ON THE BAYESIAN ESTIMATION FOR THE STATIONARY NEYMAN-SCOTT POINT PROCESSES

Jiří Kopecký, Tomáš Mrkvička, České Budějovice

(Received May 6, 2015)

Abstract. The pure and modified Bayesian methods are applied to the estimation of parameters of the Neyman-Scott point process. Their performance is compared to the fast, simulation-free methods via extensive simulation study. Our modified Bayesian method is found to be on average 2.8 times more accurate than the fast methods in the relative mean square errors of the point estimates, where the average is taken over all studied cases. The pure Bayesian method is found to be approximately as good as the fast methods. These methods are computationally affordable today.

Keywords: Bayesian method; Monte Carlo Markov chain; Neyman-Scott point process; parameter estimation; shot-noise Cox process; Thomas process

MSC 2010: 62M05, 62H12

1. Introduction

The Neyman-Scott point process [12] is a statistical model which often serves in practice as a model for spatially aggregated points, such as trees in a forest, fish in a reservoir, etc. It represents a class of models which belongs to a wider class of models called shot noise Cox processes defined, e.g., in [8]. Since the Neyman-Scott point processes have been often used in practice, many methods were developed to estimate their parameters. First, the minimum contrast methods [2], second, the composite likelihood method [5], third, the Palm likelihood method [14]. These simulation-free methods used to be preferred for their simplicity and speed. But nowadays, with the growing speed of computers, it has become less time-consuming to use methods based on the likelihood of the spatial point process. The Bayesian method became especially popular in recent years, e.g. in works [9], [11], [10] and [6].

This work was funded by the Grant Agency of the Czech Republic (Project No. 16-03708S).

An extensive simulation study was conducted in [4] comparing all simulation-free, faster-to-compute methods of estimating the parameters of the stationary Neyman-Scott point process. The authors in [4] concluded that the minimum contrast method using the pair correlation function g as a contrast function provides the best estimates in the majority of their simulated cases. Also, this method is the most frequently used method.

In the present paper, we follow the simulation study in [4] and we make an effort to extend their comparison of the performance of estimation methods with the pure Bayesian method. Since we have found surprising inaccuracy of the pure Bayesian method when compared to the simulation-free methods, we slightly modified the pure Bayesian method and compare the performance of this new, modified Bayesian method with other methods.

In the next section we introduce briefly the Neyman-Scott point process and we describe both Bayesian methods used in the study. Then we present the details of the simulation study and its results. Finally, we summarize these results and discuss the suitability of the Bayesian method for parameter estimation of the Cox process.

2. Neyman-Scott point process

The Neyman-Scott process X is a union of clusters $\bigcup_{c \in C} X_c$, where the mother process C is a Poisson process of cluster centers with intensity $\kappa > 0$. Given C, the clusters X_c, $c \in C$, are independent Poisson processes with an intensity function $\alpha k(\cdot - c, \omega)$, where $k(\cdot, \omega)$ is a probability density function parametrized by ω, which determines the spread of daughter points around their mother and α is the expected number of daughters per cluster. If the kernel $k(\cdot, \omega)$ is the density of a symmetric normal distribution with mean 0 and variance $\omega^2 > 0$, then the process is called a modified Thomas process [13] with intensity $\lambda = \kappa\alpha$.

The Neyman-Scott point process can be seen also as a Cox process [13]. The driving intensity function [7], p. 379, of the Cox process is then

$$\tilde{\lambda}(u) = \sum_{c \in C} \alpha k(u - c, \omega).$$

A detailed introduction to the theory of point processes and Neyman-Scott processes can be found for example in [7] or [13]. Two samples can be seen in Figure 1. For the definition of the pair correlation function and K-function see [7], p. 214.

