Fitting of Stochastic Telecommunication Network Models via Distance Measures and Monte–Carlo Tests

C. Gloaguen [1]   F. Fleischer [2]   H. Schmidt [3]   V. Schmidt [3]

14th November 2005

Abstract

We explore real telecommunication data describing the spatial geometrical structure of an urban region and we propose a model fitting procedure, where a given choice of different non–iterated and iterated tessellation models is considered and fitted to real data. This model fitting procedure is based on a comparison of distances between characteristics of sample data sets and characteristics of different tessellation models by utilizing a chosen metric. Examples of such characteristics are the mean length of the edge–set or the mean number of vertices per unit area. In particular, after a short review of a stochastic–geometric telecommunication model and a detailed description of the model fitting algorithm, we verify the algorithm by using simulated test data and subsequently apply the procedure to infrastructure data of Paris.

Keywords : Telecommunication network modelling, stochastic geometry, access network, random tessellations, statistical fitting, Monte–Carlo tests

[1]France Télécom R&D RESA/NET/NSO 92794 Issy Moulineaux Cedex 9, France

[2]Department of Applied Information Processing and Department of Stochastics, University of Ulm, 89069 Ulm, Germany

[3]Department of Stochastics, University of Ulm, 89069 Ulm, Germany

1 Introduction

Spatial stochastic models for telecommunication networks have been developed in recent years as an alternative to more traditional economical approaches to cost measurement and strategic planning. These models allow for incorporation of the stochastic *and* geometric features observed in telecommunication networks. By taking the geometric structure of network architectures into consideration, network models using tools of stochastic geometry offer a more relevant view to location-dependent network characteristics than conventional network models. The probabilistic setting reflects the network's variability in time and space.

Popular examples of networks where stochastic–geometric models have been considered so far are *switching networks*, *multi-cast networks*, and *mobile telecommunication systems*. These new models based on stochastic geometry include Poisson–Voronoi aggregated tessellations (Bacelli et al. (1996), Tchoumatchenko and Zuyev (2001)), superpositions of Poisson–Voronoi tessellations (Baccelli, Gloaguen and Zuyev (2000)), spanning trees (Bacccelli, Kofman and Rougier (1999), Baccelli and Zuyev (1996)), and coverage processes (Baccelli and Blaszczyszyn (2001)).

In the following, we focus on telecommunication access networks that can be regarded as the most important part of telecommunication network modelling, since roughly 50% of the total capital investment made in these networks is made in the access network. With such large investments at stake, and possibly evoluting subscriber populations, it is important to find appropriate models for cost evaluation, performance analysis, and strategic planning of access networks.

The *access network* or *local loop* is the part of the network connecting a subscriber to its corresponding Wire Center Stations (WCS). The hierarchical physical link is made via network components: a Network Interface Device (ND), secondary and primary cabinets (CS and CP) and a Service Area Station (SAI) as shown in Figure 1(a). A serving zone is associated to each WCS; the subnetwork gathering all the links between the WCS and the subcribers lying in its serving zone displays a tree structure (Figure 1(b)).

The most important feature about the access network is that it is the place where the telecommunication network fits in the town and country planning. For urban networks considered here this means the urban architecture and street system.

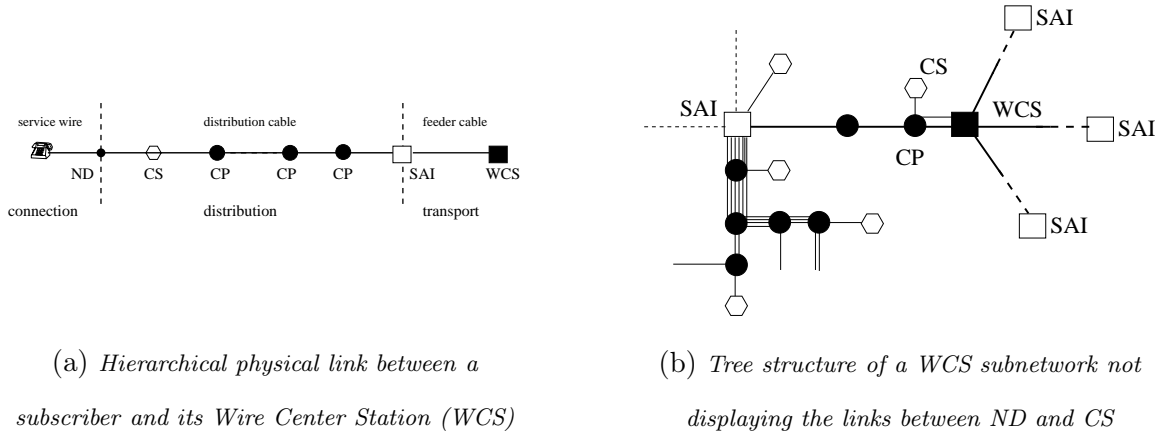Significant research in studying access networks has taken place in recent years with the so–

(a) *Hierarchical physical link between a subscriber and its Wire Center Station (WCS)*

(b) *Tree structure of a WCS subnetwork not displaying the links between ND and CS*

Figure 1: *Hierachical structure of access networks*

called *Stochastic Subscriber Line Model* (SSLM); see Gloaguen et.al. (2002) and Maier (2003). The SSLM is a stochastic–geometric model and offers tools in order to describe the spatial irregularity as well as the geometric features of access networks and allows for stochastic econometrical analysis, like the analysis of connection costs. Particularly, it provides simple mean value formulae for network characteristics used for cost evaluation, performance analysis, and strategic planning.

