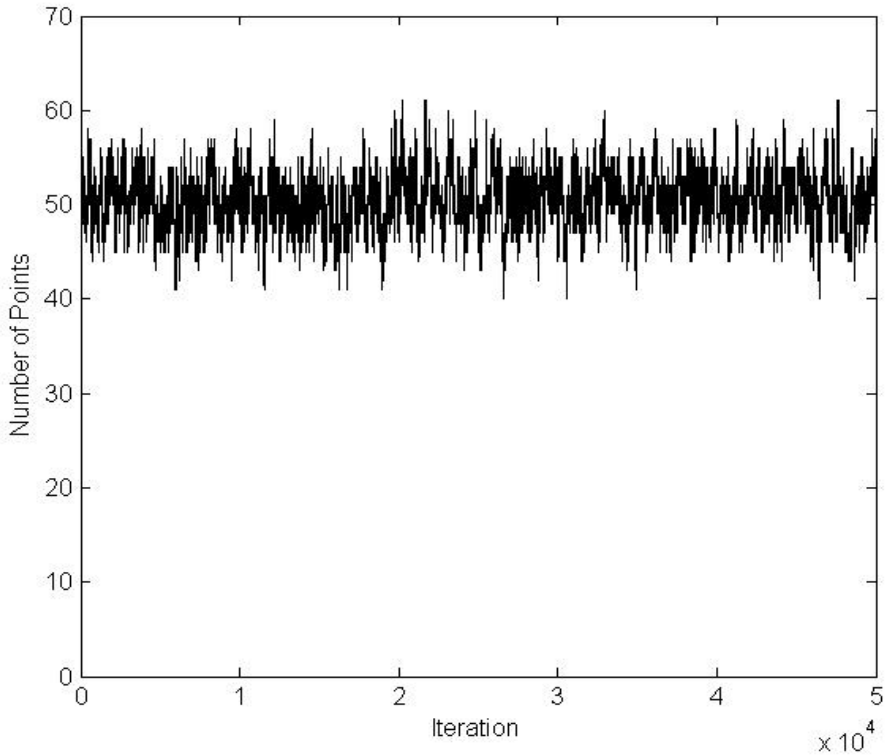

Figure D7. Trace plot of the number of points for the simulation of the EXPTP model with $(\beta = 10,\ \theta_1 = 1,\ \theta_2 = 1,\ R = 2)$.

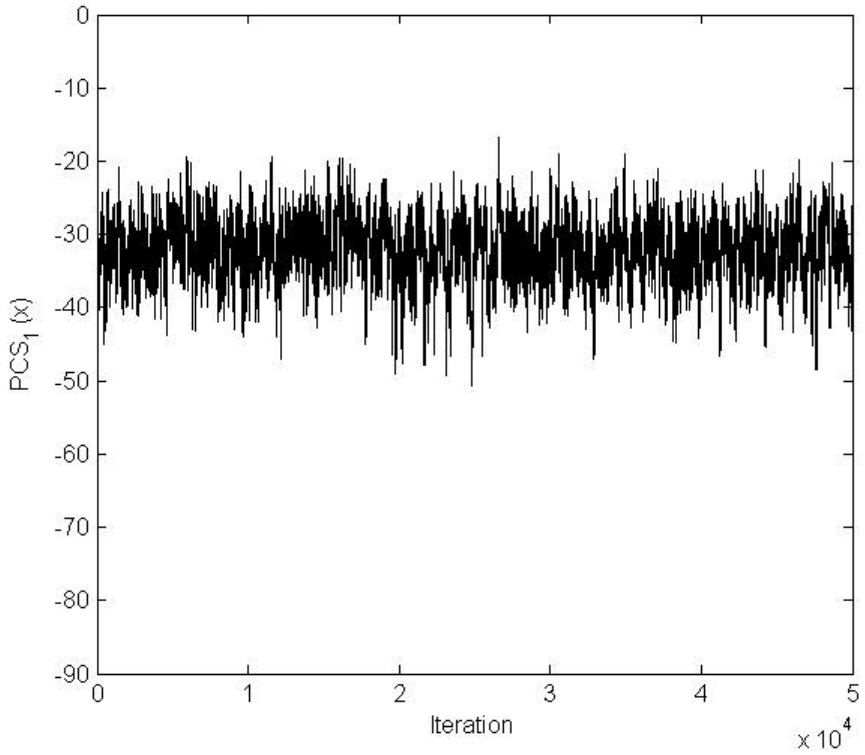

Figure D8. Trace plot of the pseudo-canonical statistic PCS_1 (**x**) for the simulation of the EXPTP model with $(\beta = 10,\ \theta_1 = 1,\ \theta_2 = 1,\ R = 2)$.

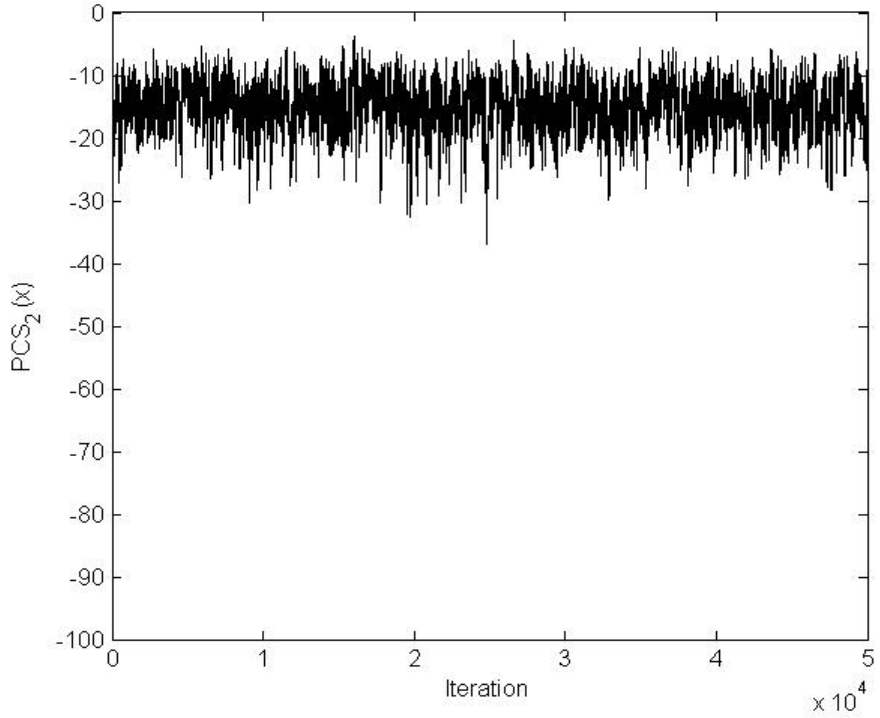

Figure D9. Trace plot of the pseudo-canonical statistic PCS_2 (**x**) for the simulation of the EXPTP model with $(\beta = 10,\ \theta_1 = 1,\ \theta_2 = 1,\ R = 2)$.

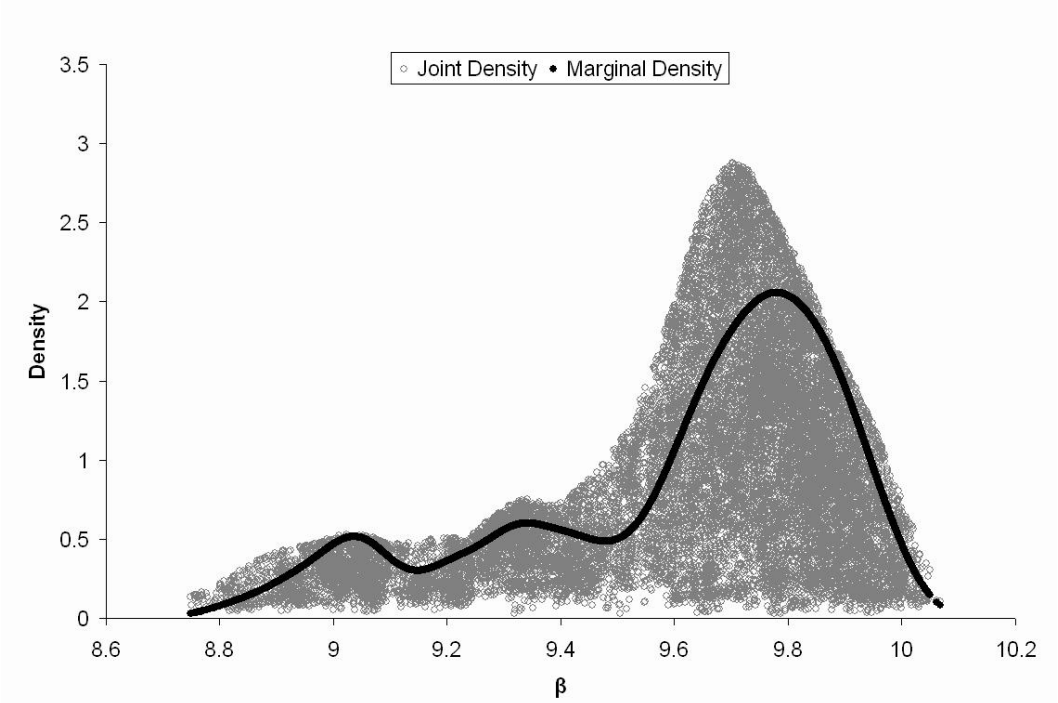

Figure D10. Joint and marginal posterior density estimation for the parameter β of the EXPTP model with true parameters $(\beta = 10,\ \theta_1 = 1,\ \theta_2 = 1,\ R = 2)$.

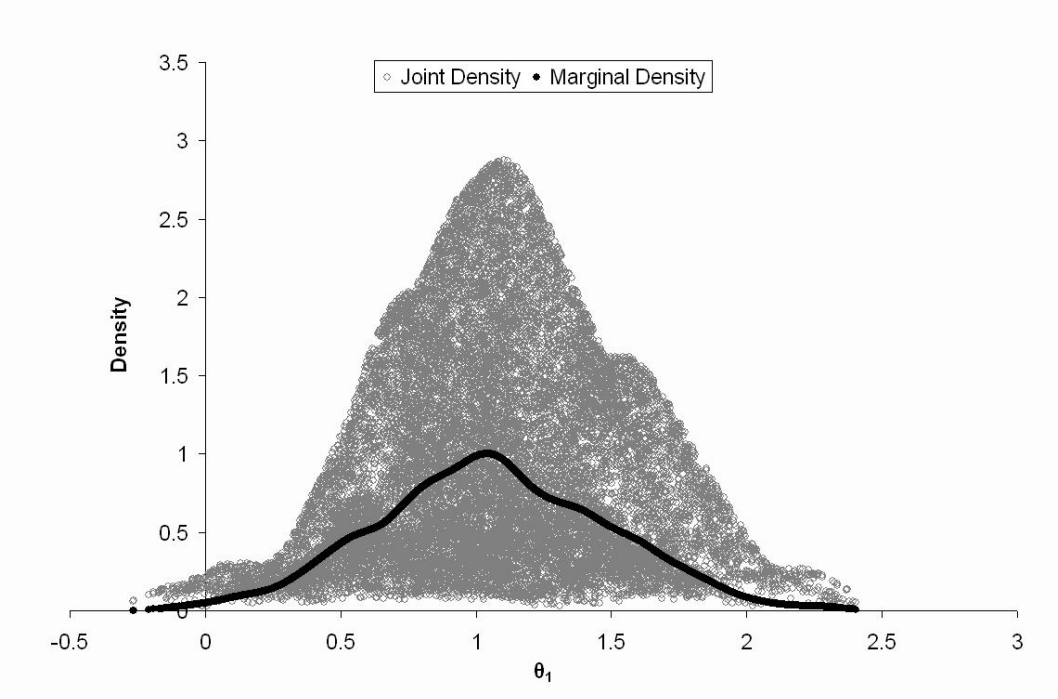

Figure D11. Joint and marginal posterior density estimation for the parameter θ_1 of the EXPTP model with true parameters $(\beta = 10,\ \theta_1 = 1,\ \theta_2 = 1,\ R = 2)$.

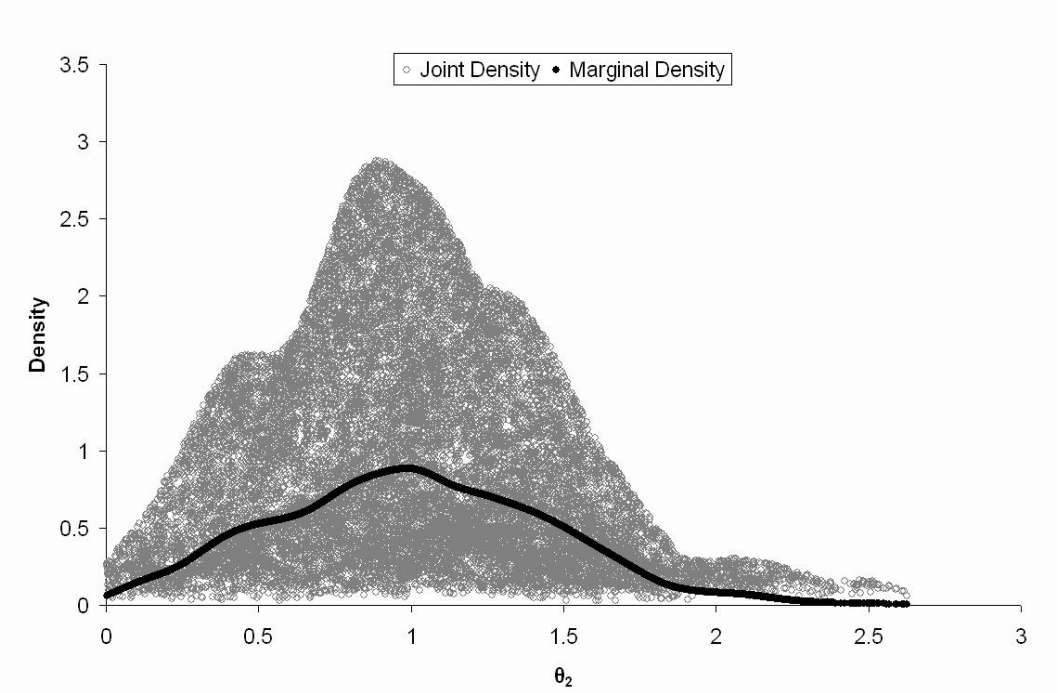

Figure D12. Joint and marginal posterior density estimation for the parameter θ_2 of the EXPTP model with true parameters $\left(\beta = 10,\ \theta_1 = 1,\ \theta_2 = 1,\ R = 2\right)$.

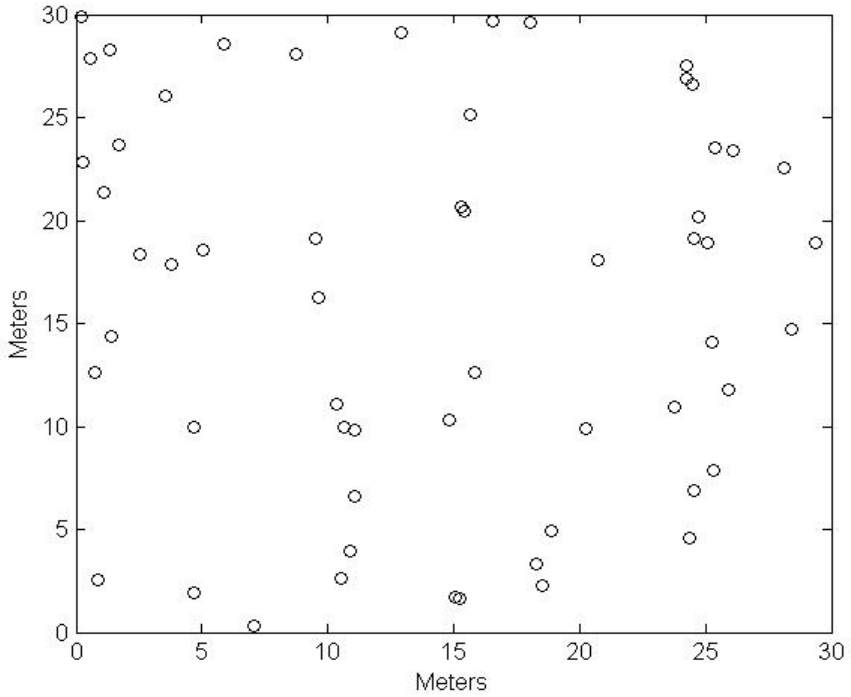

Figure D13. Configuration of simulated EXPTP model in 30 x 30 plot with parameter values $\left(\beta = 0.0625,\ \theta_1 = -2.5,\ \theta_2 = 3.5,\ R = 4\right)$

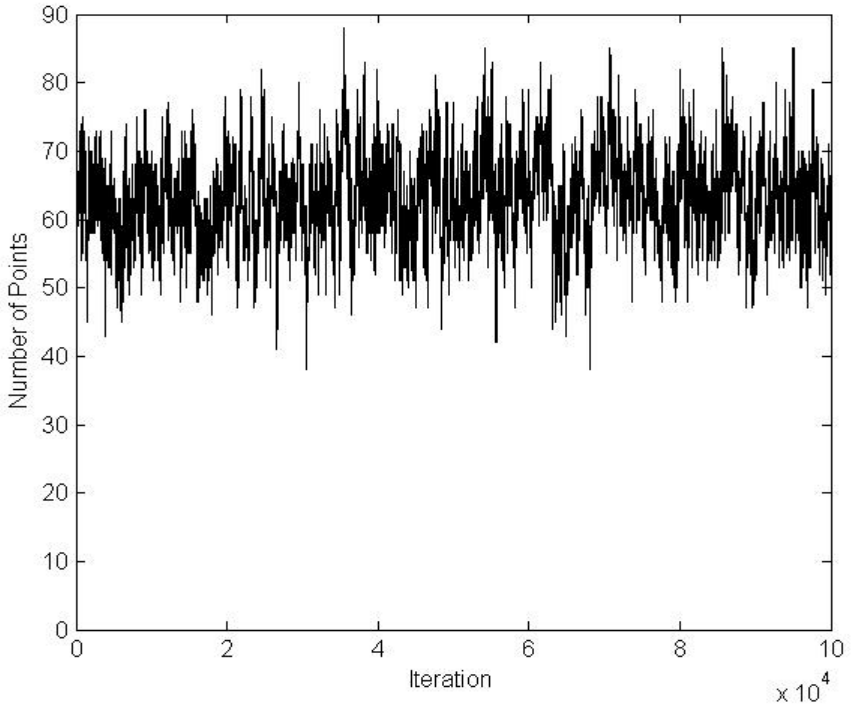

Figure D14. Trace plot of the number of points for the simulation of the EXPTP model with $\left(\beta = 0.0625,\ \theta_1 = -2.5,\ \theta_2 = 3.5,\ R = 4\right)$.