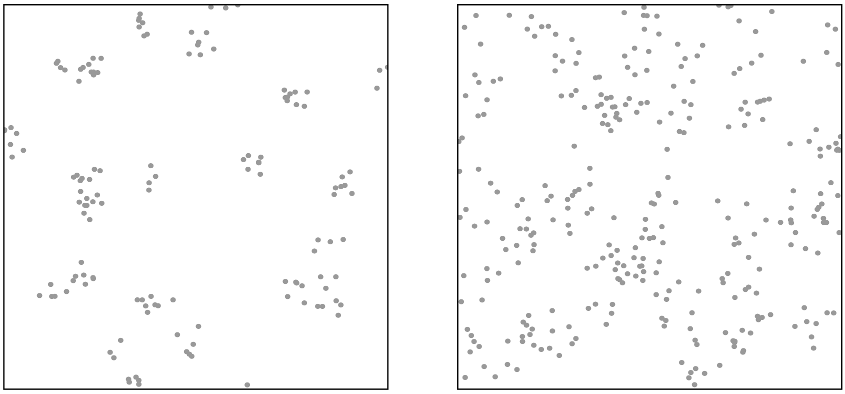

Figure 1. Two samples of the Neyman-Scott process in unit windows with strong and weak clustering in parameter settings $\alpha = 6$, $\omega = 0.02$, $\kappa = 25$ (left) and $\alpha = 4$, $\omega = 0.04$, $\kappa = 50$ (right).

3. Bayesian methods

The pure Bayesian inference for stationary Neyman-Scott point processes is considered in [9], [6], and [10]. It is considered for non-stationary Neyman-Scott point processes in [11]. This approach requires numerical computation of the likelihood function in each step of the MCMC algorithm.

3.1. Pure Bayesian method. In the Bayesian approach for Neyman-Scott point processes, the mother process of cluster centers C with intensity κ is considered an unknown parameter along with the original parameters α, ω and κ of the Neyman-Scott point process X to simplify the computations.

Let $p(C|\kappa)$ denote the probability density of the point process C, in an observation window W, under the knowledge of κ with respect to the homogeneous unit Poisson point process. And let $p(X|C, \alpha, \omega)$ denote the probability density of the point process X with respect to the homogeneous unit Poisson point process in $|W|$, under the knowledge of the point process C and all the parameters. Thus

$$p(X|C, \alpha, \omega) = \exp\left(|W| - \int_W \tilde{\lambda}(u)\,\mathrm{d}u\right) \prod_{x \in X} \tilde{\lambda}(u),$$

where $\tilde{\lambda}(u) = \alpha \sum\limits_{c \in C} k(u - c, \omega)$. The joint posterior distribution of the process C and the parameters is then obtained from the probability density $p(X|C, \alpha, \omega)$ of the point process X, from the probability density $p(C|\kappa)$ and from the prior probability densities $p(\kappa)$, $p(\alpha)$, and $p(\omega)$ in the form

$$p(C, \kappa, \alpha, \omega|X) \propto p(X|C, \alpha, \omega)p(C|\kappa)p(\kappa)p(\alpha)p(\omega).$$

Our MCMC simulation algorithm uses the Birth-Death-Move algorithm [8] for updating C and the Metropolis-Hastings algorithm for updating parameters κ, α, and ω in every step.

The Bayesian point estimates of κ, α, and ω are then the expected values of the estimated posterior distribution.

3.2. Modified Bayesian method. Since the pure Bayesian method overestimates the parameter κ (see Table 1), we modified the procedure by fixing the parameter κ. First, we estimated the intensity λ of point process X by the classical estimator $\hat{\lambda} = X(W)/|W|$, where $X(W)$ denotes the number of points of X in W and $|W|$ denotes the area of W. Second, we updated α and ω in every step of the MCMC chain. Finally, κ is calculated in every step by the formula $\kappa = \hat{\lambda}/\alpha$.

The joint posterior distribution of the process C is then

$$p(C, \alpha, \omega | X) \propto p(X | C, \alpha, \omega) p(C|\alpha) p(\alpha) p(\omega),$$

where $p(C|\alpha)$ denotes the probability density of the point process C under the knowledge of α with respect to the homogeneous unit Poisson point process.

Three different samples of resultant various posterior distributions for the modified Bayesian method from the simulated study are in Figure 2.

