# NMST543
# Spatio-temporal clustering

Daniel Jahn
jahn@karlin.mff.cuni.cz

January 13, 2019

## 1 Introduction

## 2 Interaction and clustering

In order to understand the term *spatio-temporal clustering*, care has to be taken to carefully tease apart the related concepts of interaction and clustering.

We distinguish between spatial and temporal clustering. **Spatial clustering** results in the inhomogeneity of spatial distribution. An example is the crime rate being higher in densely populated areas. **Temporal clustering** results in the inhomogeneity of temporal distribution. An example are seasonal trends, such as an increased incidence of disease during winter.

Interaction is typically understood to be a mechanism through which the location of one or more points influences the locations of others. For example, certain tree species prefer to grow further from other trees, thus resulting in repulsion between the individual locations. In the temporal domain, we have e.g. the refractory period for neurons, where one neuron remains inactive for a small period after firing.

Both interaction and clustering result in inhomogeneity of the point patterns. From a statistical perspective, their difference is often purely interpretational, as they cannot distinguished through data only, but require domain knowledge to interpret the origin of the inhomogeneity.

These two terms blur together somewhat when we deal with data with both spatial and temporal component. A mere presence of both spatial and temporal clustering do not suffice for spatio-temporal clustering in the sense of the word as used in Diggle et al. [1995]. They give an example of a disease, which may cluster in densely populated areas and be more prevalent during cold weather. However, the disease only manifests spatio-temporal clustering if it is contagious. This is reflected in the fact that the terms interaction and clustering are used interchangeably in the paper.

# 3 Description of the method

In this text, we investigate the spatio-temporal clustering method proposed in Diggle et al. [1995]. Let $X$ be a **stationary** simple spatio-temporal point process on $\mathbb{R}^2 \times \mathbb{R}$. We only observe the events $X \cap (W \times [0,T])$, where $W$ is bounded Borel with $|W| > 0$ and $T > 0$. We define its projections to the spatial and temporal domain,

$$X_1 = \{x \in W : (x,t) \in X \cap (W \times [0,T])\}, \quad X_2 = \{t \in [0,T] : (x,t) \in X \cap (W \times [0,T])\},$$

and we assume $X_1$ and $X_2$ to be simple. We denote the intensity of $X$ by $\lambda$ and the intensities of $X_1$ and $X_2$ by $\lambda_1$ and $\lambda_2$, respectively. Note that $\lambda_1 = \lambda T$ and $\lambda_2 = \lambda |W|$.

We utilize the *K-function*, which can be defined by

$$\lambda K(s,t) = \mathrm{E} \sum_{(x,t_1),(y,t_2) \in X}^{\neq} \frac{1_A(x) 1_{[0,s]}(\|x-y\|) 1_S(t_1) 1_{[0,t]}(|t_1 - t_2|)}{\lambda |A||S|}$$

where $A \subset \mathbb{R}^2, S \subset \mathbb{R}$ are arbitrary Borel sets with finite and positive measure. We can interpret the term $\lambda K(s,t)$ as the number of further events occurring within distance $s$ and time $t$ of an arbitrary event.

Under the assumption of spatio-temporal independence, i.e. the independence of $X_1$ and $X_2$, we obtain the factorization

$$K(s,t) = K_1(s) K_2(t),$$

where $K_1$ and $K_2$ are the K-functions of $X_1$ and $X_2$, respectively. Analogously, their interpretation is the number of further events occurring withing distance $s$, respectively within time $t$, of an arbitrary event.

Not really defined

The K-functions serve as a basis for the method proposed by Diggle et al. [1995]. The functions $K, K_1, K_2$ will be used not only for testing for spatio-temporal interaction, but also to measure the extent and nature of it.