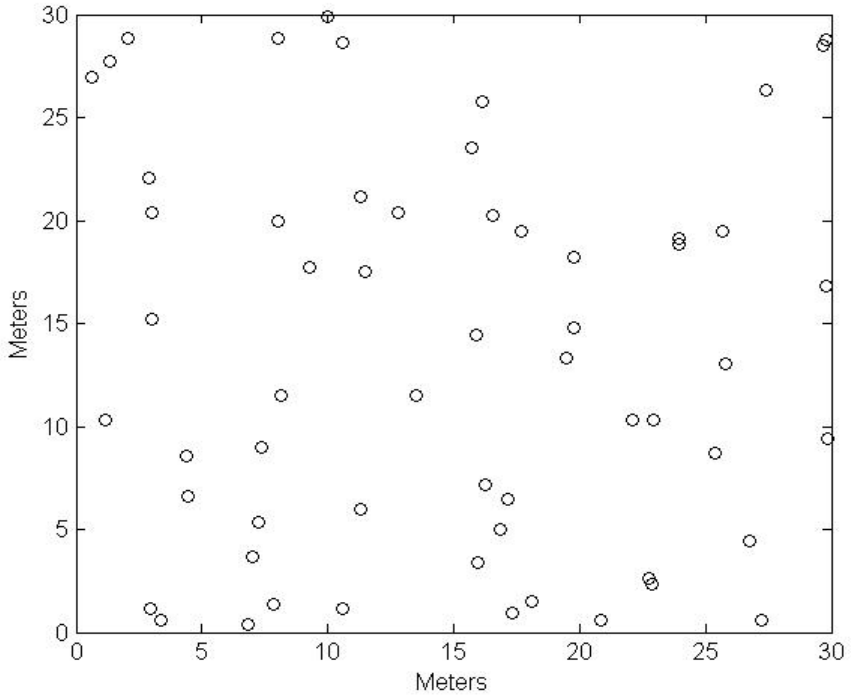

Figure D15. Configuration of simulated EXPTP model in 30 x 30 plot with parameter values $(\beta = 0.0625,\ \theta_1 = -2.5,\ \theta_2 = 3.5,\ R = 6)$

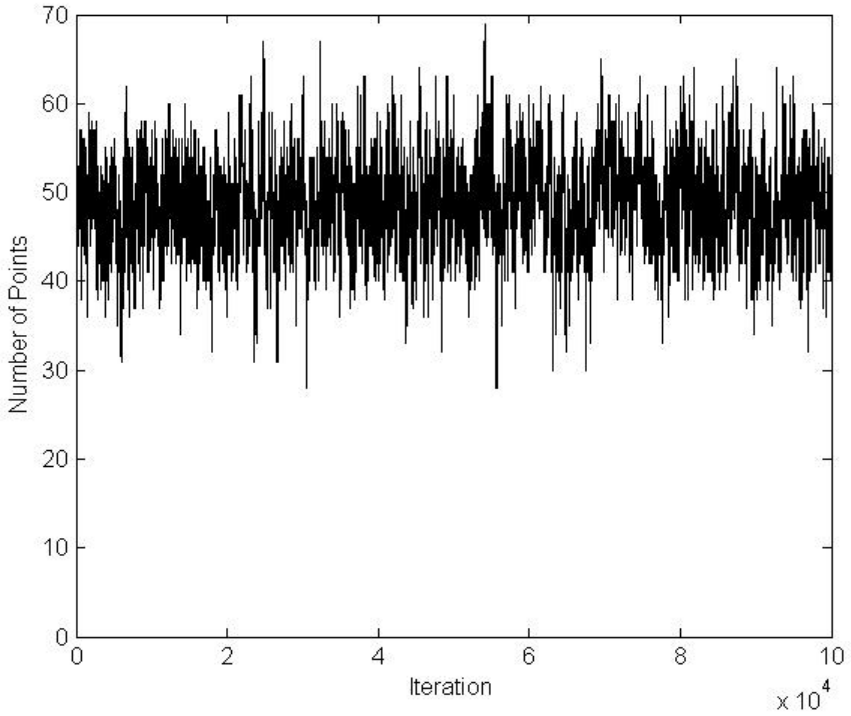

Figure D16. Trace plot of the number of points for the simulation of the EXPTP model with $(\beta = 0.0625,\ \theta_1 = -2.5,\ \theta_2 = 3.5,\ R = 6)$.

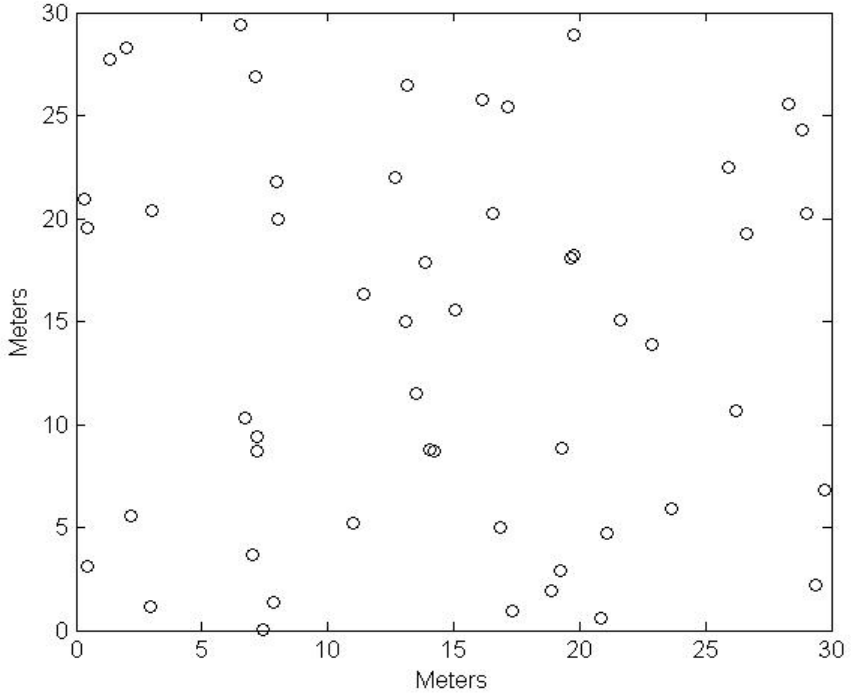

Figure D17. Configuration of simulated EXPTP model in 30 x 30 plot with parameter values $(\beta = 0.0625,\ \theta_1 = -2.5,\ \theta_2 = 3.5,\ R = 8)$

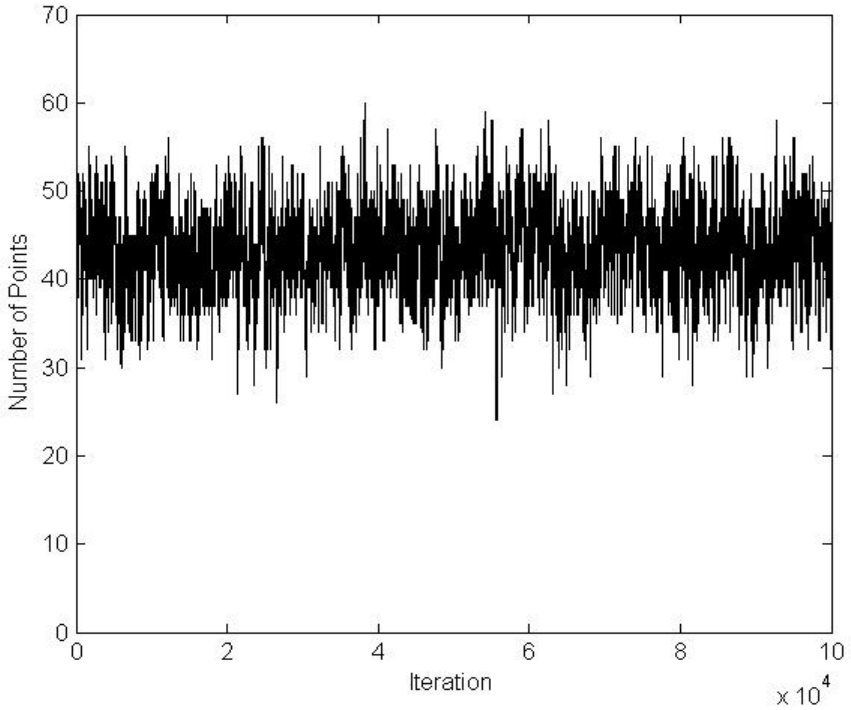

Figure D18. Trace plot of the number of points for the simulation of the EXPTP model with $(\beta = 0.0625,\ \theta_1 = -2.5,\ \theta_2 = 3.5,\ R = 8)$.

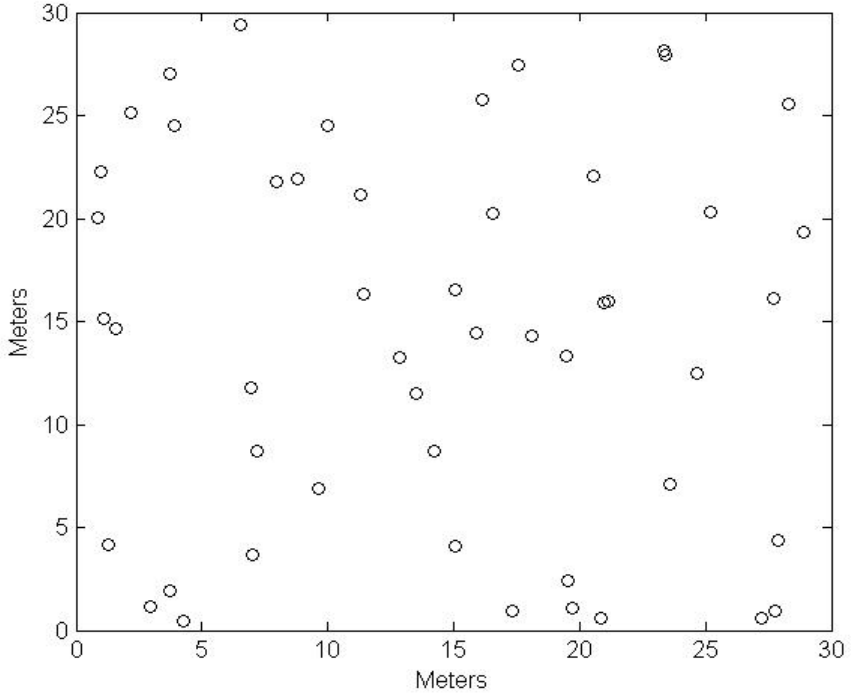

Figure D19. Configuration of simulated EXPTP model in 30 x 30 plot with parameter values $(\beta = 0.0625,\ \theta_1 = -2.5,\ \theta_2 = 3.5,\ R = 10)$

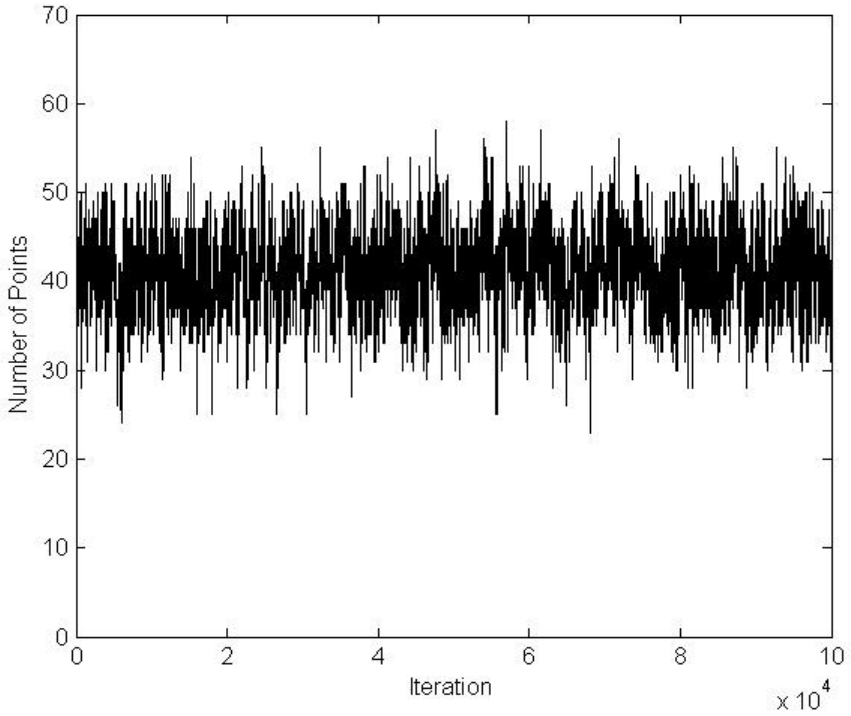

Figure D20. Trace plot of the number of points for the simulation of the EXPTP model with $(\beta = 0.0625,\ \theta_1 = -2.5,\ \theta_2 = 3.5,\ R = 10)$.

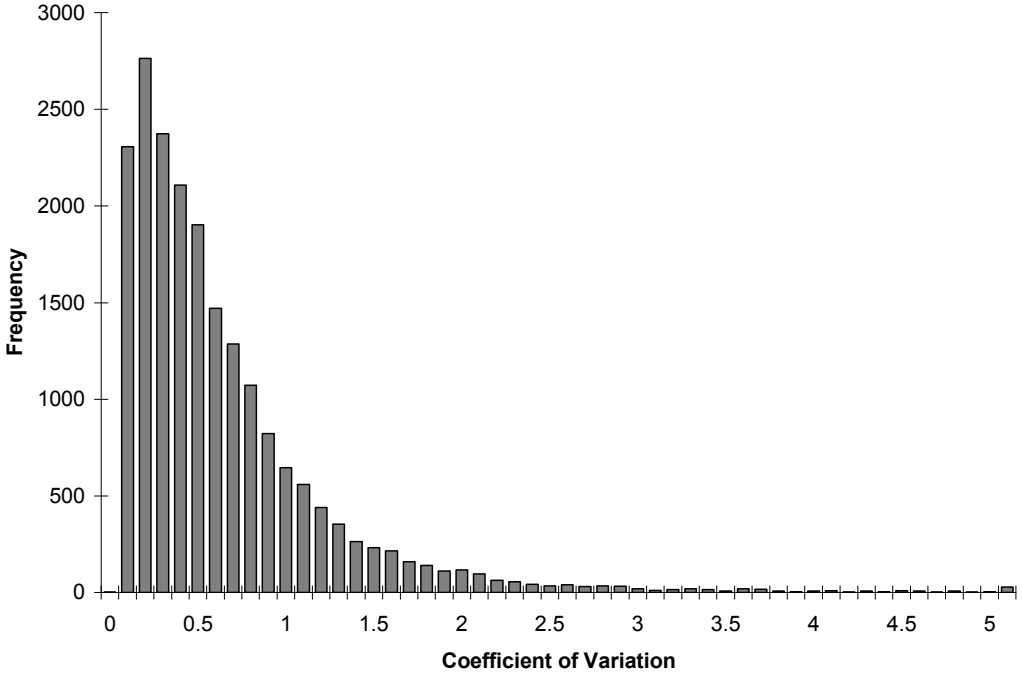

Figure D21. Histogram of the coefficient of variation of the importance sampling weights from Bognar's algorithm on the 4m interaction distance simulated data. Note: a CV of 1.0 corresponds to 100%.

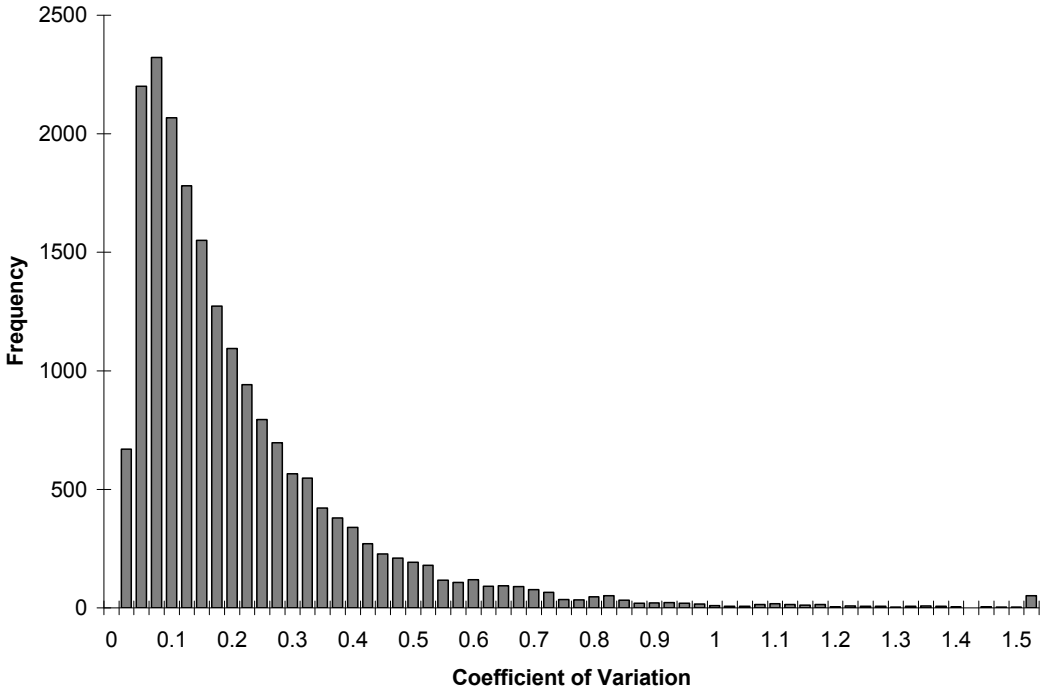

Figure D22. Histogram of the coefficient of variation of the importance sampling weights from Bognar's algorithm on the 6m interaction distance simulated data. Note: a CV of 1.0 corresponds to 100%.

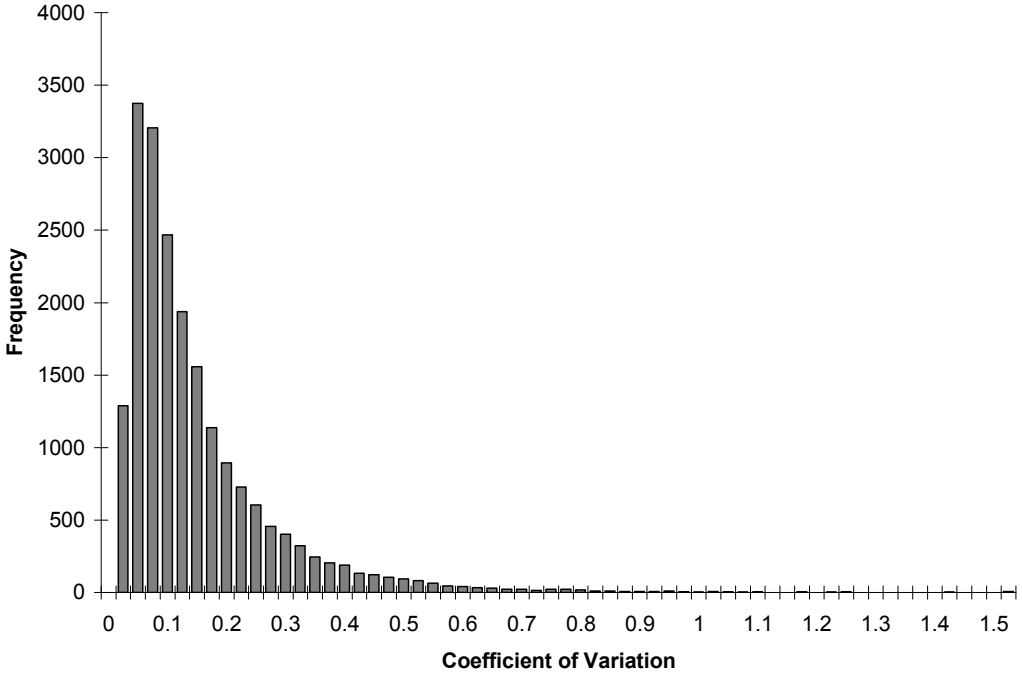

Figure D23. Histogram of the coefficient of variation of the importance sampling weights from Bognar's algorithm on the 8m interaction distance simulated data. Note: a CV of 1.0 corresponds to 100%.

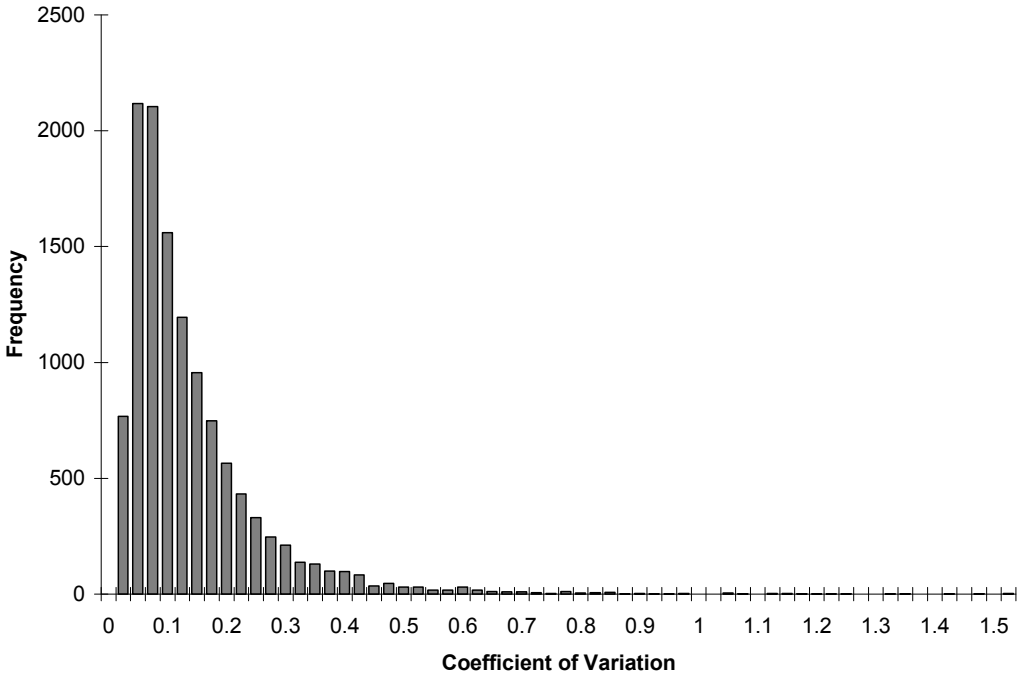

Figure D24. Histogram of the coefficient of variation of the importance sampling weights from Bognar's algorithm on the 10m interaction distance simulated data. Note: a CV of 1.0 corresponds to 100%.