4. Simulation study

Our study compares the results of the two Bayesian methods mentioned above with the results of the minimum contrast method based on the pair correlation function g. The latter method performed best in the simulation study in [4]. The authors in [4] conducted a comparison of the performance of the minimum contrast estimation based on the K-function and on the pair correlation function g, the composite likelihood estimation in different settings, and the Palm likelihood estimation method also with different choices of the tuning parameter.

They assessed the performance of the estimators in middle-sized to large point patterns exhibiting different degrees of clustering. They estimated the parameters of the Thomas process for eight different combinations of parameters using the three moment estimation methods. They considered also the log-Gaussian Cox process. The results are summarized in [3] in the tables of the relative mean biases and relative mean squared errors (MSEs) of these methods.

The minimum contrast method using the pair correlation function g provides the best estimates in the majority of cases in their study and it is also the most frequently used method. To facilitate the comparison of the Bayesian methods, we chose the

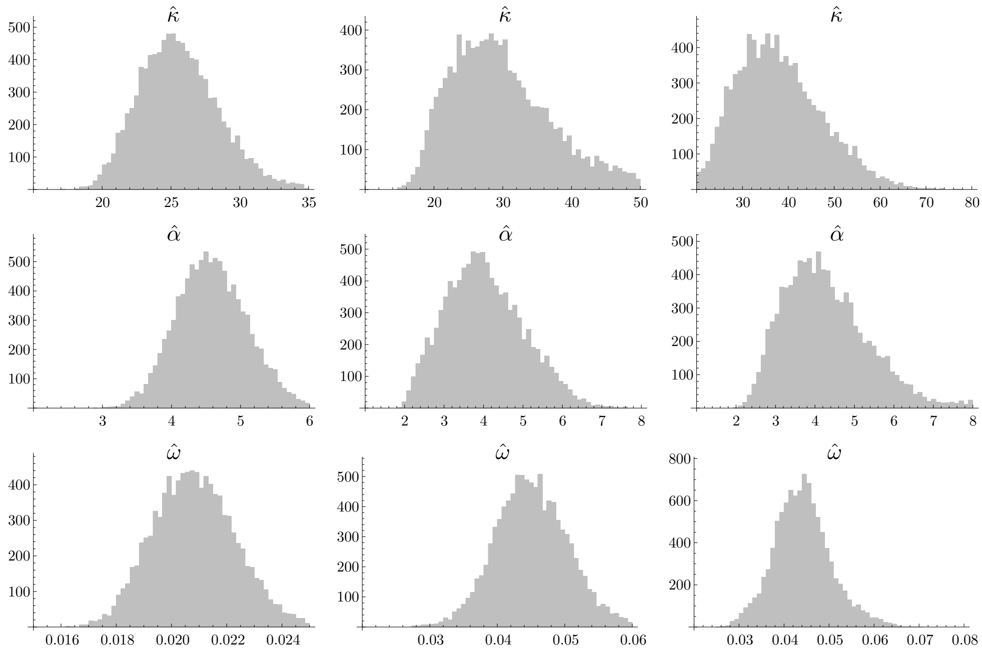

Figure 2. Examples of chosen posterior distributions from the simulation study calculated by the modified Bayesian method for three different parameter combinations. The true parameter values α, κ, ω of the first sample are 25, 4, 0.02 (left column), 25, 4, 0.04 for the second sample (middle column), and 50, 4, 0.04 for the third sample (right column).

minimum contrast method based on the pair correlation function g as a benchmark. Note that we follow [8] in the notation. I.e., the parameters α and ω correspond to the parameters μ and σ in [4] and [3].

4.1. Tuning parameters of the Bayesian methods. Like other estimating methods, the Bayesian method has some tuning parameters. Here it is especially the choice of the priors and the standard deviation h of the Gaussian distributions which suggest new parameter values when updating them inside the Metropolis-Hastings algorithm.

Since the posterior distribution of a parameter is highly sensitive to its prior distribution, we used the completely flat proper prior as in [9] (e.g., uniform prior on a bounded interval) for all parameters, i.e., $p(\alpha) \sim 1_{[0.03,20]}(\alpha)$, $p(\omega) \sim 1_{[0.001,0.2]}(\omega)$ and in the case of the pure Bayesian method $p(\kappa) \sim 1_{[1,300]}(\kappa)$ as well.