### 3.0.1 Estimates

Let $\{(x_i, t_i) : i = 1, \ldots, n\}$ denote the locations and times of all events within $W \times [0,T]$. We introduce the estimates

$$\hat{K}(s,t) = |A| T (n(n-1))^{-1} \sum_{j \neq i} w_{ij} v_{ij} I_{[d_{ij} \leq s]} I_{[u_{ij} \leq t]},$$

$$\hat{K}_1(s) = |A| (n(n-1))^{-1} \sum_{j \neq i} w_{ij} I_{[d_{ij} \leq s]},$$

$$\hat{K}_2(t) = T(n(n-1))^{-1} \sum_{j \neq i} v_{ij} I_{[u_{ij} \leq t]},$$

where $d_{ij} = \|x_i - x_j\|$ and $u_{ij} = |t_i - t_j|$, and $w_{ij}$ and $v_{ij}$ are edge-effect corrections.

We now introduce the three visual diagnostics which can be used to describe the extent and nature of the spatio-temporal interaction.

### 3.0.2 Diagnostic surface plot

$$\hat{D}(s,t) = \hat{K}(s,t) - \hat{K}_1(s)\hat{K}_2(t)$$

### 3.0.3 Residual plot

### 3.0.4 Monte-Carlo test

Interpretation

Intractable distribution. MC test.

# 4 Simulated data

# 5 Real data

# References

P.J. Diggle, A.G. Chetwynd, R. Häggkvist, and S.E. Morris. Second-order analysis of space-time clustering. *Statistical Methods in Medical Research*, 4(2):124–136, 1995. doi: 10.1177/096228029500400203.

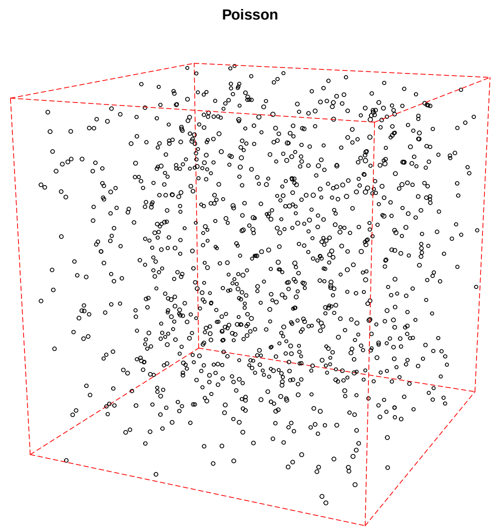


Figure 1: Realization of a Poisson point process. Intensity . Number of points: 1021

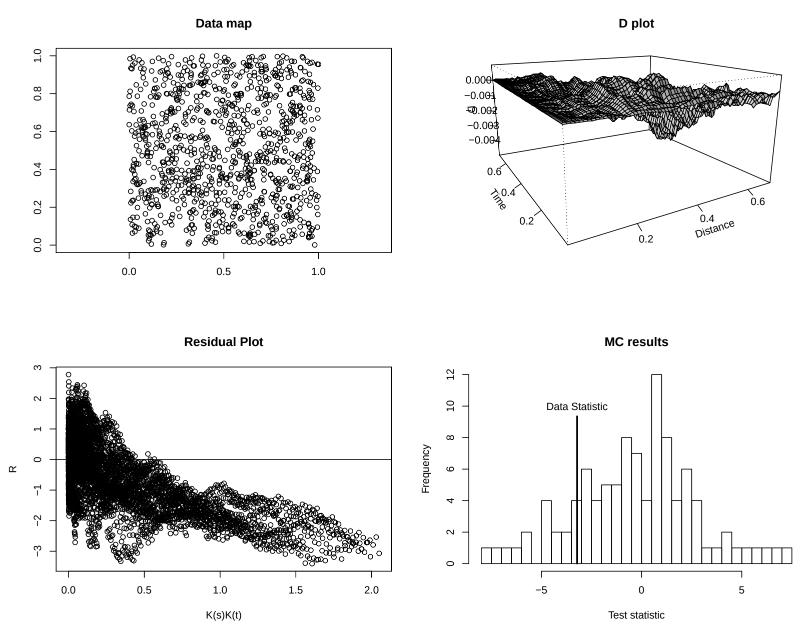


Figure 2: Diagnostic plots for a realization of a Poisson point process. Intensity . Number of points: 1021