Appendix E

In this appendix are trace plots and plots of marginal vs. joint posterior densities for Bayesian inference of the center 30 x 30m of Coweeta Plot 318 with Bognar's algorithm.

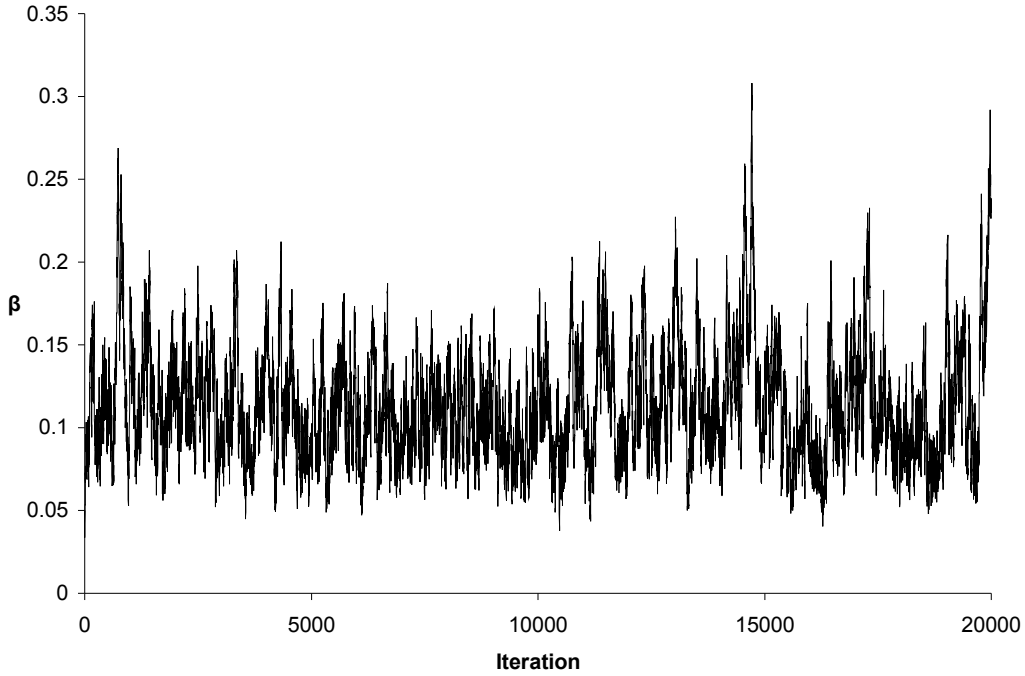

Figure E1. Trace plot of the parameter β of the EXPTP model for the center 30 x 30m of Coweeta Plot 318.

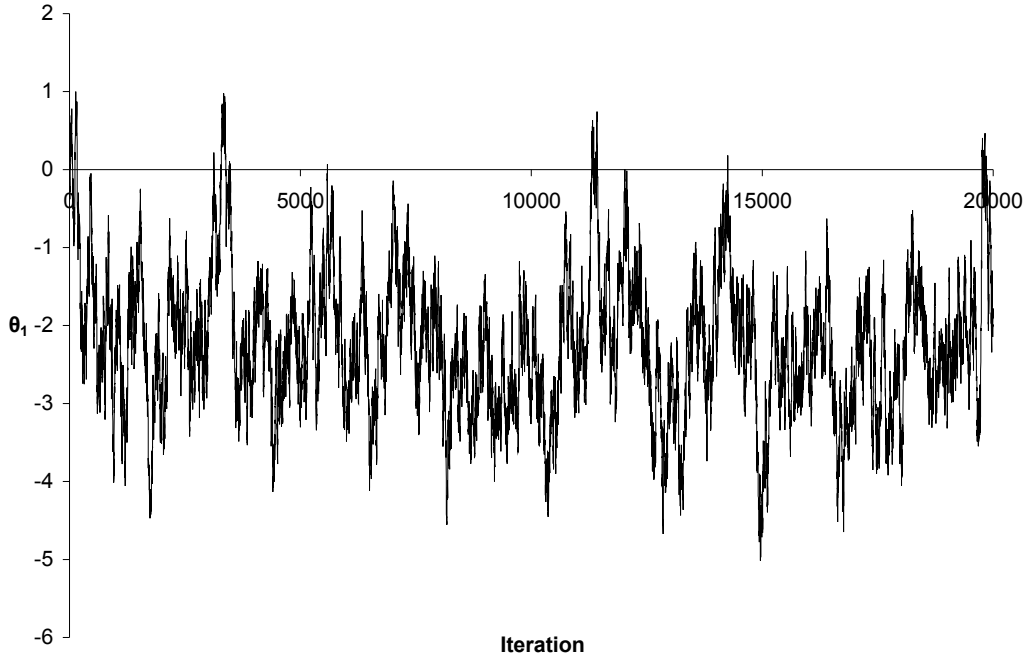

Figure E2. Trace plot of the parameter θ_1 of the EXPTP model for the center 30 x 30m of Coweeta Plot 318.

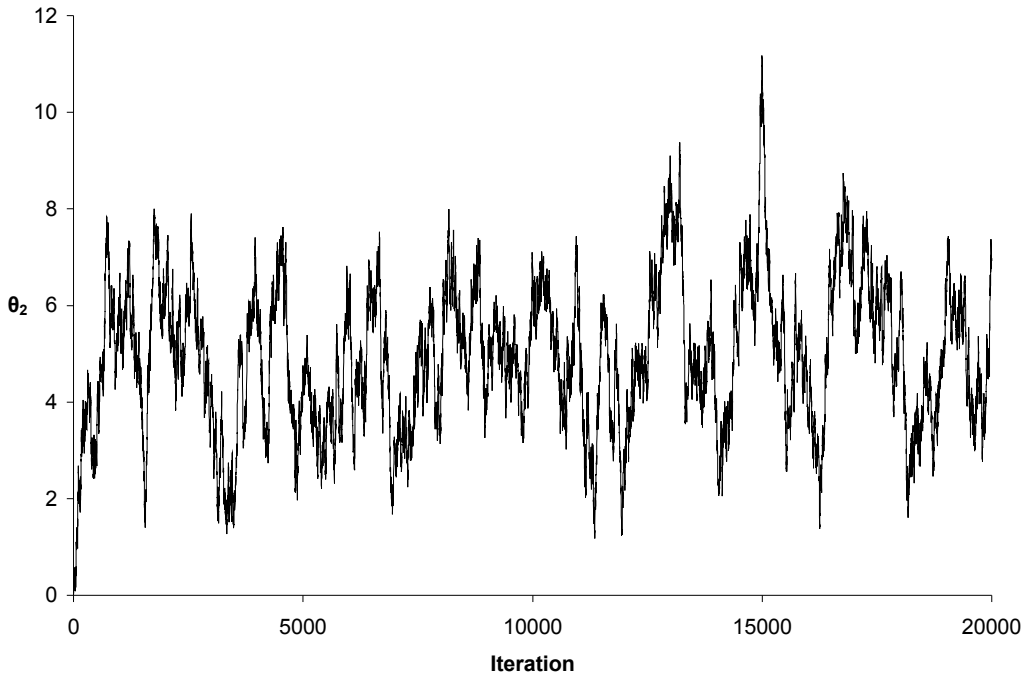

Figure E3. Trace plot of the parameter θ_2 of the EXPTP model for the center 30 x 30m of Coweeta Plot 318.

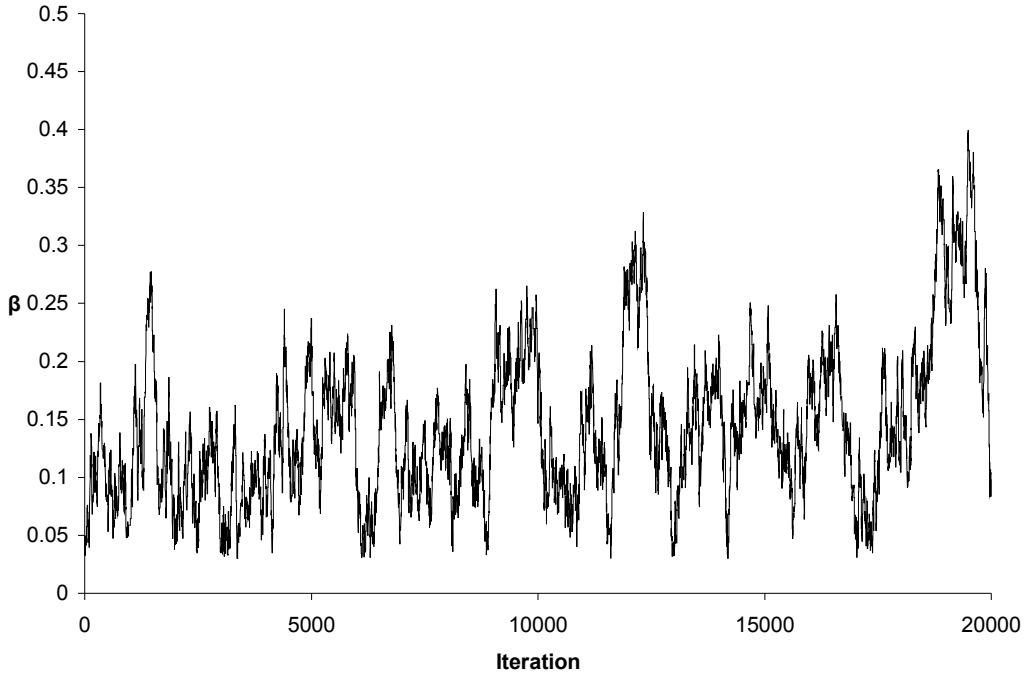

Figure E4. Trace plot of the parameter β of the GTP model for the center 30 x 30m of Coweeta Plot 318.

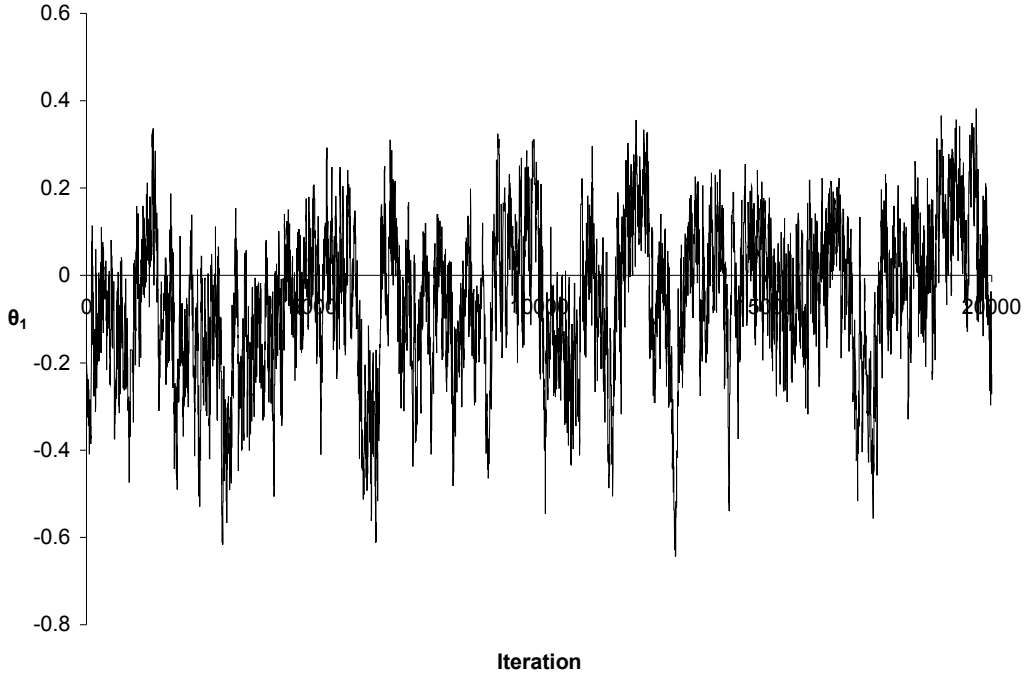

Figure E5. Trace plot of the parameter θ_1 of the GTP model for the center 30 x 30m of Coweeta Plot 318.

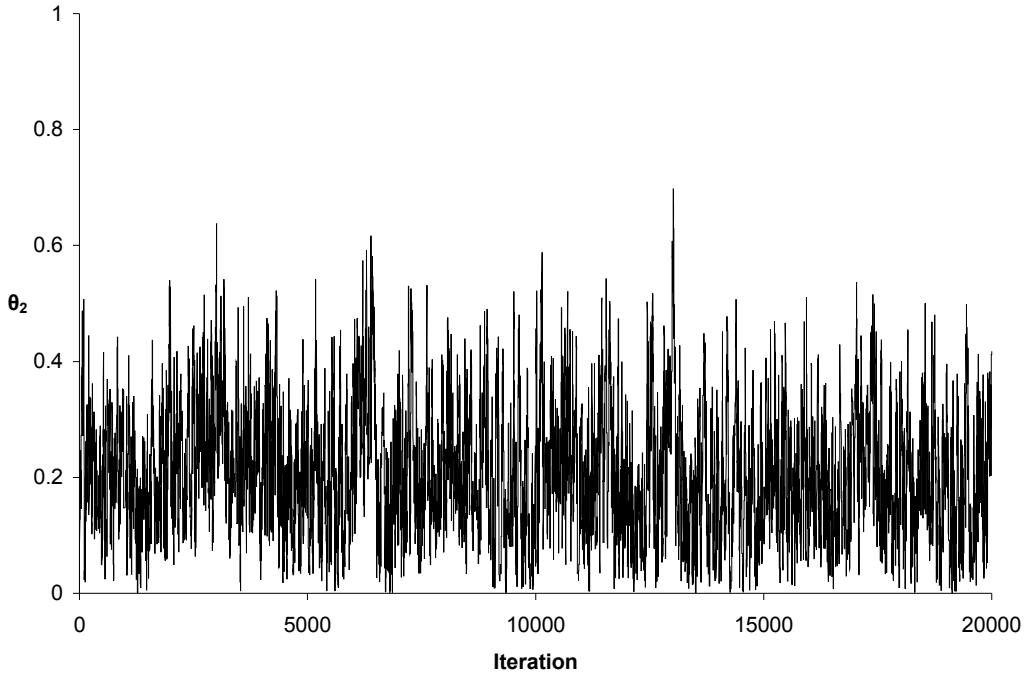

Figure E6. Trace plot of the parameter θ_2 of the GTP model for for the center 30 x 30m of Coweeta Plot 318.

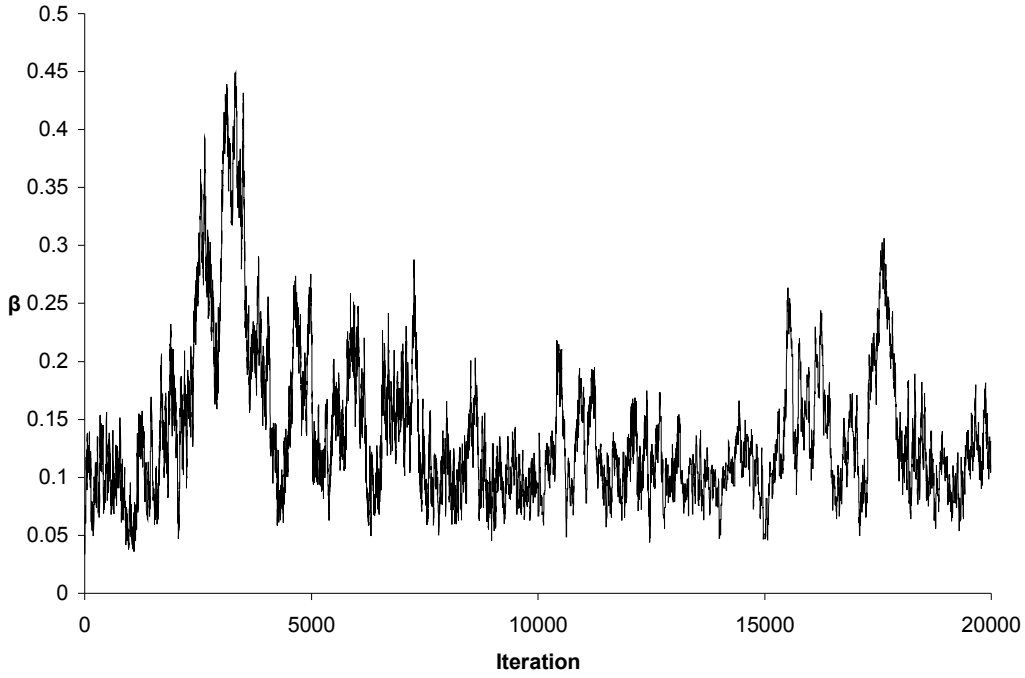

Figure E7. Trace plot of the parameter β of the SPTP model for the center 30 x 30m of Coweeta Plot 318.

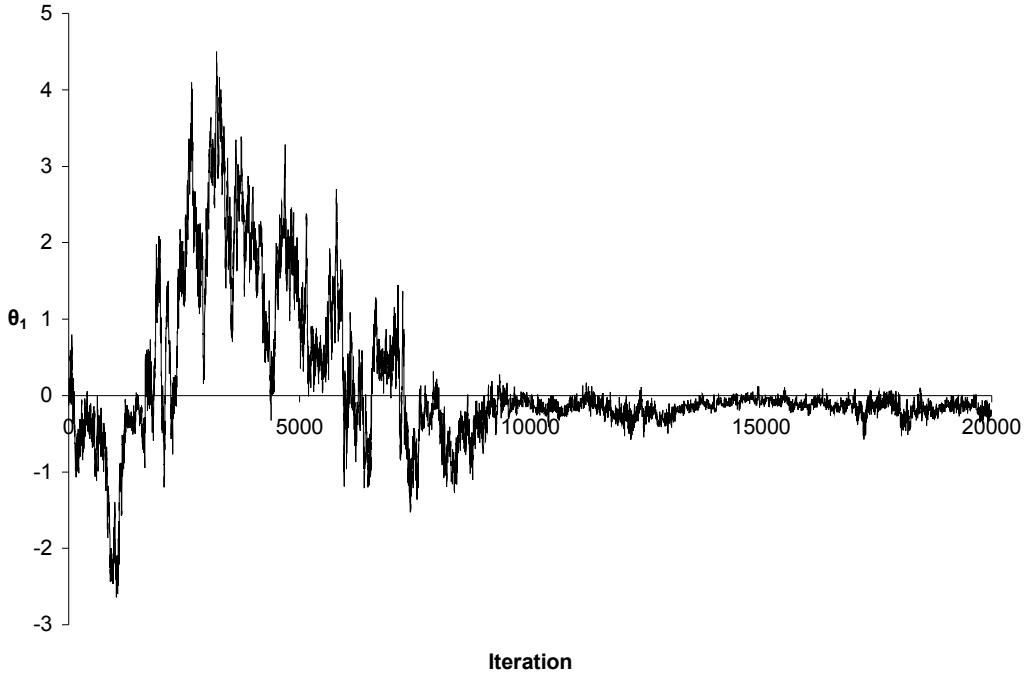

Figure E8. Trace plot of the parameter θ_1 of the SPTP model for the center 30 x 30m of Coweeta Plot 318.