In order to compare the two Bayesian methods, we had to use the same proposal variances for both of them. Proper values have to be chosen more carefully when using flat priors instead of the usual vague or informative priors. Especially for the pure Bayesian method, it is important to maintain sufficient parameter mixing, i.e., to allow the model parameters to explore the whole domain of possible values. If the proposal variances h_κ and h_α are too small, the algorithm becomes unstable and the estimate of parameters can run into meaningless values as seen in Figure 3, where the limited parameter changes defined by h in each individual step are not able to reverse the wrong trend of estimation.

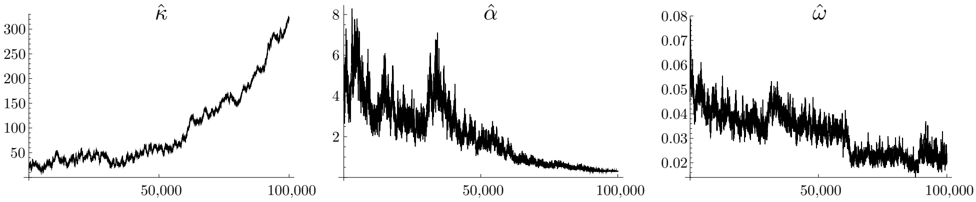

Figure 3. Sample run of the pure Bayesian method MCMC algorithm with insufficient parameter mixing. The estimation of parameter κ rises disregarding the true value of 25, while the estimation of parameter $\alpha = 4$ decreases too much. For numerical reasons the algorithm was stopped after 100,000 steps.

The proposal distributions were set to be symmetric normal distributions with standard deviations $h_\kappa = 2.5$, $h_\alpha = 0.25$, and $h_\omega = 0.005$. Note that in using the modified Bayesian method there were no problems for a large percentage of tested settings that failed with the pure method. More effects of standard deviations on the posterior distribution will be discussed in the conclusion.

Another tuning parameter is the length of the MCMC chain which is discarded and the total length of the chain. Initially we discarded 50,000 of 150,000 steps, since the MCMC chain appeared to be stationary then as seen in Figure 4. However, the study results and the subsequent analysis of posteriors revealed that some "difficult" parameter combinations (i.e., $\kappa = 50$, $\alpha = 4$, and $\omega = 0.04$) could yield even better results if we extend the length of the MCMC chain. This may be indicated by comparing histograms in Figures 2 and 5. Finally, we set the total length of the MCMC chain to 250,000 steps and the discard step to 100,000. The posterior characteristics were calculated from every tenth step of the chain to avoid substantial dependencies.

4.2. Design. In our simulation study we generated 200 independent realizations of the Thomas process in unit square observation windows for eight combinations of parameter values identically as in [4]: $\kappa = 25$ or 50, $\alpha = 4$ or 6 and $\omega = 0.02$ or 0.04. These different sets of parameters represent relatively strong and weak clustering. The intensity of the process ranges from 100 to 300.

Realizations were simulated using the package *spatstat* for R (for reference see [1]). Using the same package we re-estimated the parameters of all processes by the mini-

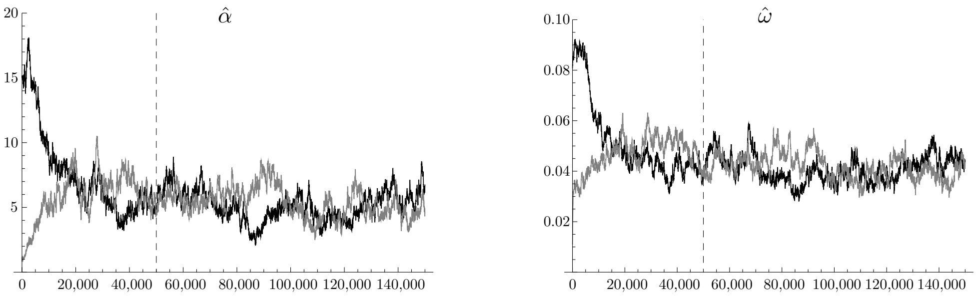

Figure 4. Convergence of two runs of the Metropolis-Hastings algorithm on the selected sample with different initial values of the parameters (α—left, ω—right). The black starts at a very high value and the grey takes off with a too low value relative to the equilibrium. The discard step is illustrated by a vertical dashedline.

mum contrast method using the pair correlation function g. Note that the results are consistent with [4].