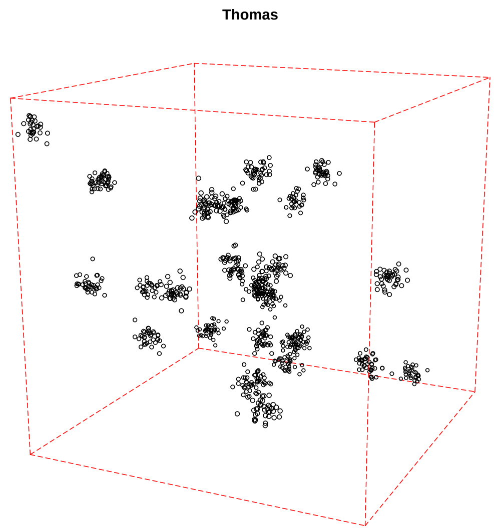


Figure 3: Realization of a Thomas point process. Intensity of parent process 20. Child process: intensity 40, distribution $N(0, 0.02)$. Number of points: 949.

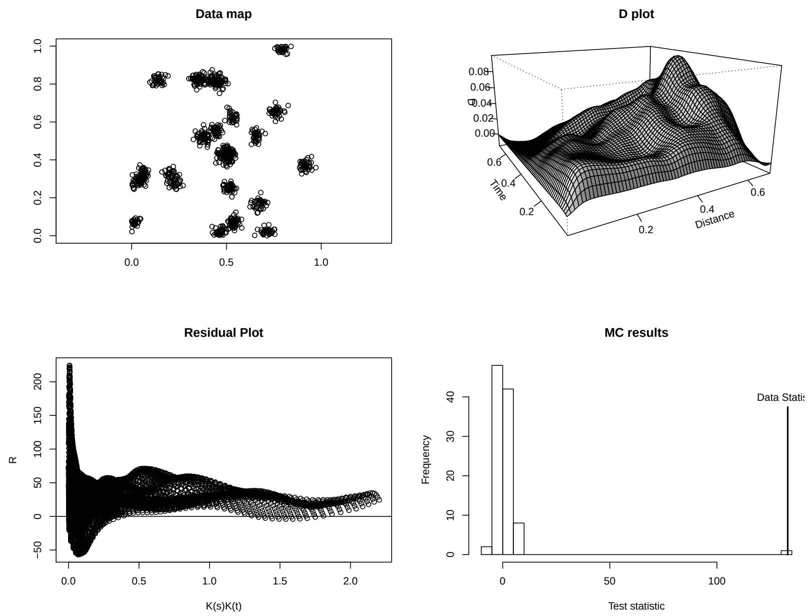


Figure 4: Diagnostic plots for a realization of a Thomas point process. Intensity of parent process 20. Child process: intensity 40, distribution $N(0, 0.02)$. Number of points: 949.

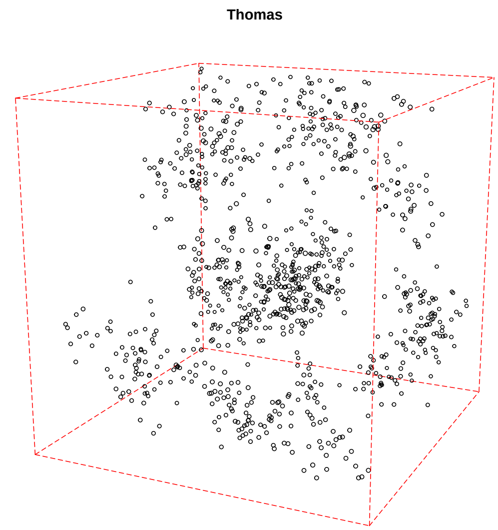


Figure 5: Realization of a Thomas point process. Intensity of parent process 40. Child process: intensity 20, distribution $N(0, 0.05)$. Number of points: 929.