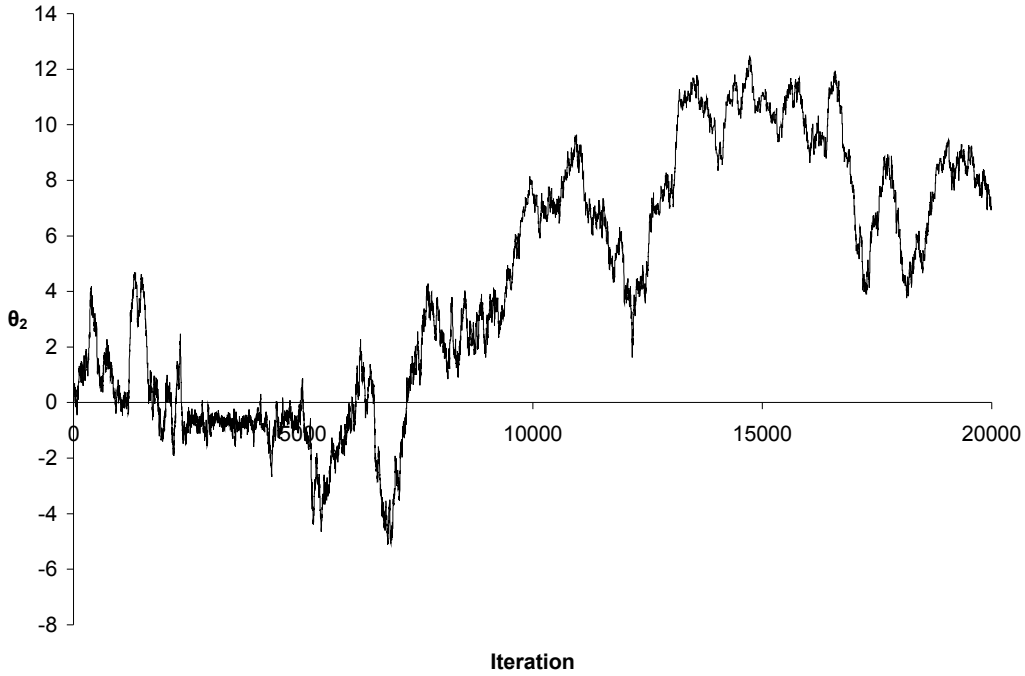

Figure E9. Trace plot of the parameter θ_2 of the SPTP model for the center 30 x 30m of Coweeta Plot 318.

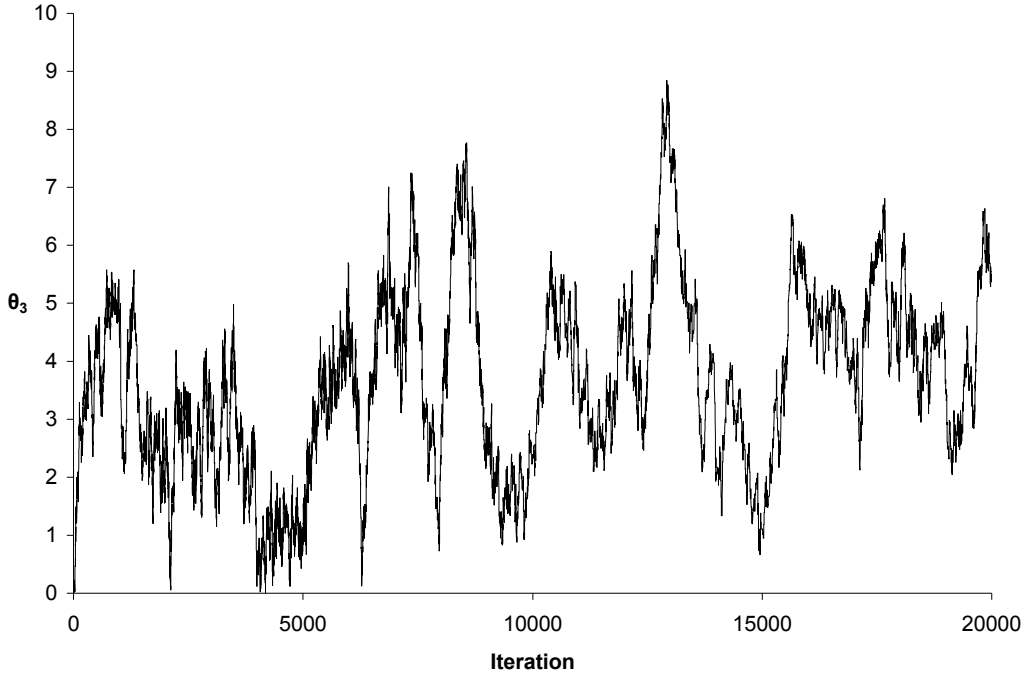

Figure E10. Trace plot of the parameter θ_3 of the SPTP model for the center 30 x 30m of Coweeta Plot 318.

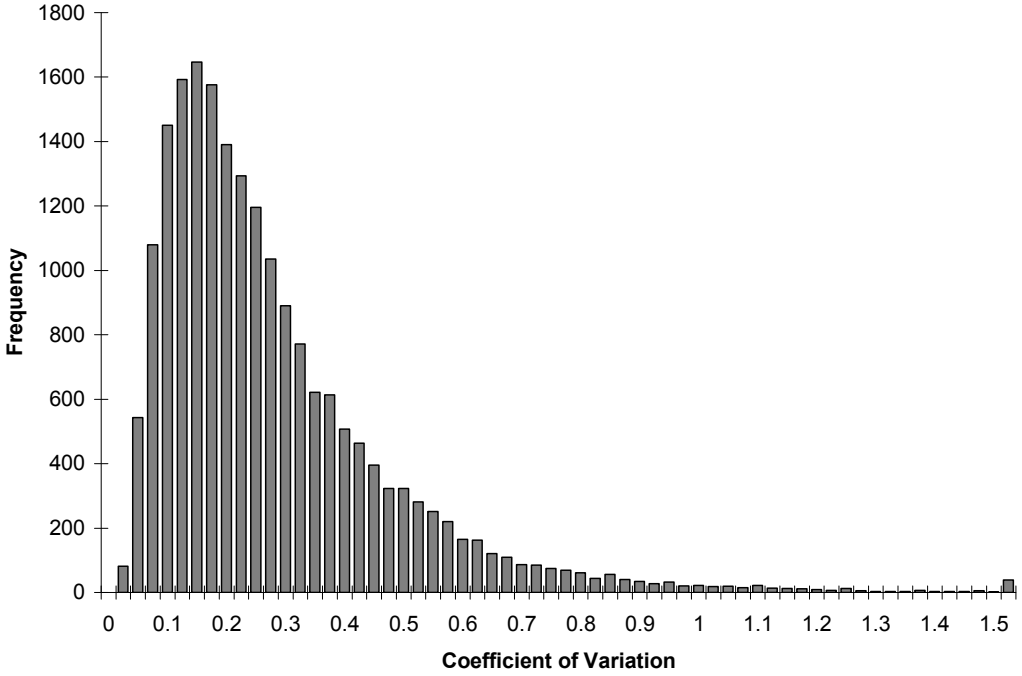

Figure E11. Histogram of the coefficient of variation of the importance sampling weights for the EXPTP model on the center 30 x 30m of Coweeta Plot 318. Note: a CV of 1.0 corresponds to 100%.

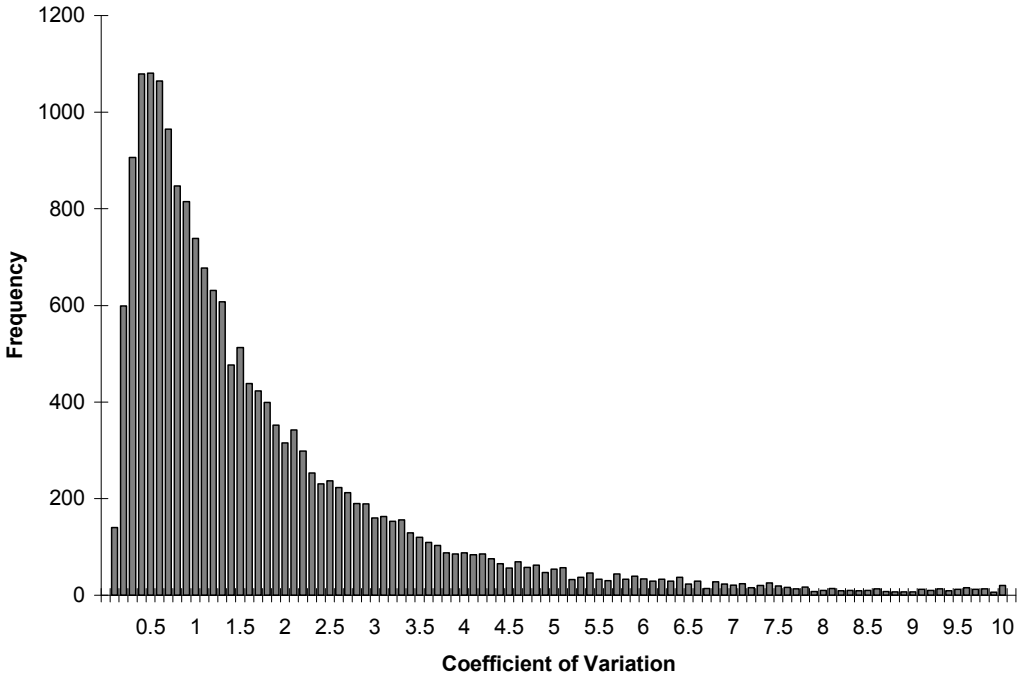

Figure E12. Histogram of the coefficient of variation of the importance sampling weights for the GTP model on the center 30 x 30m of Coweeta Plot 318. Note: a CV of 1.0 corresponds to 100%.

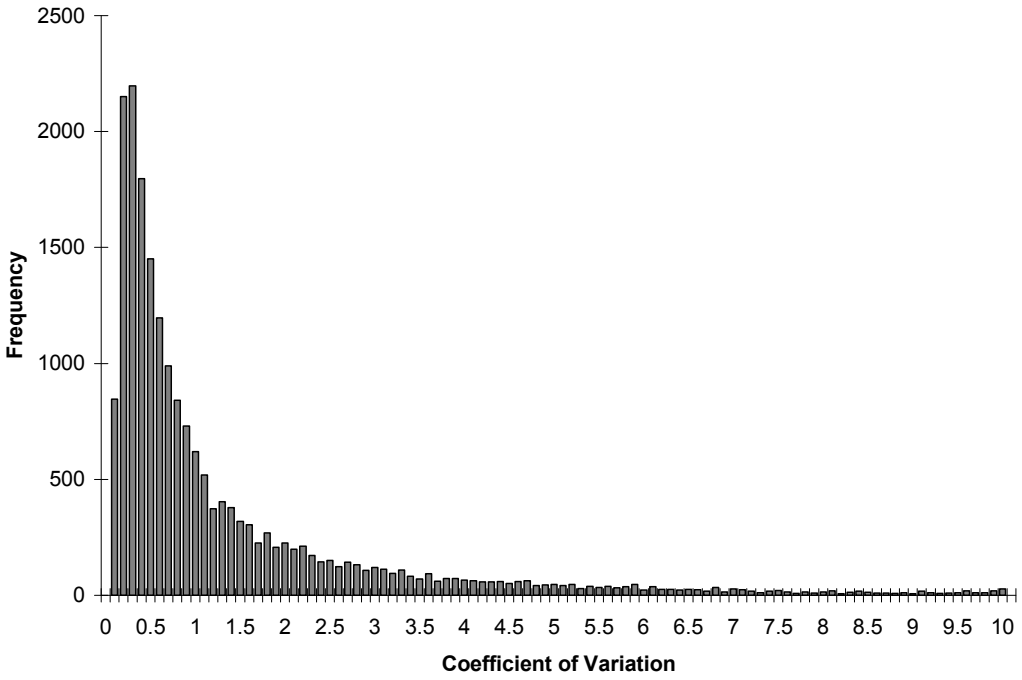

Figure E13. Histogram of the coefficient of variation of the importance sampling weights for the SPTP model on the center 30 x 30m of Coweeta Plot 318. Note: a CV of 1.0 corresponds to 100%.

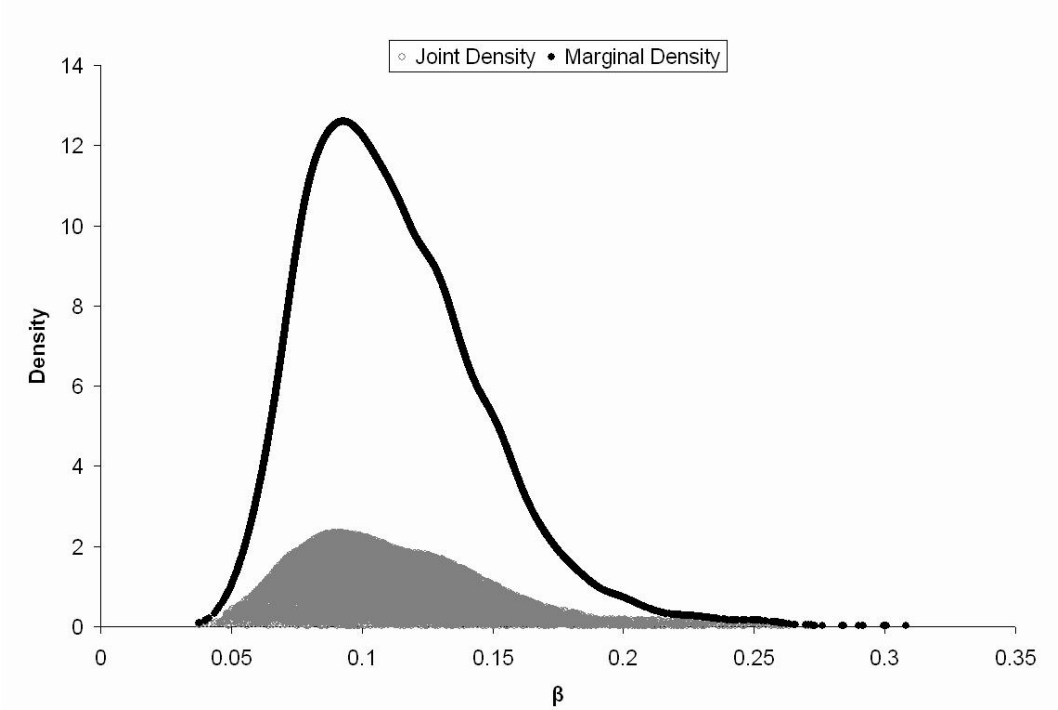

Figure E14. Joint and marginal posterior density estimation for the parameter β of the EXPTP model for the center 30 x 30m of Coweeta Plot 318.

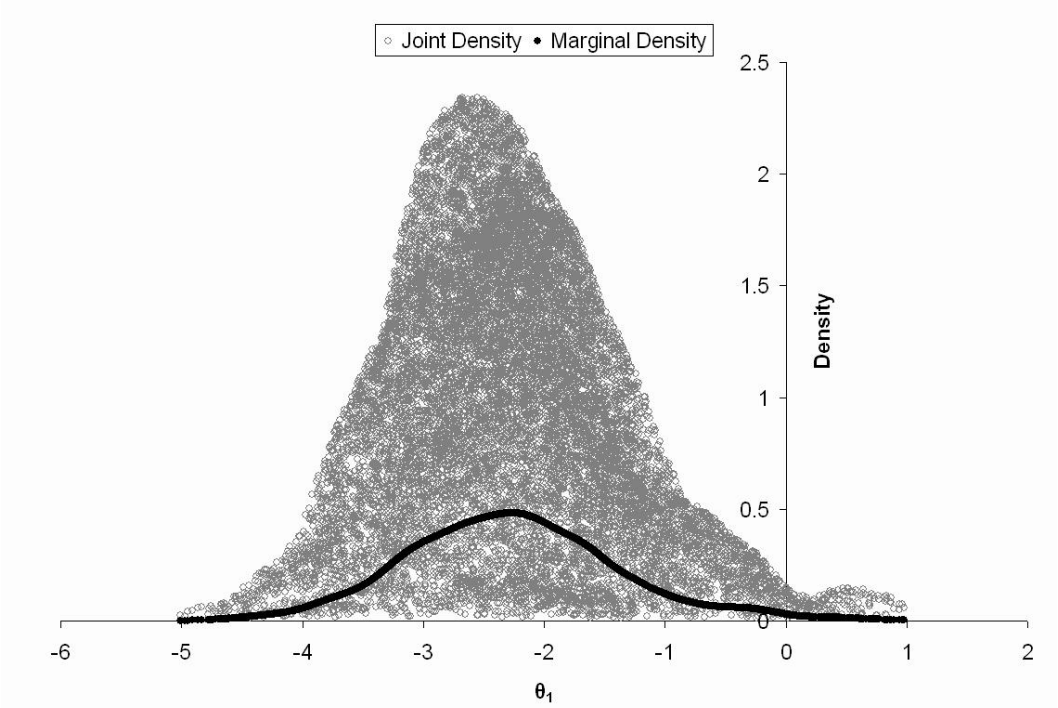

Figure E15. Joint and marginal posterior density estimation for the parameter θ_1 of the EXPTP model for the center 30 x 30m of Coweeta Plot 318.

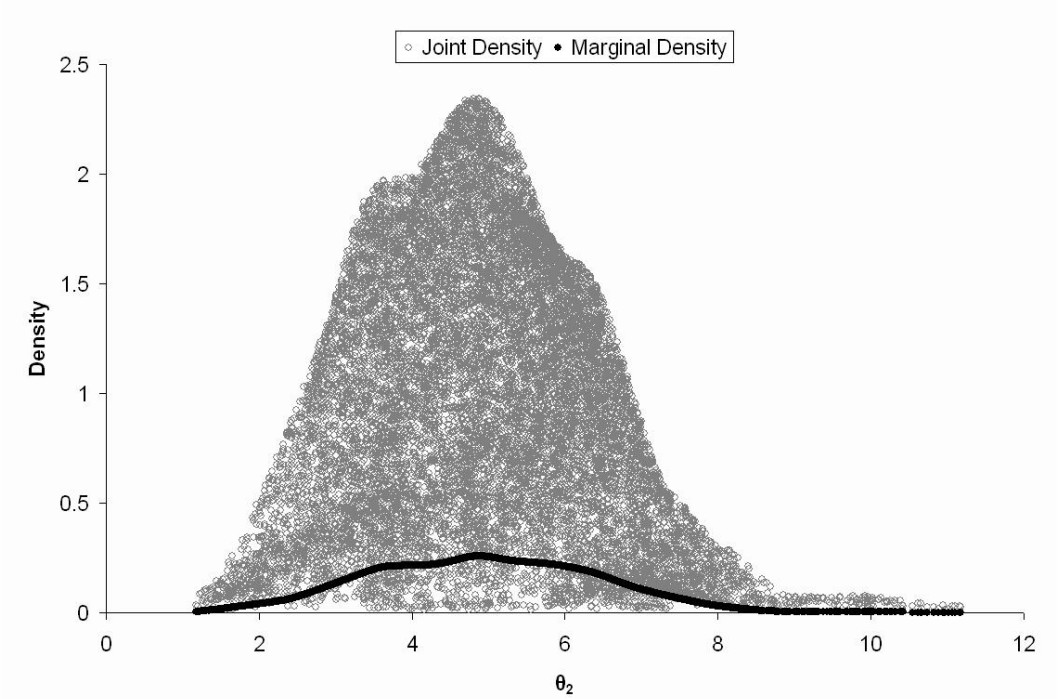

Figure E16. Joint and marginal posterior density estimation for the parameter θ_2 of the EXPTP model for the center 30 x 30m of Coweeta Plot 318.