For each realization we also estimated the parameters by the pure Bayesian method and the modified Bayesian method described above. The algorithm was implemented in the software Mathematica 8, see [15].

4.3. Results. The results of the simulation study are presented in Table 1 and Table 2. The tables show relative mean squared error and relative bias (rel. MSE was obtained by dividing by the square of the true parameter value while the rel. bias was obtained by dividing by the true parameter value). All the statistics were obtained from the middle 95% of the estimated values as in [4].

The mean square error of the modified Bayesian method is smaller than the mean square error of the simulation-free moment estimation. In our simulated cases, the modified Bayesian method outperforms the moment method for all parameter combinations. We have calculated from the result table that the modified Bayesian method is on average 2.8 times more accurate than the minimum contrast estimation based on the pair correlation function g if comparing mean square errors. Here the average was calculated over all studied cases and over all estimated parameters. Specifically, the result for parameter κ was 2.77 times more accurate, the result for α was 2.74 times more accurate and the result for ω was 3.02 times more accurate.

Considering the pure Bayesian method, it was less accurate than the modified Bayesian method especially for parameter κ. Specifically, the result for parameter κ of the modified Bayesian method was 12 times more accurate on average than the result for the pure Bayesian method. This huge number was caused by two extreme cases where $\alpha = 4$ and $\omega = 0.04$. The other parameters seem to be estimated with similar precision.

	κ	α	ω	MCE_g	B_{pur}	B_{mod}
$\hat{\kappa}$	25	4	.02	.059	.048	.034
			.04	.202	1.31	.060
		6	.02	.073	.050	.033
			.04	.177	.130	.057
	50	4	.02	.069	.050	.024
			.04	.198	3.23	.054
		6	.02	.046	.023	.021
			.04	.181	.497	.062
$\hat{\alpha}$	25	4	.02	.040	.013	.013
			.04	.124	.098	.074
		6	.02	.044	.009	.010
			.04	.095	.042	.049
	50	4	.02	.046	.018	.016
			.04	.112	.201	.105
		6	.02	.037	.008	.007
			.04	.106	.087	.053
$\hat{\omega}$	25	4	.02	.015	.005	.005
			.04	.039	.027	.019
		6	.02	.014	.003	.003
			.04	.033	.009	.010
	50	4	.02	.016	.005	.005
			.04	.040	.043	.024
		6	.02	.012	.003	.003
			.04	.032	.014	.010

Table 1. Summary of simulation results—relative mean square errors for the pure Bayesian method (B_{pur}) and the modified Bayesian method (B_{mod}) estimation compared with the minimum contrast estimation based on the pair correlation function g (MCE_g).

The comparison of biases is shown in Table 2. The table shows significant positive bias of the estimates of κ both for the pure Bayesian method and for the minimum contrast method.

Finally, we also observed correlations between estimated parameters and we have found that α and κ are usually strongly negatively correlated for both pure and modified Bayesian method, whereas the parameters α and ω are usually positively correlated. The parameters ω and κ show only weak negative correlation. The size of the correlation is slightly smaller for the modified Bayesian method.