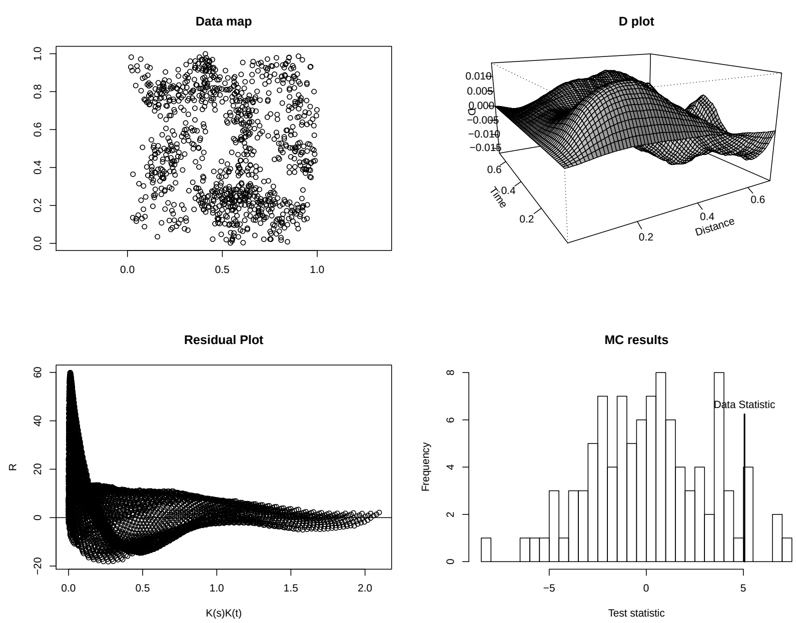


Figure 6: Diagnostic plots for a realization of a Thomas point process. Intensity of parent process 40. Child process: intensity 20, distribution $N(0, 0.05)$. Number of points: 929.

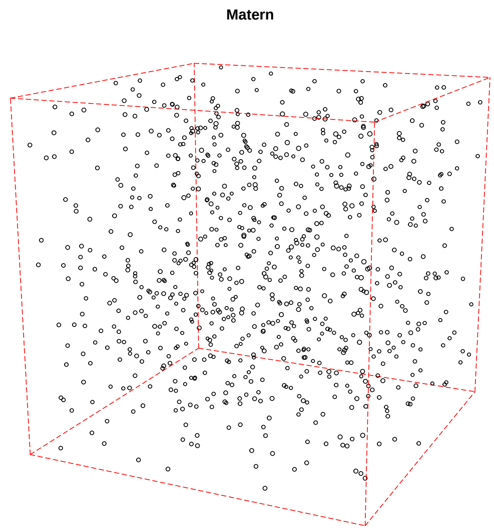


Figure 7: Realization of a Matern II point process. Intensity 1000. Minimum distance 0.05. Number of points: 818.

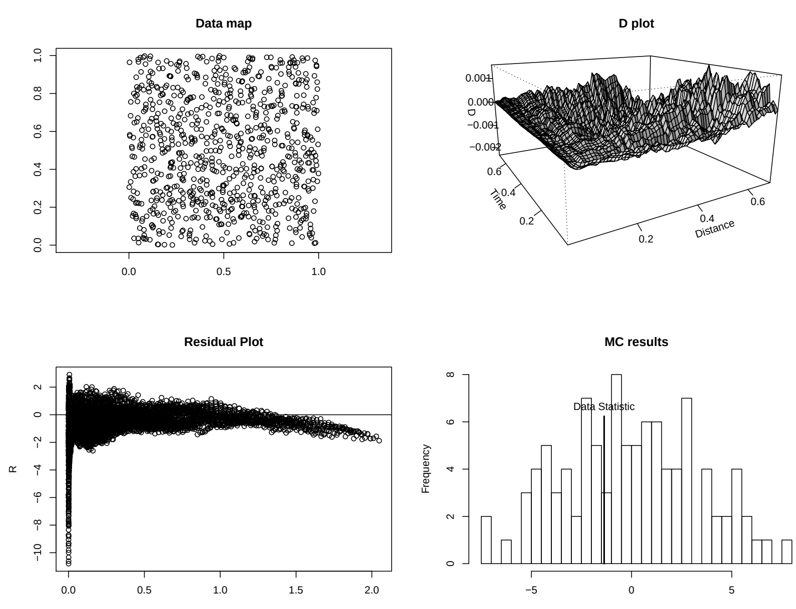


Figure 8: Diagnostic plots for a realization of a Matern II point process. Intensity 1000. Minimum distance 0.05. Number of points: 818.