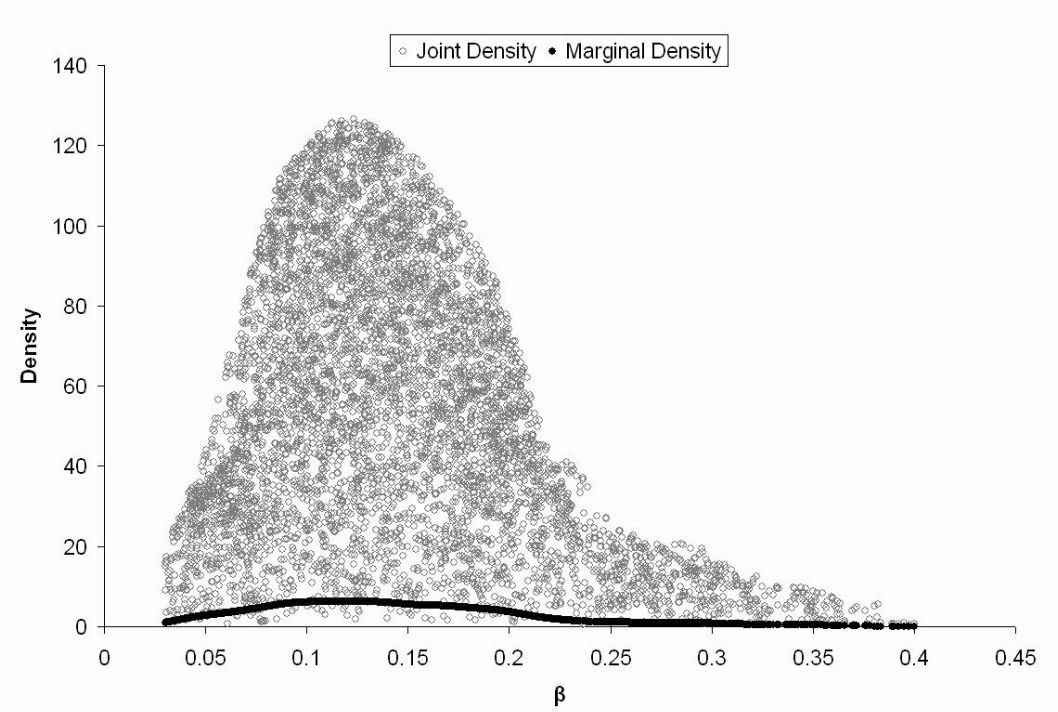

Figure E17. Joint and marginal posterior density estimation for the parameter β of the GTP model for the center 30 x 30m of Coweeta Plot 318.

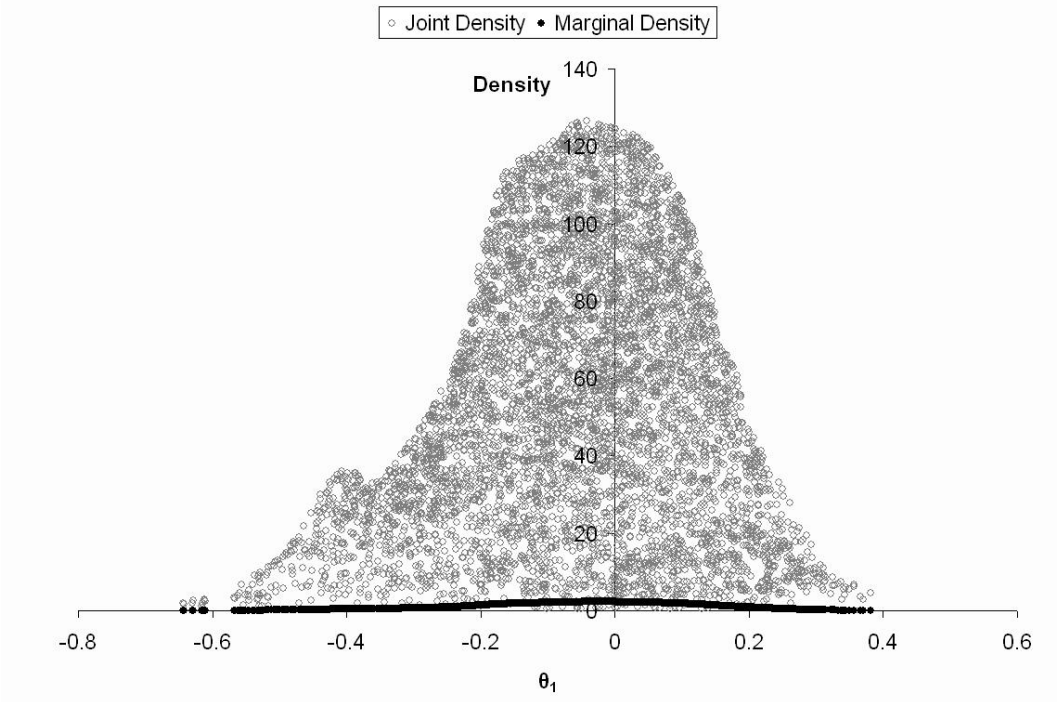

Figure E18. Joint and marginal posterior density estimation for the parameter θ_1 of the GTP model for the center 30 x 30m of Coweeta Plot 318.

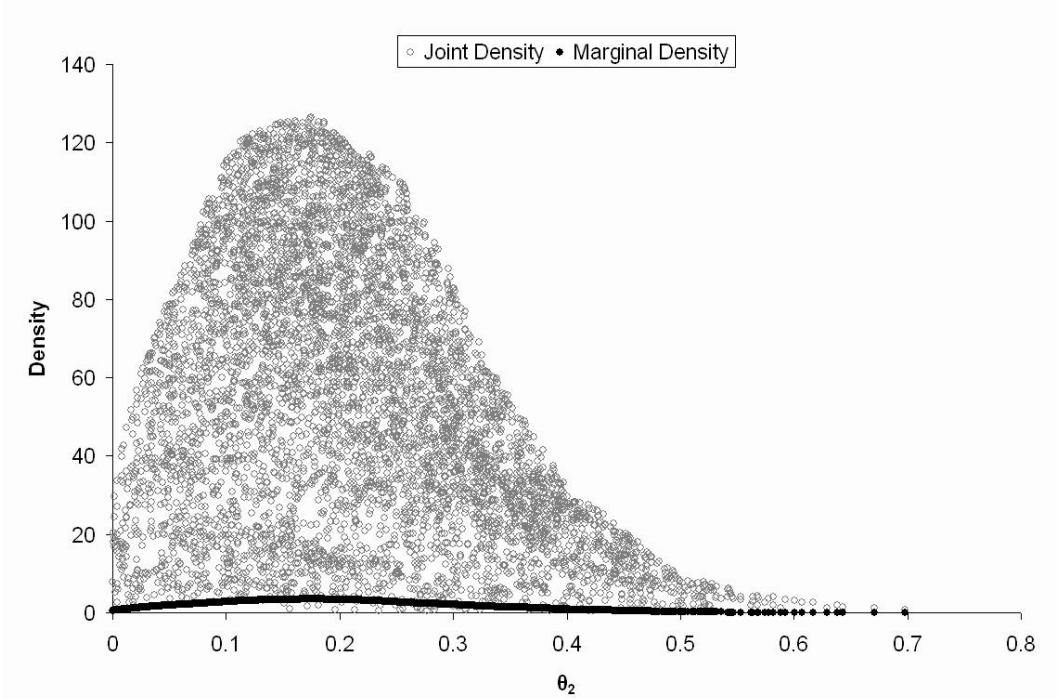

Figure E19. Joint and marginal posterior density estimation for the parameter θ_2 of the GTP model for the center 30 x 30m of Coweeta Plot 318.

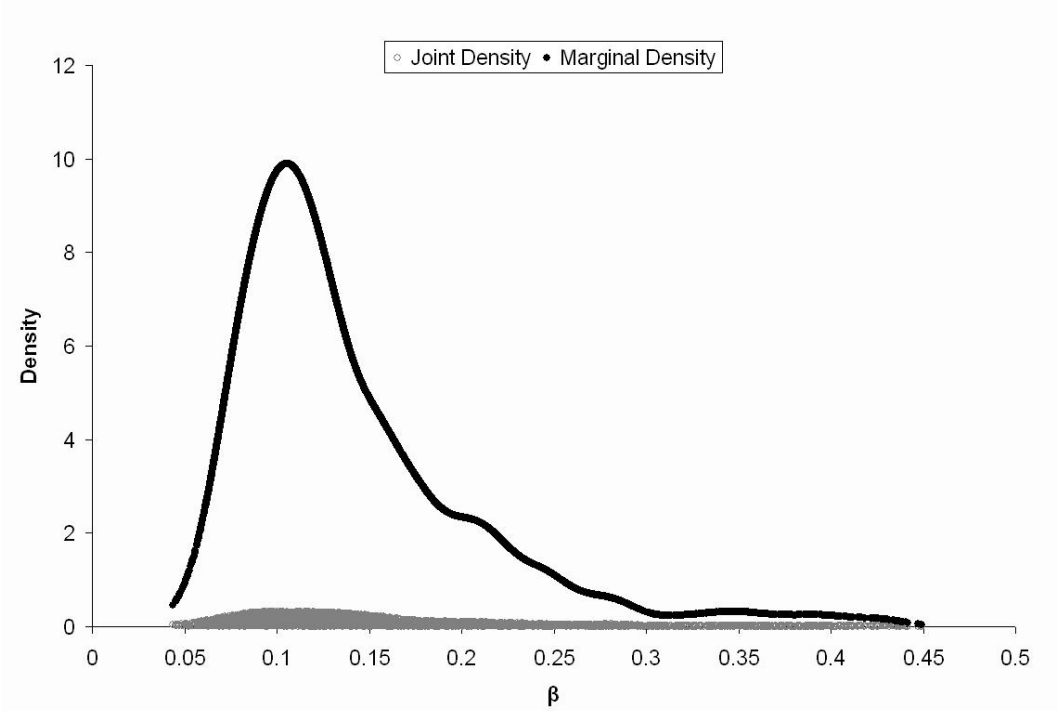

Figure E20. Joint and marginal posterior density estimation for the parameter β of the SPTP model for the center 30 x 30m of Coweeta Plot 318.

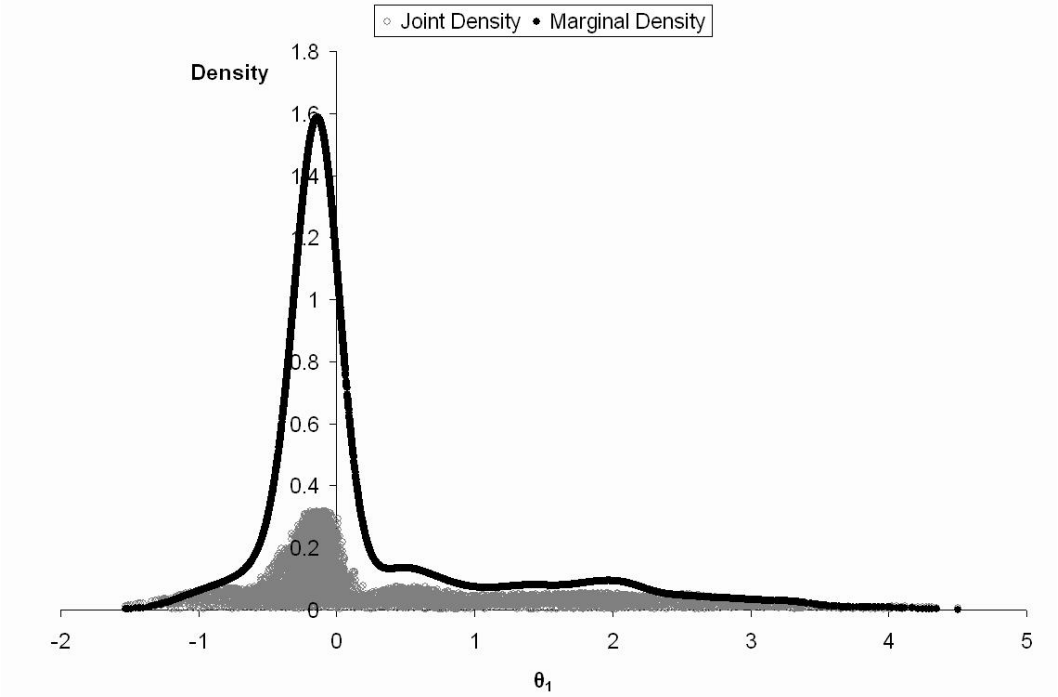

Figure E21. Joint and marginal posterior density estimation for the parameter θ_1 of the SPTP model for the center 30 x 30m of Coweeta Plot 318.

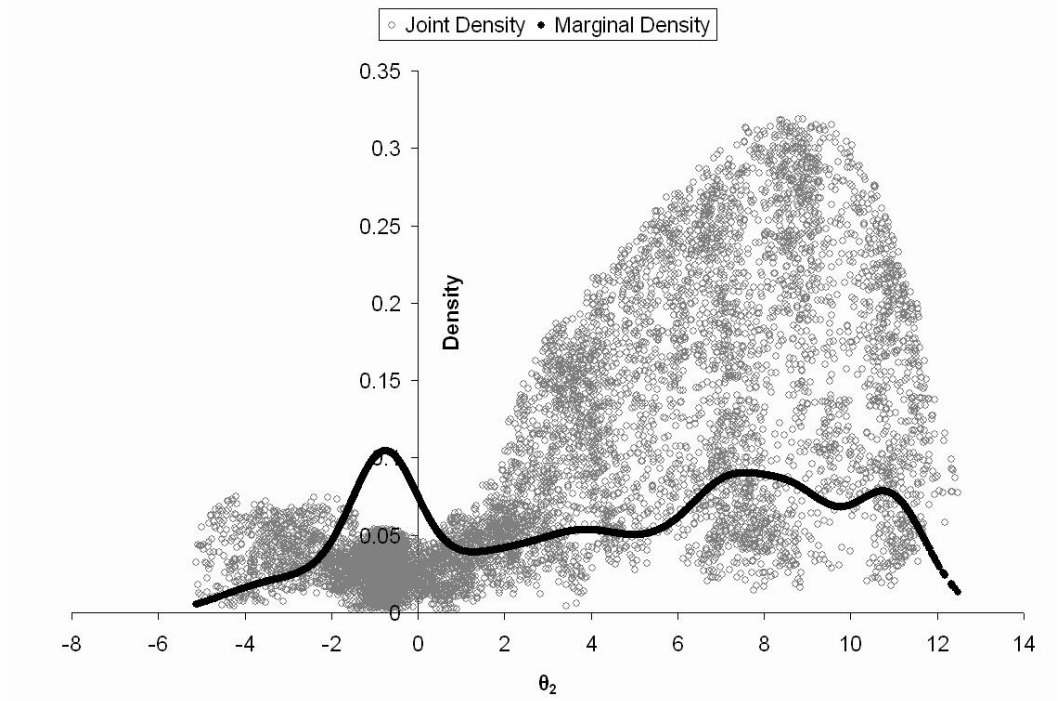

Figure E22. Joint and marginal posterior density estimation for the parameter θ_2 of the SPTP model for the center 30 x 30m of Coweeta Plot 318.

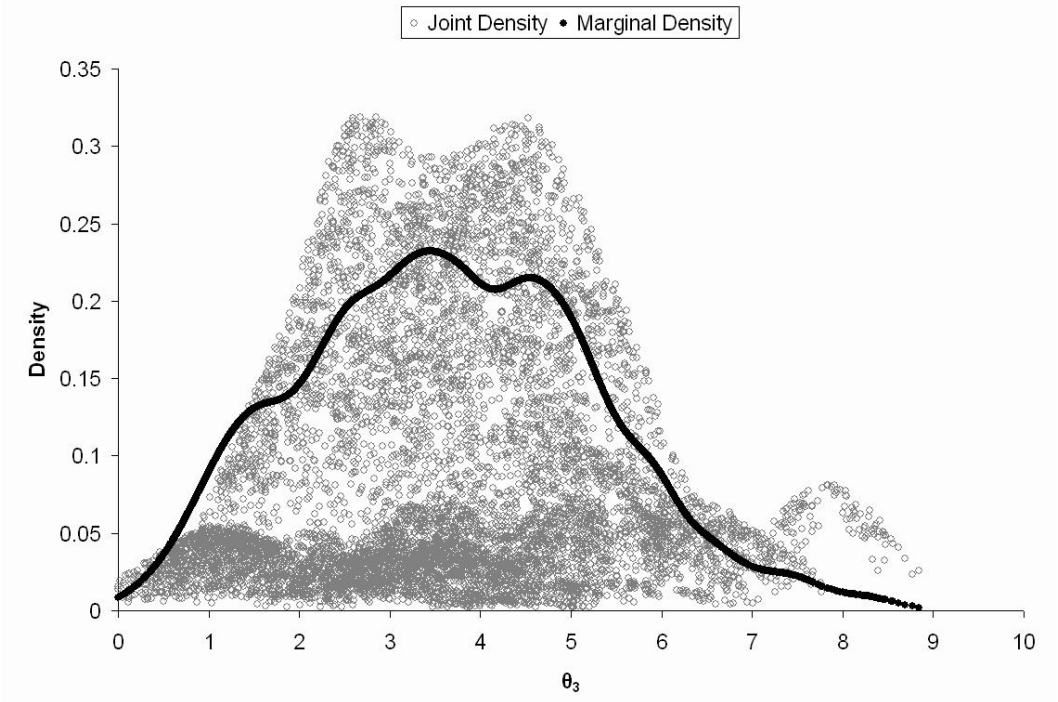

Figure E23. Joint and marginal posterior density estimation for the parameter θ_3 of the SPTP model for the center 30 x 30m of Coweeta Plot 318.

Appendix F

In this appendix are trace plots and plots of marginal vs. joint posterior densities for Bayesian inference of the center 30 x 30m of Coweeta Plot 318 with Møller et al.'s algorithm.

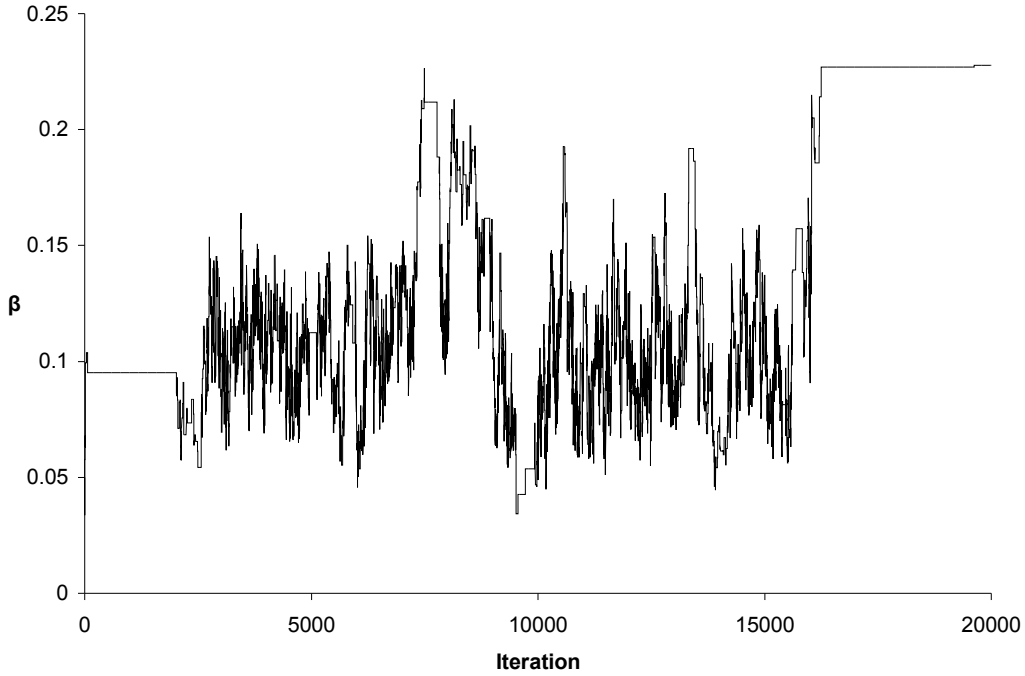

Figure F1. Trace plot of the parameter β of the EXPTP model for the center 30 x 30m of Coweeta Plot 318.