	κ	α	ω	MCE_g	B_{pur}	B_{mod}
$\hat{\kappa}$	25	4	.02	.070	.113	−.003
			.04	.175	.492	−.028
		6	.02	.060	.121	.020
			.04	.163	.204	−.009
	50	4	.02	.072	.140	.012
			.04	.222	1.22	−.015
		6	.02	.004	.045	−.044
			.04	.194	.471	.011
$\hat{\alpha}$	25	4	.02	−.023	.007	.029
			.04	−.070	−.078	.125
		6	.02	−.013	.013	.026
			.04	−.084	−.028	.072
	50	4	.02	−.039	−.011	.036
			.04	−.085	−.282	.167
		6	.02	−.004	.003	.029
			.04	−.103	−.168	.071
$\hat{\omega}$	25	4	.02	−.060	.008	.017
			.04	−.099	−.049	.054
		6	.02	−.057	−.003	.001
			.04	−.081	−.004	.036
	50	4	.02	−.066	−.008	.015
			.04	−.098	−.122	.074
		6	.02	−.054	−.005	.003
			.04	−.087	−.056	.034

Table 2. Summary of simulation results—the relative bias of the estimates for the pure Bayesian method (B_{pur}) and the modified Bayesian method (B_{mod}) estimation compared with the minimum contrast estimation based on the pair correlation function g (MCE_g).

5. Discussion and conclusion

In this paper we described two Bayesian methods for parameter estimation of the Neyman-Scott point process and we made a comparative study of their results with the results of the minimum contrast estimation method based on the pair correlation function g.

The extensive simulation study shows that the pure Bayesian method is on average more precise (comparing mean square errors) than the simulation-free methods but in some cases it is much worse. On the other hand, it was shown that the introduced modified Bayesian method produces much more accurate results than other methods.

It improves the estimation of all parameters compared to the minimum contrast method, but it only improves the estimation of parameter κ compared to the pure Bayesian method.

Both methods are easily applicable for more general processes such as for the inhomogeneous [11] and shot noise Cox processes. Also it is possible to use the Bayesian estimation method for the log-Gaussian Cox processes [8], p. 200. Since the pure Bayesian method was the core procedure of the methods described in works [6] and [10], it would improve these methods if the modified Bayesian method was used.

The main advantages of the Bayesian methods include the fact that the posterior distribution is calculated together with the point estimates of the parameters and they are more precise than simulation-free methods. Furthermore, modifying the pure Bayesian method brings benefits to tuning the algorithm. It is more stable than the pure Bayesian method when choosing prior distributions and variances.

Subsequent analysis of posteriors indicates that there were no significant differences in the variance of the posterior distribution between the two methods, except for the parameter κ which the pure method is unable to deal with satisfactorily. No significant differences were detected when increasing the standard deviation h of each step. The analysis considered long chains of 500,000 steps. The comparison on one sample is in Figure 5.

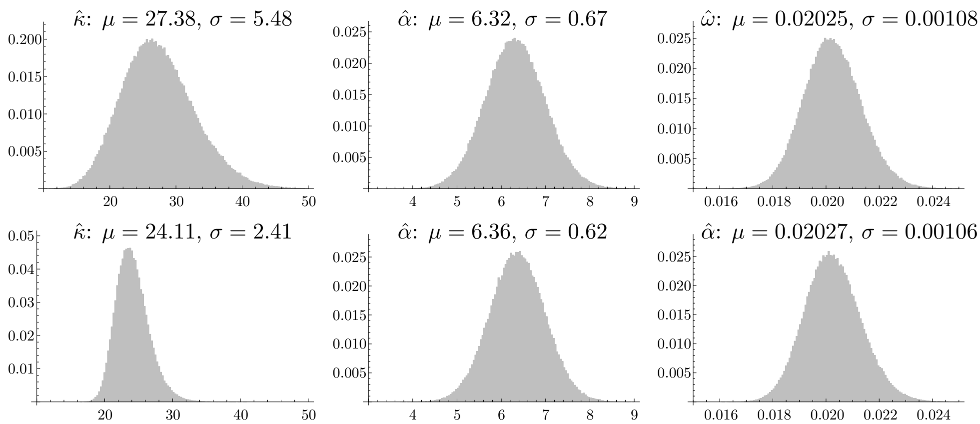

Figure 5. Posterior probability density histograms with the mean values and standard deviations of the pure method (first row) and the modified Bayesian method (second row) applied to the same sample.

Nevertheless, this approach is time-consuming. All the simulations were run on a computer with Intel Core i5 3.20 GHz. The average time of estimating a process ranges from 1 to 5 hours depending on the number of points and the type of the process. It is assumed that the total time required for estimation would be greatly reduced by implementing algorithms in C++.