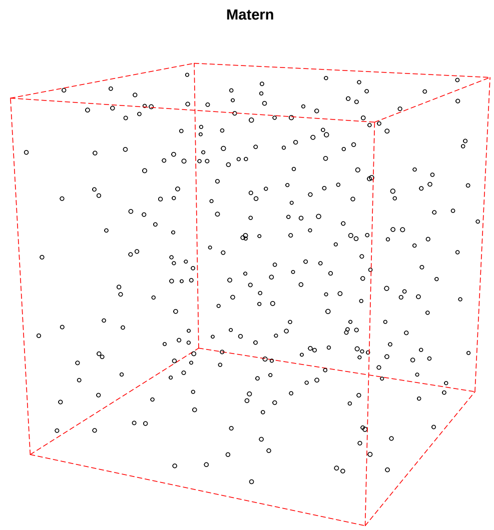


Figure 9: Realization of a Matern II point process. Intensity 1000. Minimum distance 0.1. Number of points: 275.

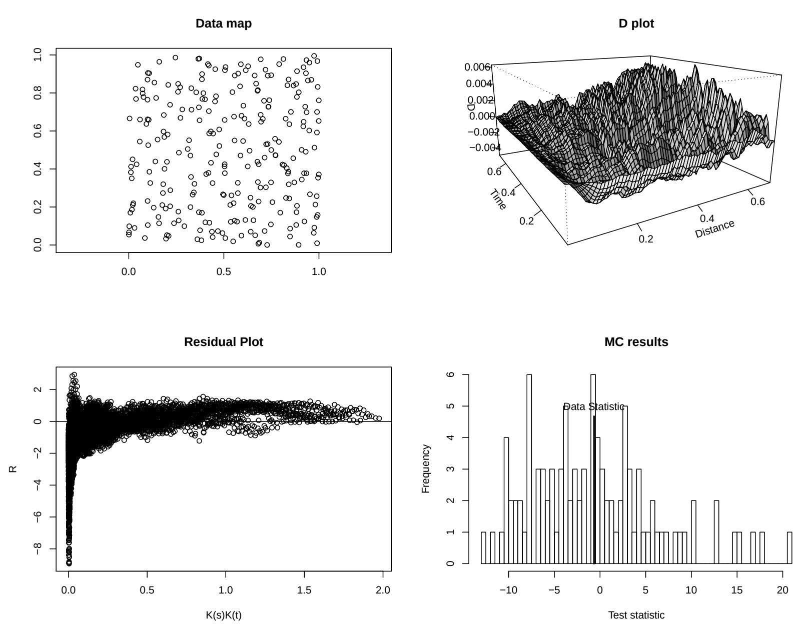


Figure 10: Diagnostic plots for a realization of a Matern II point process. Intensity 1000. Minimum distance 0.1. Number of points: 275.

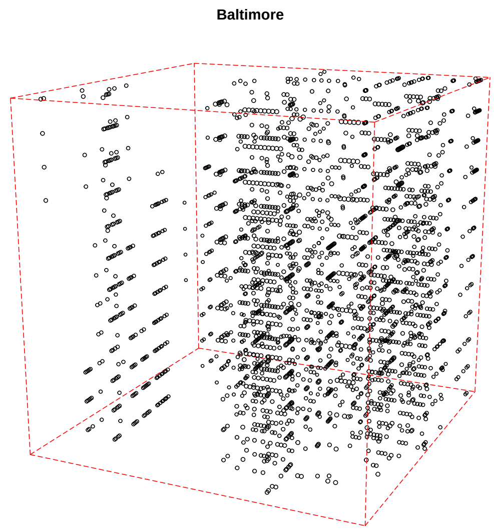


Figure 11: Vacant flats represented as a spatio-temporal point process. First two dimension indicate location in coordinates, third dimension indicates year when flat was vacant. Number of points: 2538.