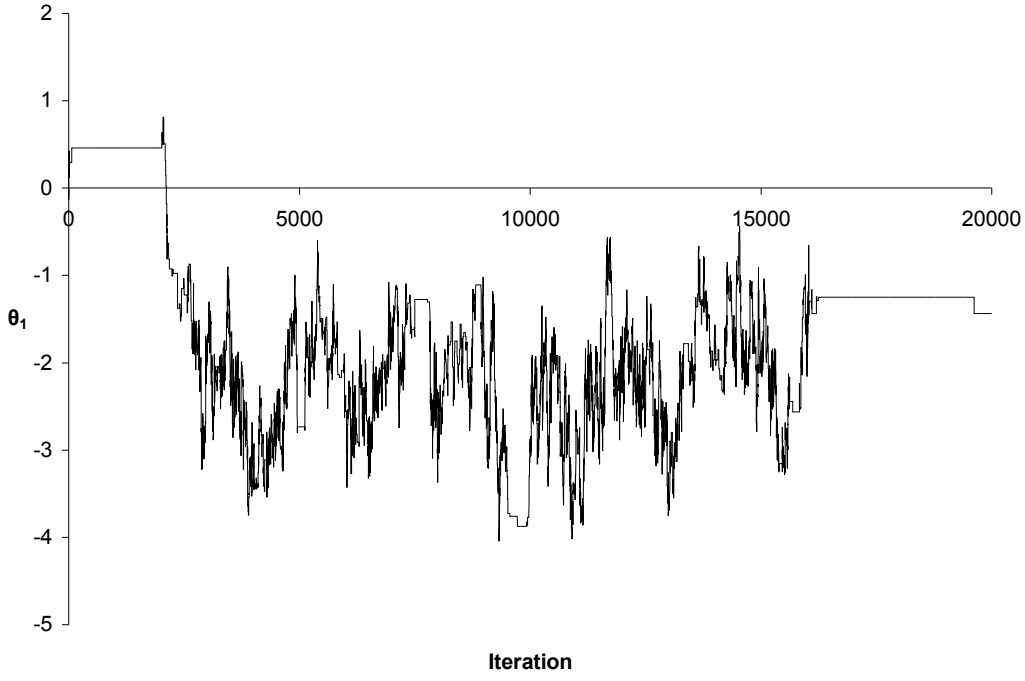

Figure F2. Trace plot of the parameter θ_1 of the EXPTP model for the center 30 x 30m of Coweeta Plot 318.

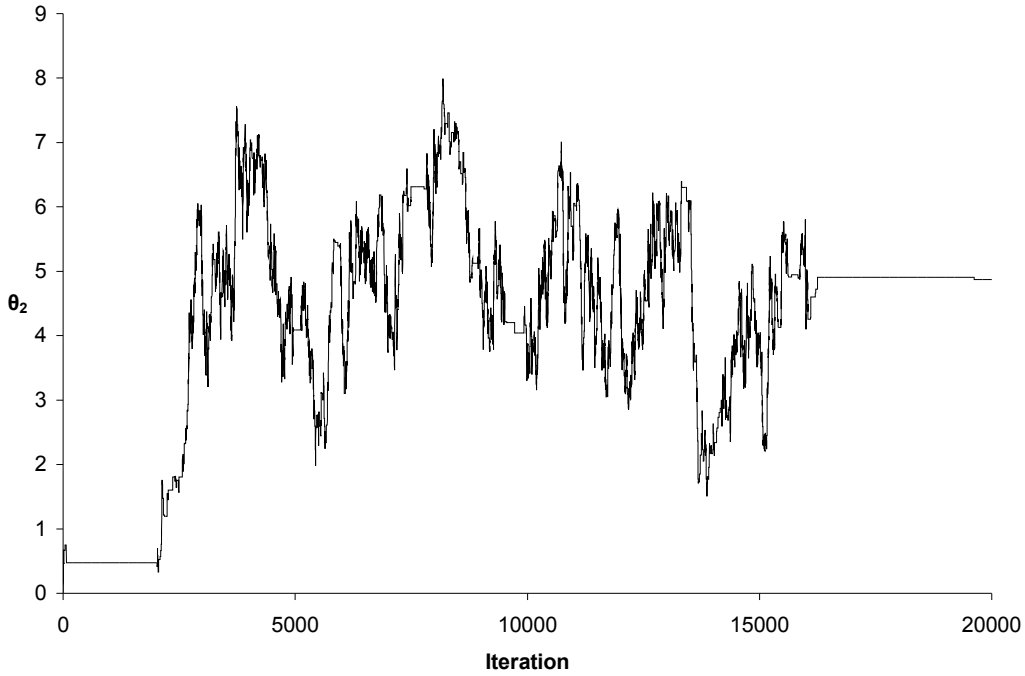

Figure F3. Trace plot of the parameter θ_2 of the EXPTP model for the center 30 x 30m of Coweeta Plot 318.

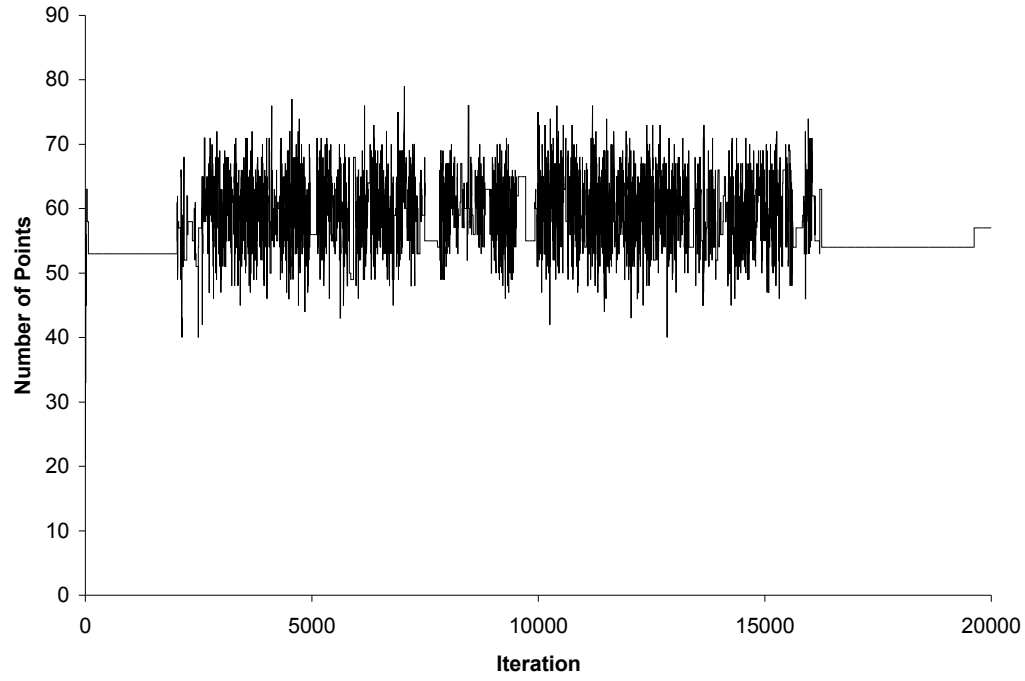

Figure F4. Trace plot of the number of points of the auxiliary point process of the EXPTP model for the center 30 x 30m of Coweeta Plot 318.

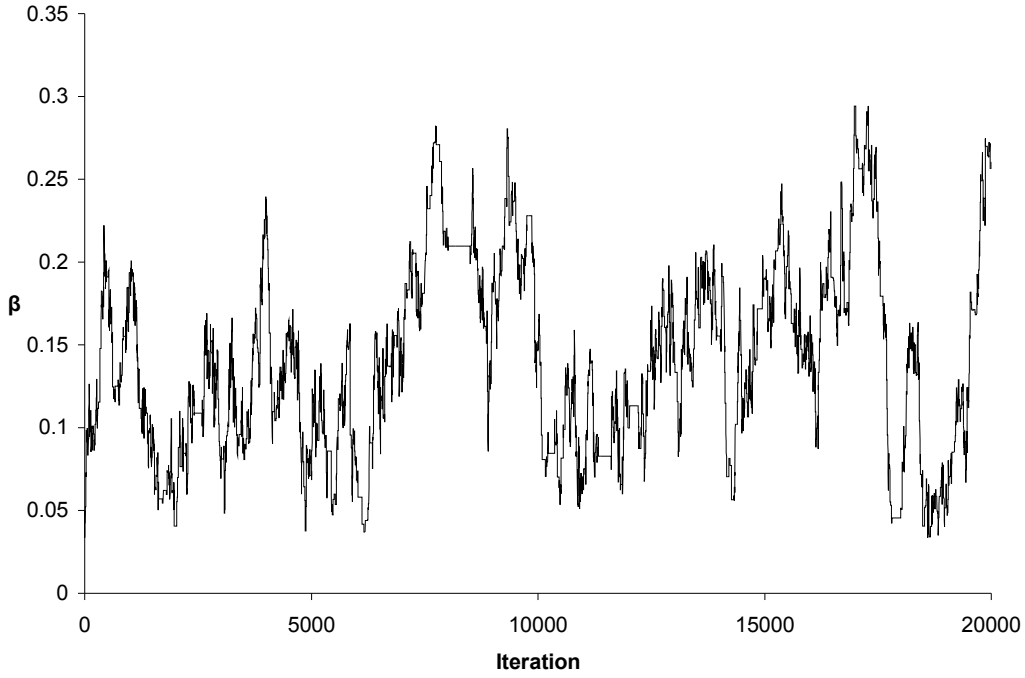

Figure F5. Trace plot of the parameter β of the GTP model for the center 30 x 30m of Coweeta Plot 318.

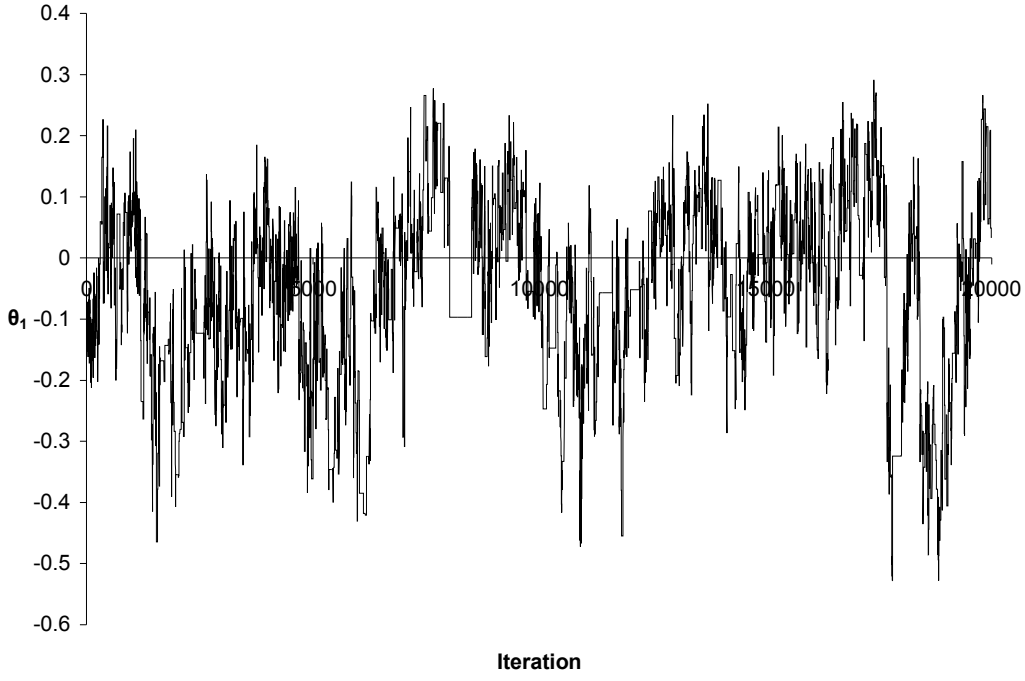

Figure F6. Trace plot of the parameter θ_1 of the GTP model for the center 30 x 30m of Coweeta Plot 318.

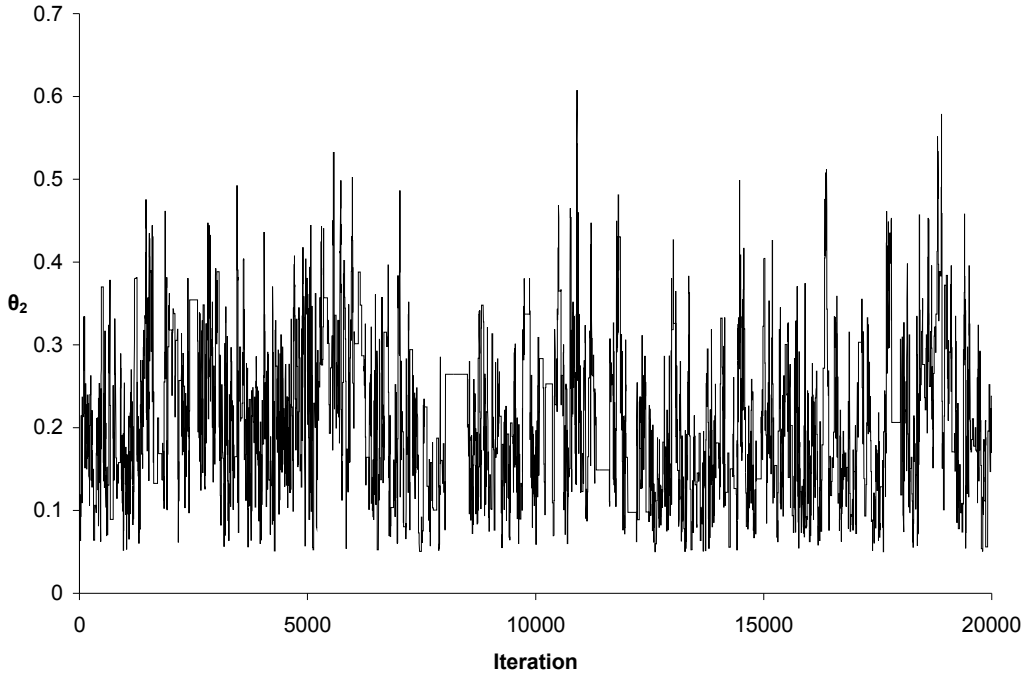

Figure F7. Trace plot of the parameter θ_2 of the GTP model for the center 30 x 30m of Coweeta Plot 318.

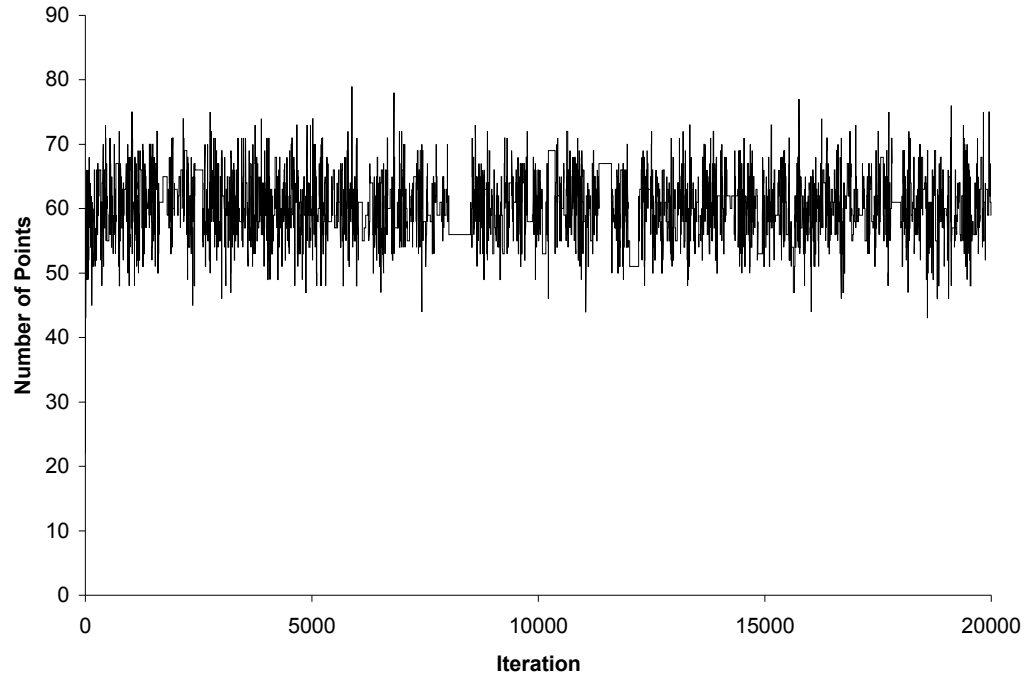

Figure F8. Trace plot of the number of points of the auxiliary point process of the GTP model for the center 30 x 30m of Coweeta Plot 318.

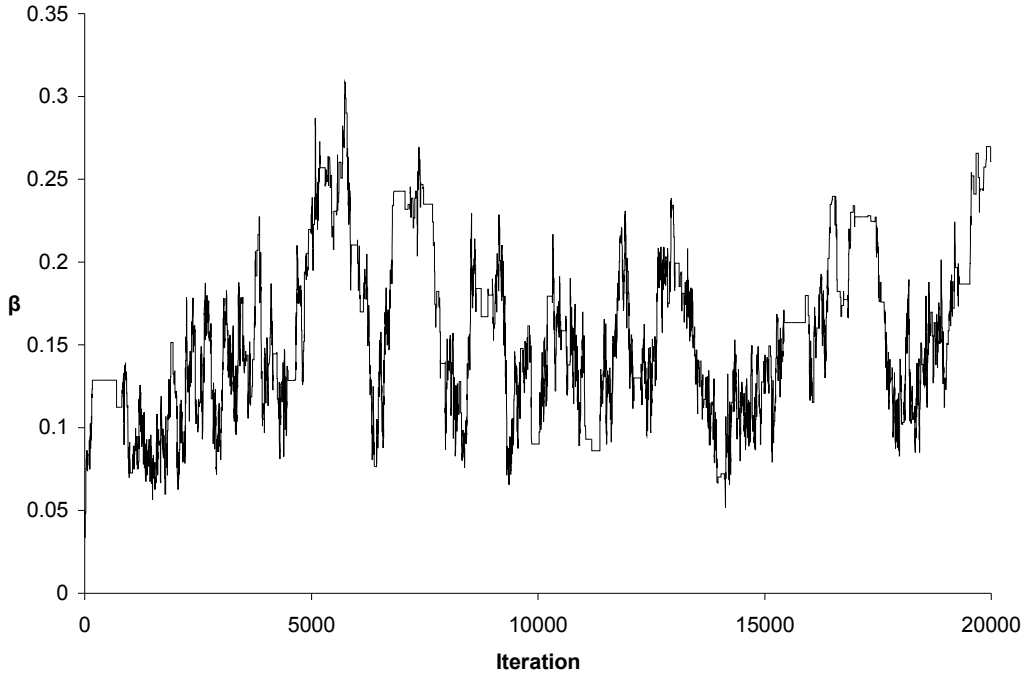

Figure F9. Trace plot of the parameter β of the SPTP model for the center 30 x 30m of Coweeta Plot 318.

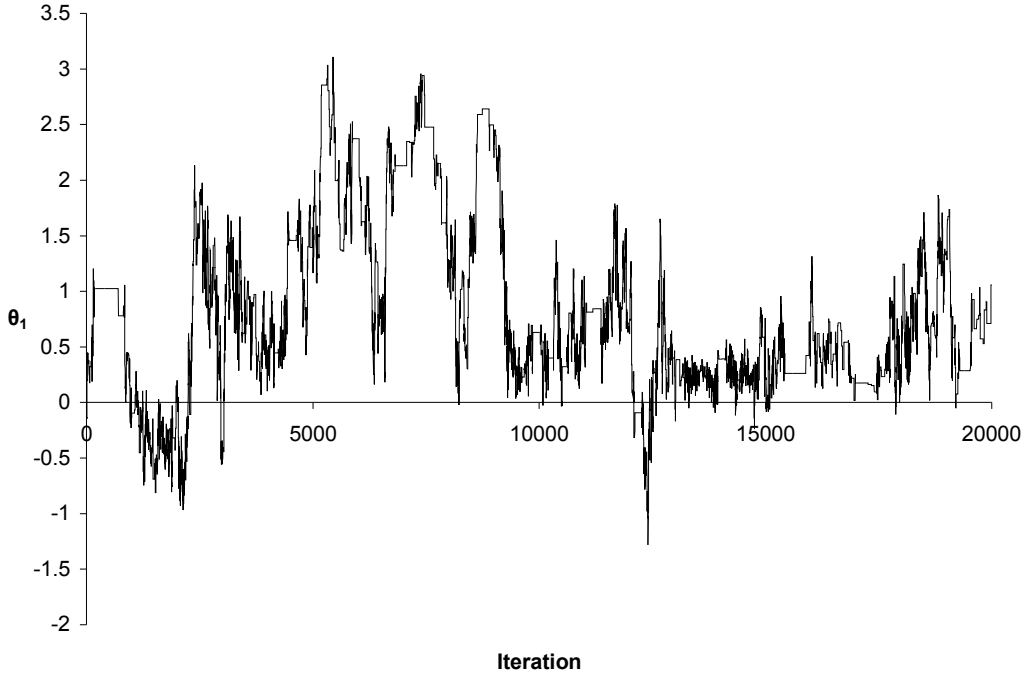

Figure F10. Trace plot of the parameter θ_1 of the SPTP model for the center 30 x 30m of Coweeta Plot 318.