We find interesting the problem of the inaccuracy of the pure Bayesian method and its improvement with the elimination of the estimation of the parameter κ in the algorithm. We explain this by the fact that the likelihood cannot easily identify a combination of the parameters α and κ, but it can easily identify one parameter when $\alpha \cdot \kappa$ is fixed. Further, the estimate of $\alpha \cdot \kappa$ used in the modified Bayesian method seems to be more efficient than the procedure contained in the pure Bayesian method.

A c k n o w l e d g e m e n t. We would like to thank Anonymous Referee for useful comments and detailed corrections, which we found very constructive and helpful to improve the quality of our paper.

References

[1] *A. Baddeley, R. Turner*: An R Package for analyzing spatial point patterns. J. Stat. Software *12* (2005), 1–42.

[2] *P. J. Diggle, R. J. Gratton*: Monte Carlo methods of inference for implicit statistical models. J. R. Stat. Soc., Ser. B *46* (1984), 193–227. zbl MR

[3] *J. Dvořák, M. Prokešová*: Moment estimation methods for stationary spatial Cox processes—a simulation study. Preprint (2011), `http://www.karlin.mff.cuni.cz/~dvorak/CoxEstimation/CoxEstimation_SimulationStudy.pdf`.

[4] *J. Dvořák, M. Prokešová*: Moment estimation methods for stationary spatial Cox processes—a comparison. Kybernetika *48* (2012), 1007–1026. zbl MR

[5] *Y. Guan*: A composite likelihood approach in fitting spatial point process models. J. Am. Stat. Assoc. *101* (2006), 1502–1512. zbl MR

[6] *P. Guttorp, T. L. Thorarinsdottir*: Bayesian inference for non-Markovian point processes. Advances and Challenges in Space-Time Modelling of Natural Events (J. M. Montero, M. Schlather, eds.). Lecture Notes in Statistics 207, Springer, Berlin, 2012, pp. 79–102.

[7] *J. Illian, A. Penttinen, H. Stoyan, D. Stoyan*: Statistical Analysis and Modelling of Spatial Point Patterns. Statistics in Practice. John Wiley & Sons, Chichester, 2008. zbl MR

[8] *J. Møller, R. P. Waagepetersen*: Statistical Inference and Simulation for Spatial Point Processes. Monographs on Statistics and Applied Probability 100, Chapman and Hall/CRC, Boca Raton, 2004. zbl MR

[9] *J. Møller, R. P. Waagepetersen*: Modern statistics for spatial point processes. Scand. J. Stat. *34* (2007), 643–684. zbl MR

[10] *T. Mrkvička*: Distinguishing different types of inhomogeneity in Neyman-Scott point processes. Methodol. Comput. Appl. Probab. *16* (2014), 385–395. zbl MR

[11] *T. Mrkvička, M. Muška, J. Kubečka*: Two step estimation for Neyman-Scott point process with inhomogeneous cluster centers. Stat. Comput. *24* (2014), 91–100. zbl MR

[12] *J. Neyman, E. L. Scott*: A theory for the spatial distribution of galaxies. Astrophys. J. *116* (1952), 144–163.

[13] *D. Stoyan, W. S. Kendall, J. Mecke*: Stochastic Geometry and Its Applications. Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, Chichester, 1995. zbl

[14] *U. Tanaka, Y. Ogata, D. Stoyan*: Parameter estimation and model selection for Neyman-Scott point processes. Biom. J. *50* (2008), 43–57. MR

[15] *Wolfram Research, Inc.*: Mathematica, Version 8.0. Champaign, IL (2010).

Authors' addresses: *Jiří Kopecký*, University of South Bohemia, Faculty of Education, Jeronýmova 10, 370 01 České Budějovice, Czech Republic, e-mail: `jkopecky@pf.jcu.cz`; *Tomáš Mrkvička*, University of South Bohemia, Faculty of Economics, Branišovská 31, 370 05 České Budějovice, Czech Republic, e-mail: `mrkvicka.toma@gmail.com`.