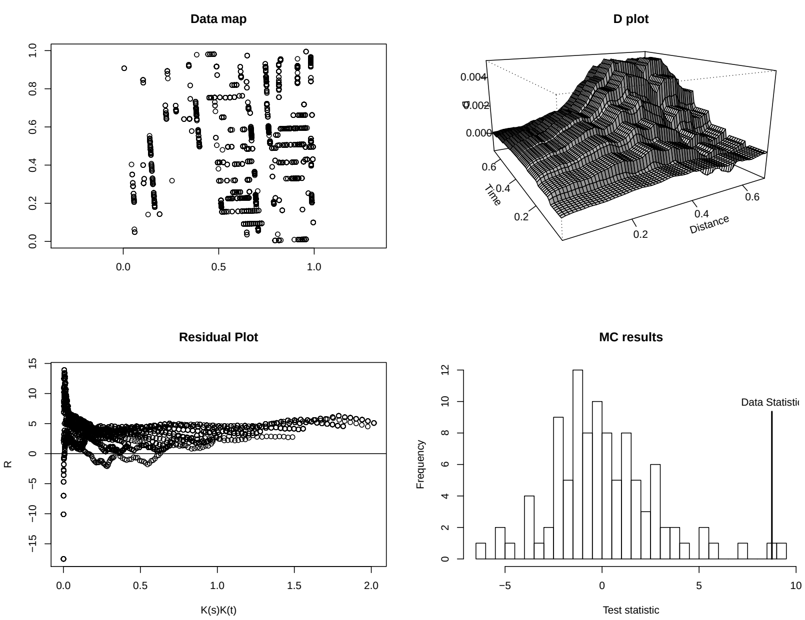


Figure 12: Diagnostic plots for vacant flats represented as a spatio-temporal point process. Number of points: 2538.

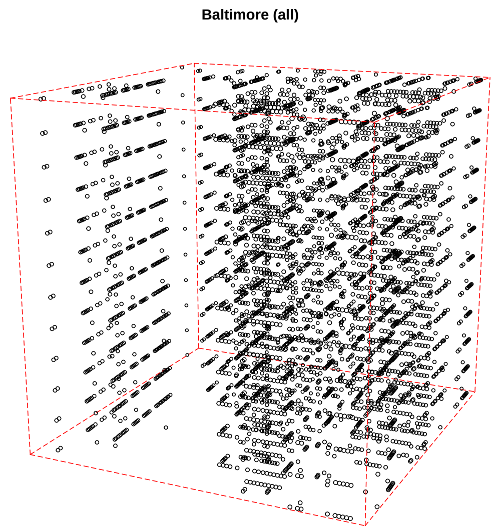


Figure 13: Spatio-temporal point process composed of locations of all flats repeated in each year, regardless of whether it was vacant or not. Number of points: 4872.

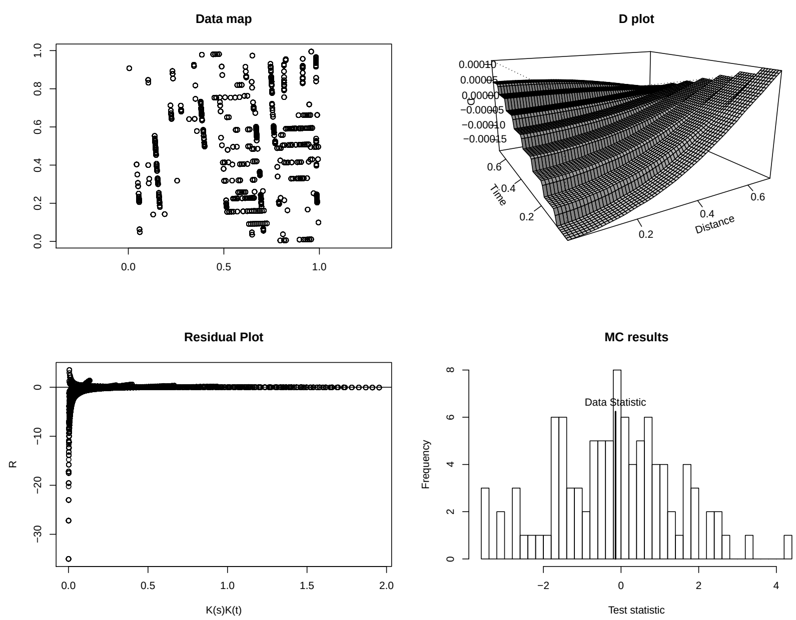


Figure 14: Diagnostic plots for repeated locations of flats. Number of points: 4872.