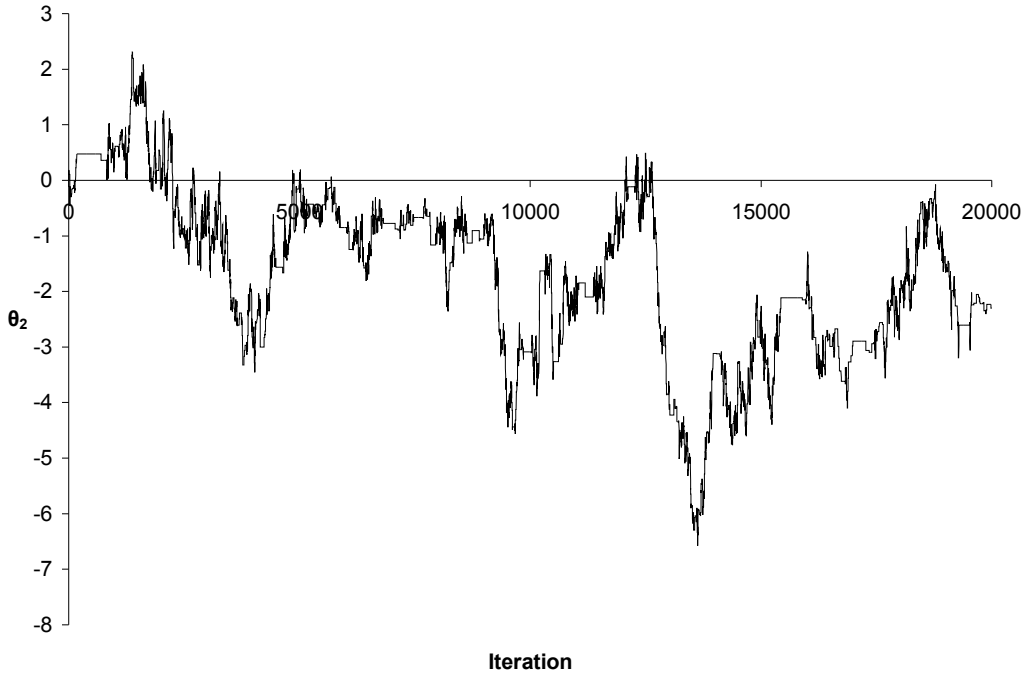

Figure F11. Trace plot of the parameter θ_2 of the SPTP model for the center 30 x 30m of Coweeta Plot 318.

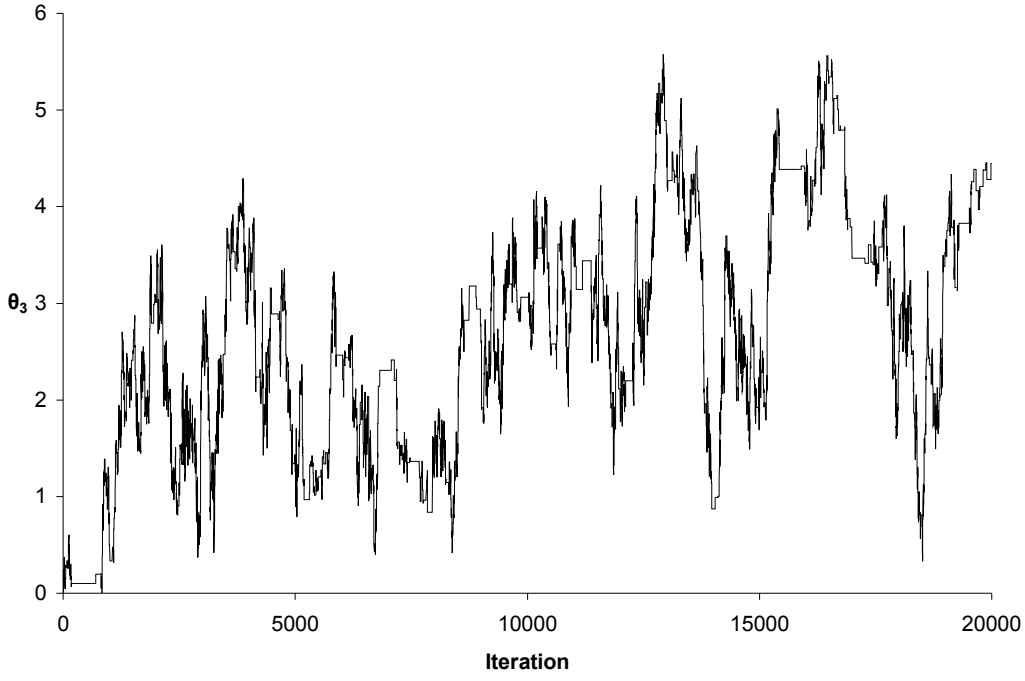

Figure F12. Trace plot of the parameter θ_3 of the SPTP model for the center 30 x 30m of Coweeta Plot 318.

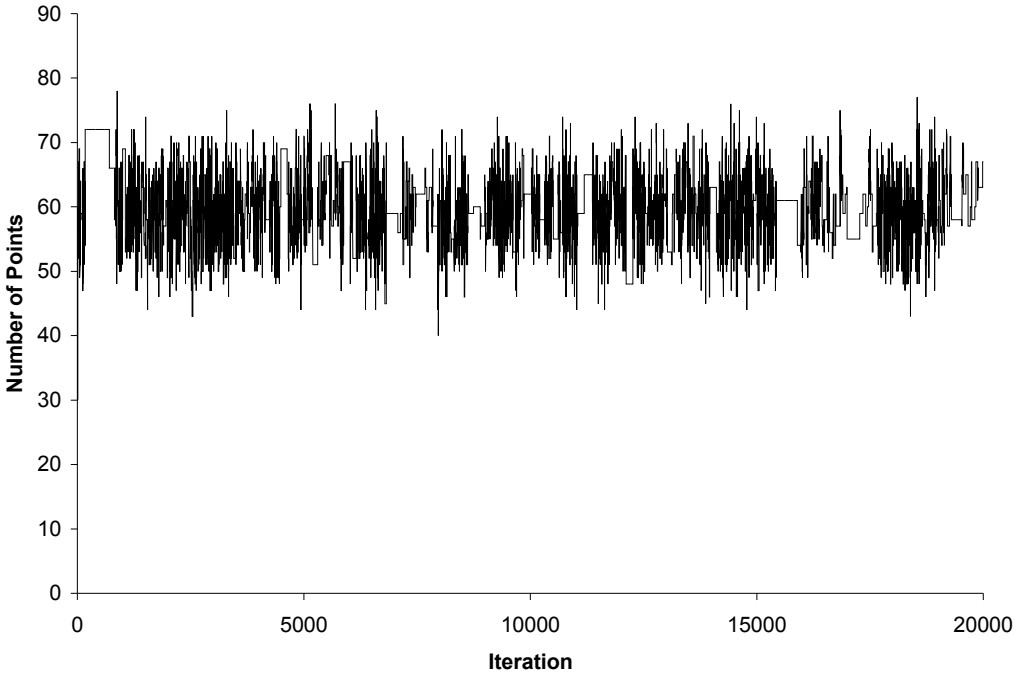

Figure F13. Trace plot of the number of points of the auxiliary point process of the SPTP model for the center 30 x 30m of Coweeta Plot 318.

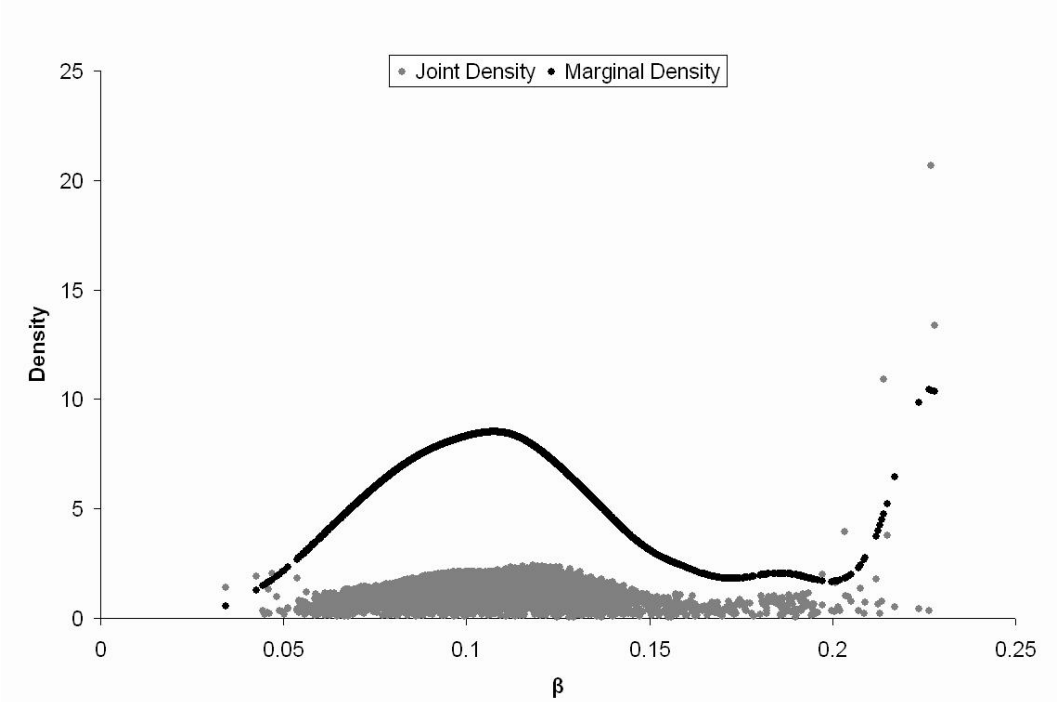

Figure F14. Joint and marginal posterior density estimation for the parameter β of the EXPTP model for the center 30 x 30m of Coweeta Plot 318.

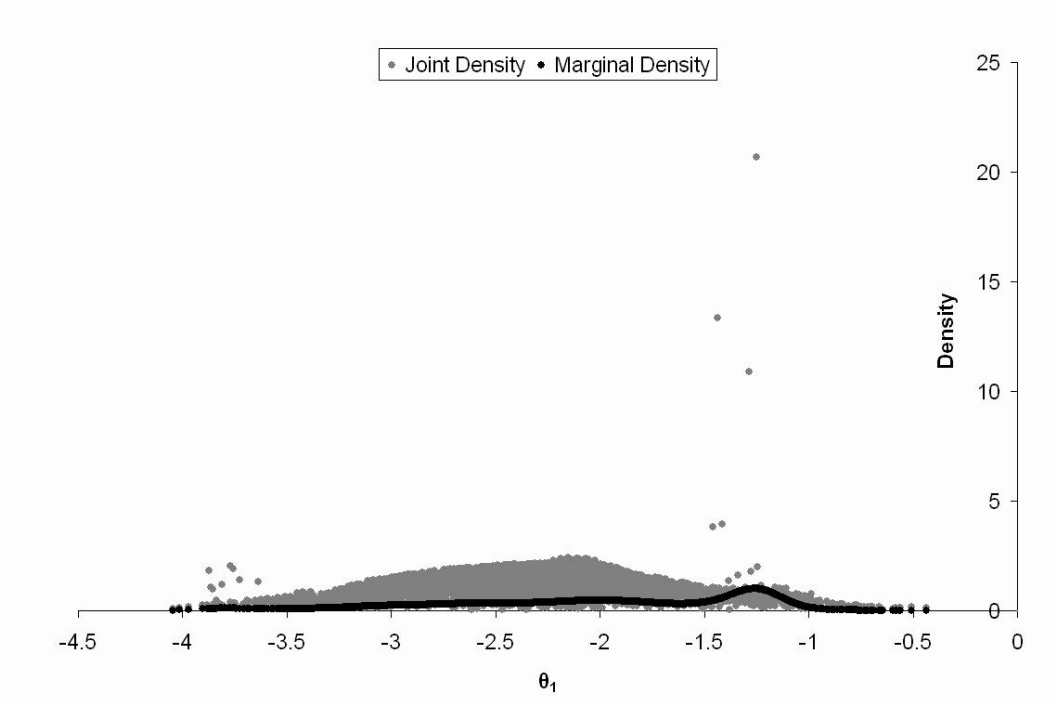

Figure F15. Joint and marginal posterior density estimation for the parameter θ_1 of the EXPTP model for the center 30 x 30m of Coweeta Plot 318.

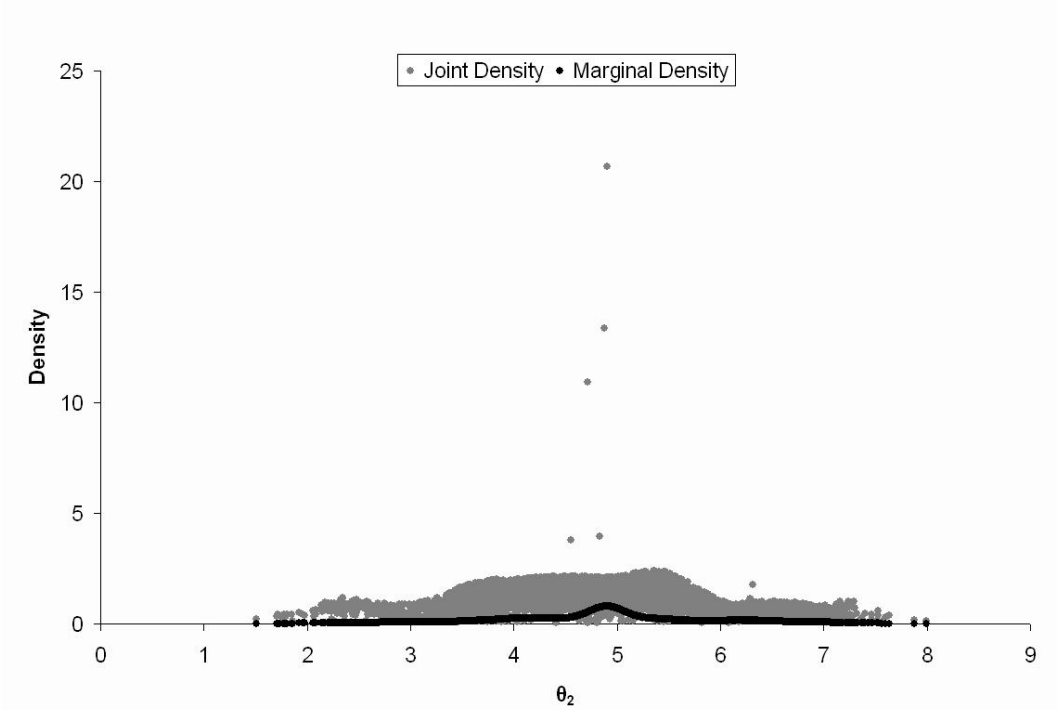

Figure F16. Joint and marginal posterior density estimation for the parameter θ_2 of the EXPTP model for the center 30 x 30m of Coweeta Plot 318.

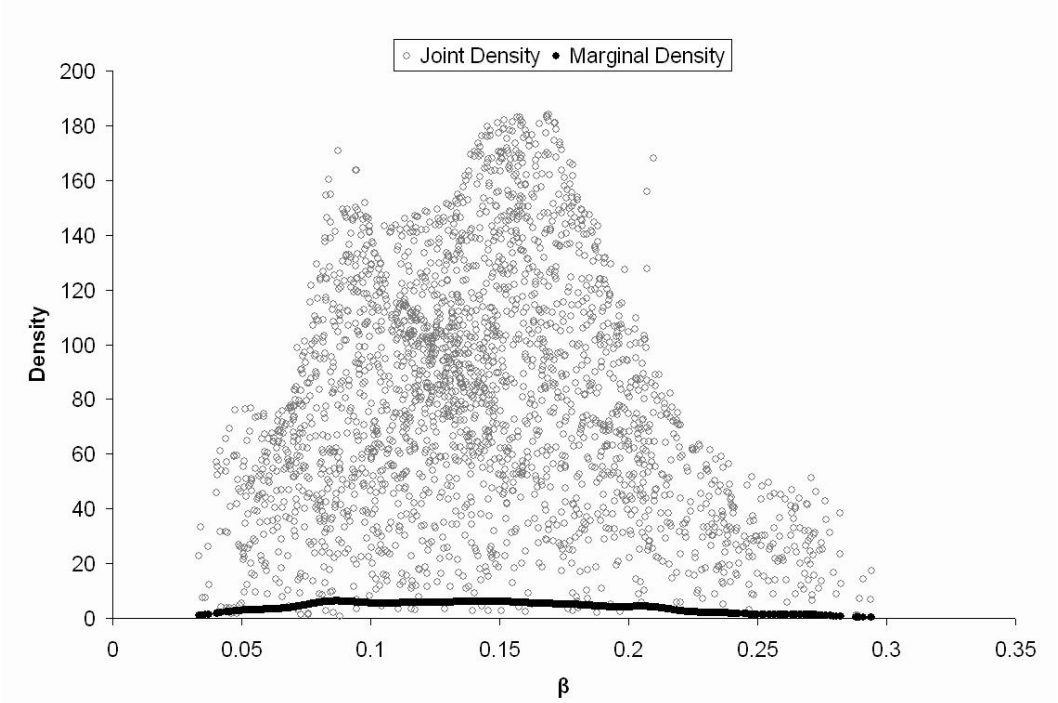

Figure F17. Joint and marginal posterior density estimation for the parameter β of the GTP model for the center 30 x 30m of Coweeta Plot 318.

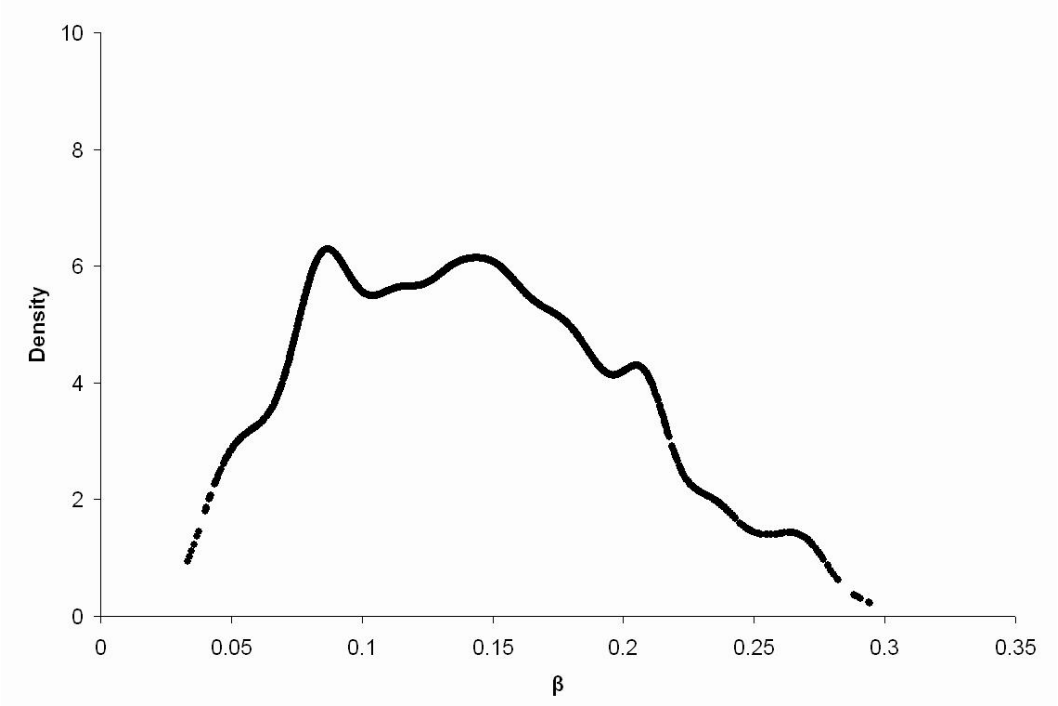

Figure F18. Marginal posterior density estimation for the parameter β of the GTP model for the center 30 x 30m of Coweeta Plot 318.

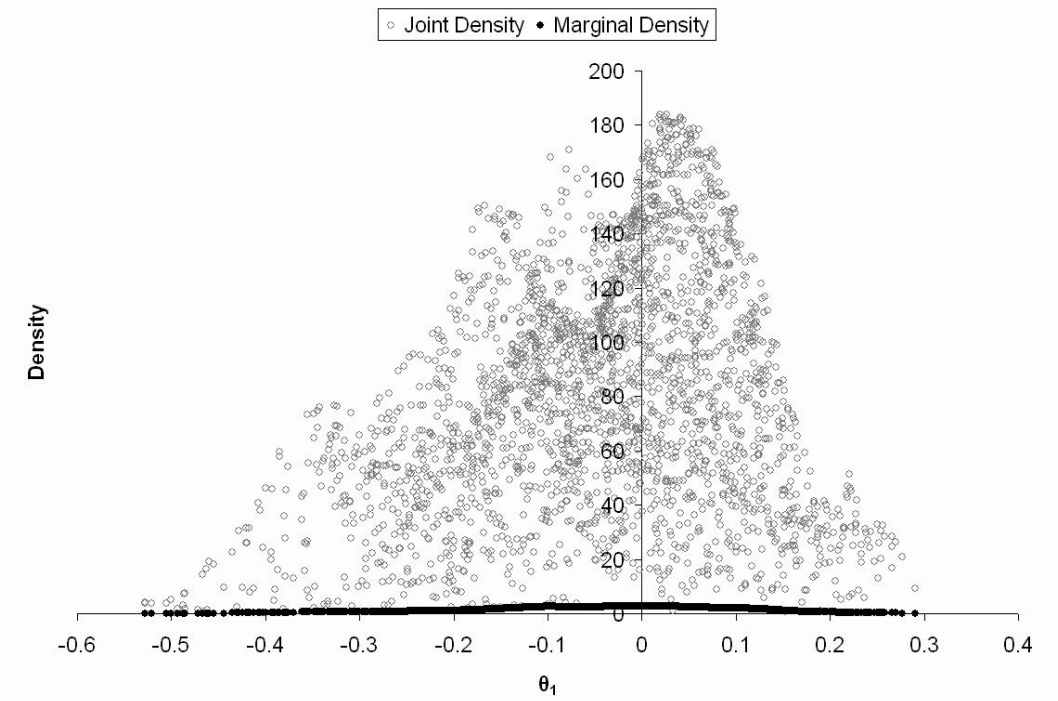

Figure F19. Joint and marginal posterior density estimation for the parameter θ_1 of the GTP model for the center 30 x 30m of Coweeta Plot 318.

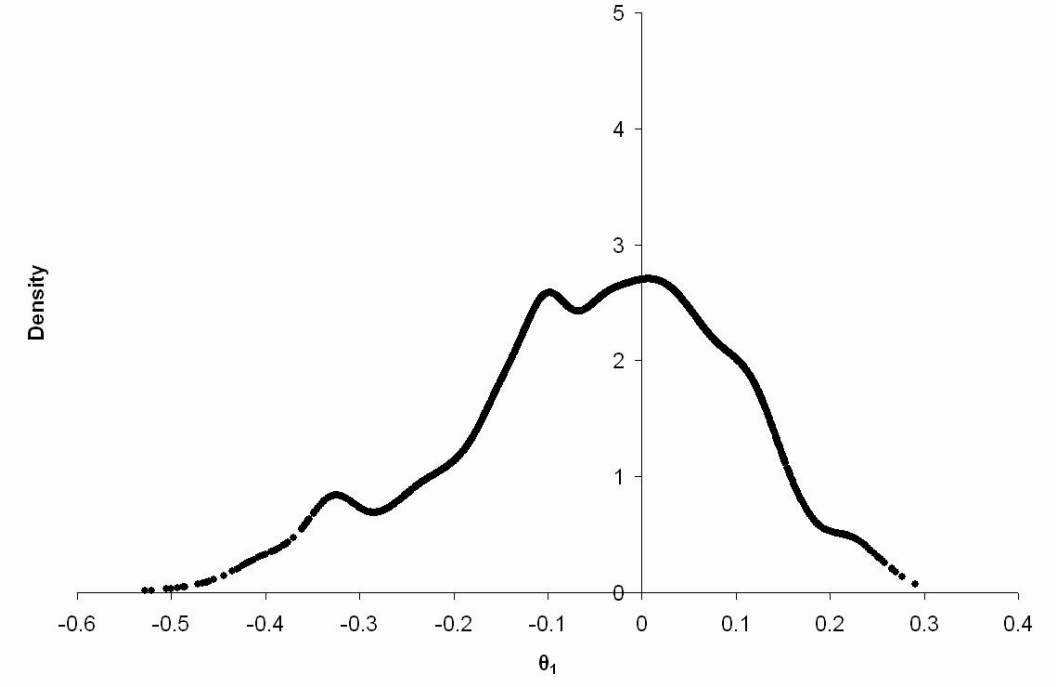

Figure F20. Marginal posterior density estimation for the parameter θ_1 of the GTP model for the center 30 x 30m of Coweeta Plot 318.

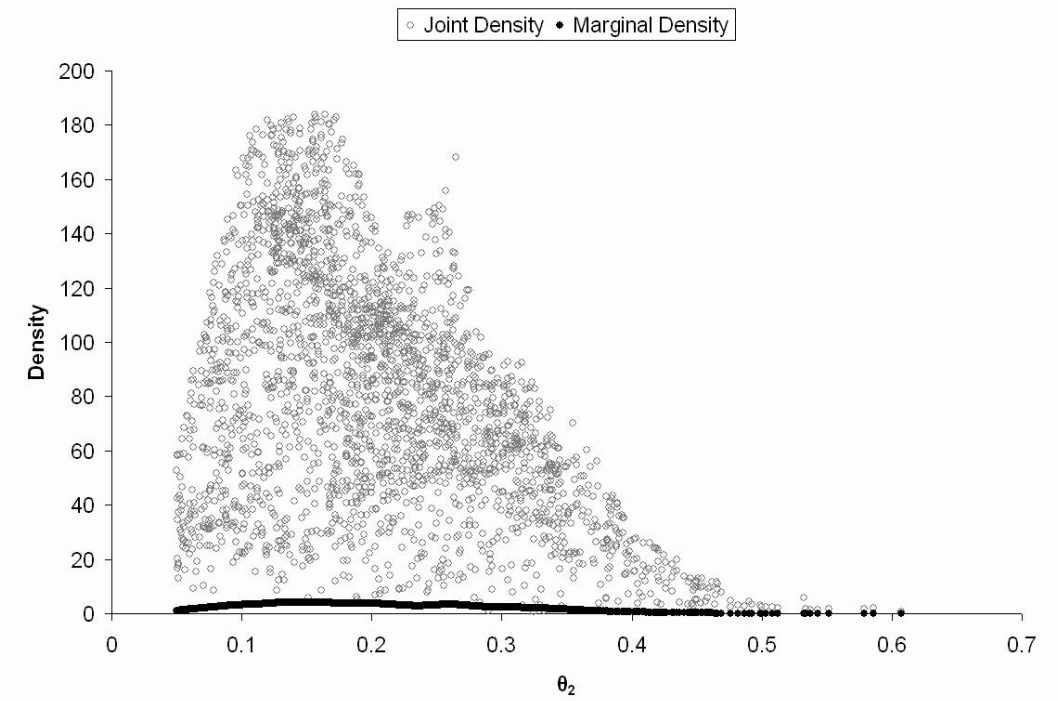

Figure F21. Joint and marginal posterior density estimation for the parameter θ_2 of the GTP model for the center 30 x 30m of Coweeta Plot 318.

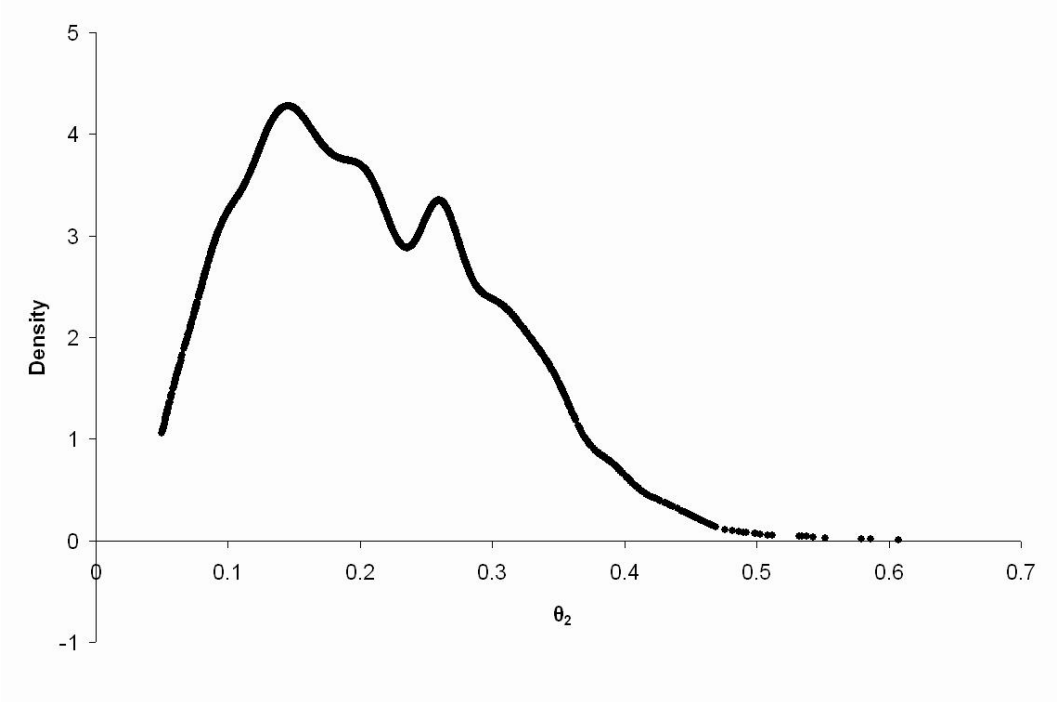

Figure F22. Marginal posterior density estimation for the parameter θ_2 of the GTP model for the center 30 x 30m of Coweeta Plot 318.

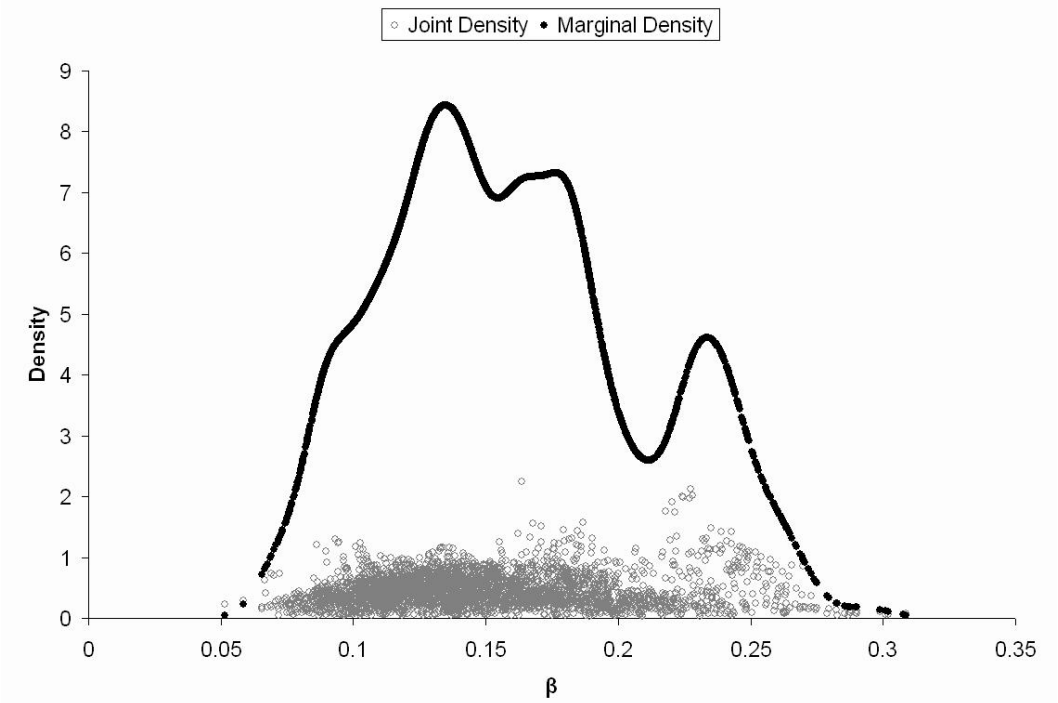

Figure F23. Joint and marginal posterior density estimation for the parameter β of the SPTP model for the center 30 x 30m of Coweeta Plot 318.

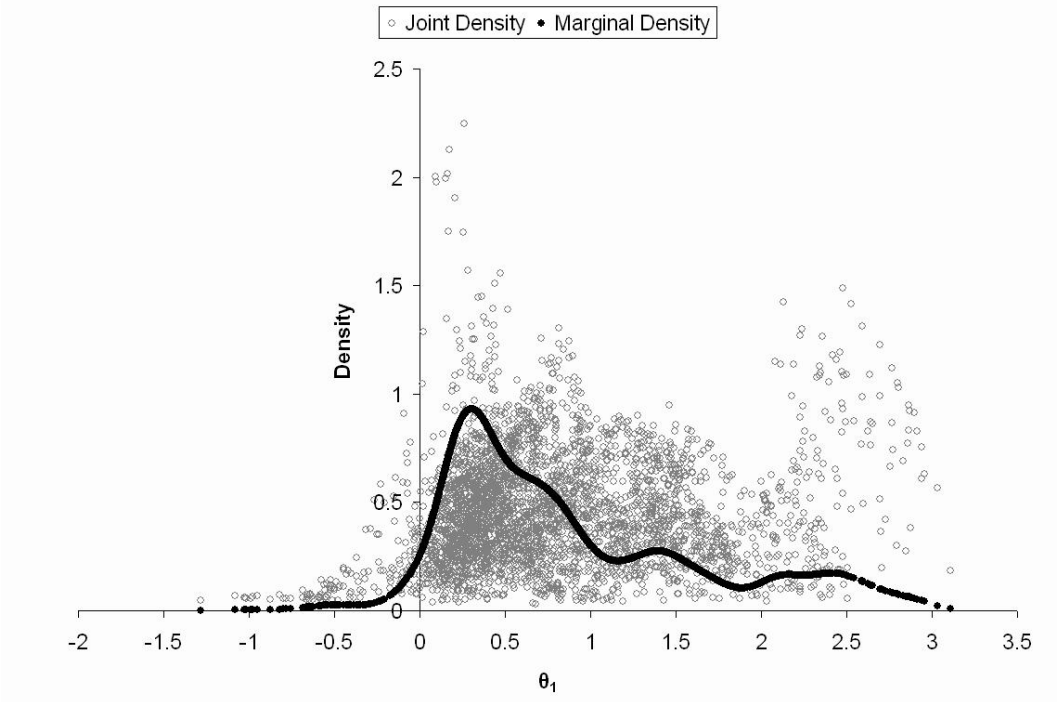

Figure F24. Joint and marginal posterior density estimation for the parameter θ_1 of the SPTP model for the center 30 x 30m of Coweeta Plot 318.

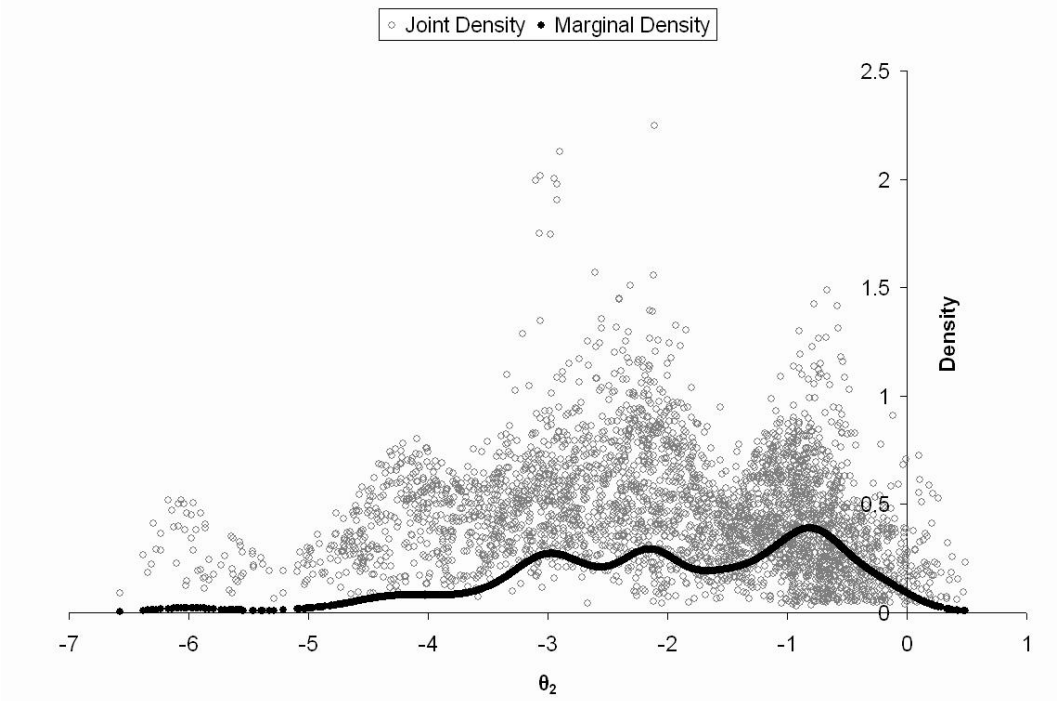

Figure F25. Joint and marginal posterior density estimation for the parameter θ_2 of the SPTP model for the center 30 x 30m of Coweeta Plot 318.

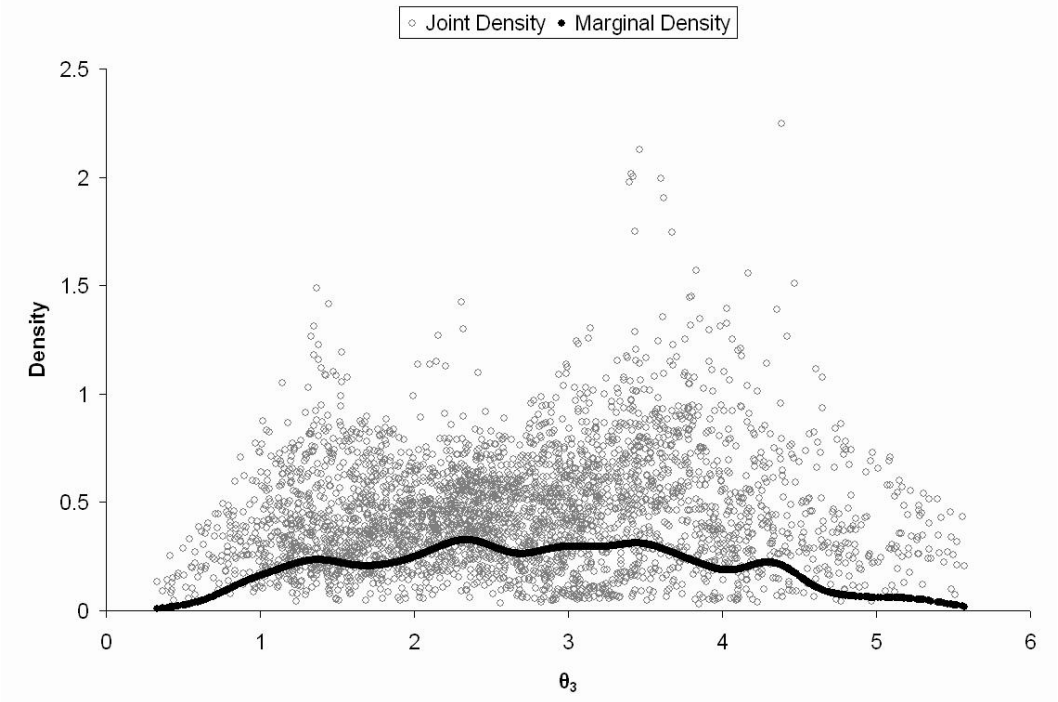

Figure F26. Joint and marginal posterior density estimation for the parameter θ_3 of the SPTP model for the center 30 x 30m of Coweeta Plot 318.