The modelling framework of the SSLM is subdivided into the *Network Geometry Model*, the *Network Component Model* and the *Network Topology Model*. The Network Geometry Model represents the cable trench system, which is located along the infrastructure system of a city or of a country. Random iterated tessellations (see e.g., Maier and Schmidt (2003)) can be used to describe this cable trench system. Subsequently the Network Component Model localizes the technical network components on the geometry using Poisson processes on lines or in the plane (Figure 2 (a)). To complete the picture, the Network Topology Model builds up the link between a subscriber and the corresponding WCS following the shortest path along the trench system (Figure 2 (b)).

Since the geometry of the infrastructure, i.e., the road system, is the basis of the access network, an important task is the choice of an appropriate tessellation model given simulated or real infrastructure data. In particular, three basic Poisson–type tessellation models are considered, out of which iterated tessellation models can be constructed. These basic models are called Poisson line tessellations (PLT), Poisson–Voronoi tessellations (PVT), and Poisson–Delaunay tessellations (PDT).

In the present paper, an approach for a model choice is presented, which is based on the
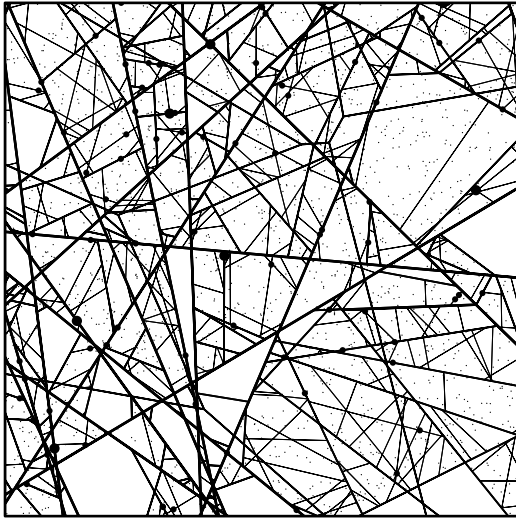
(a) *Realization of Network Geometry and Network Component Models*
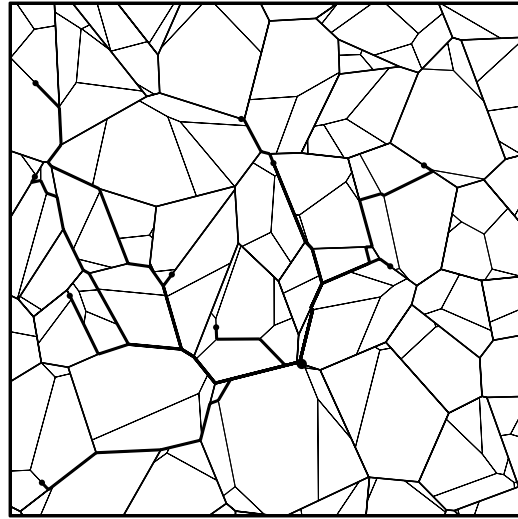
(b) *Shortest path analysis in the frame of the Network Topology Model*

Figure 2: *Realization of the Stochastic Subscriber Line Model*

minimization of distance measures between characteristics of input data and computed values of these characteristics using theoretical formulae valid for random tessellation models. Input data can be estimated characteristics both from real infrastructure data as well as from realizations of random tessellations. The latter part is important in order to verify the correctness of the model choice procedure.

In particular, in Section 2, a brief account of some basic notions of stochastic geometry is given and the theoretical tessellation models we are going to use are presented.

In Section 3, the model choice procedure is described. To compare input data and theoretical tessellation models, characteristics that describe the structural properties of the considered data are used. Examples of such characteristics are the expected number of vertices or the expected total length of the edges. Therefore, we need appropriate estimators for these characteristics first. Notice that, subsequent to the identification of the optimal model, this choice can be tested by using well–known Monte–Carlo test techniques (Stoyan and Stoyan (1994)). The section closes with numerical examples, where input data is derived from simulated realizations of random tessellations.

Finally, in Section 4, we consider real infrastructure data of Paris (see Figure 3). The data consist of line segments. Each line segment has an attached mark describing the type of road this segment belongs to. Hence for example, it is possible to distinguish between main roads

and side streets. A preprocessing of raw data is necessary in order to obtain a tessellation that consists of polygonal cells. Subsequently, it is possible to measure characteristics similar to those described above for simulated data.
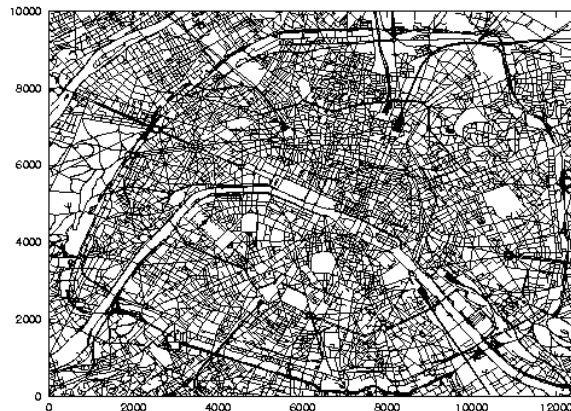

Figure 3: *Real infrastructure data of Paris*

Notice that, after having chosen an optimal model for the road system, the next logical step is to apply shortest paths algorithms to analyze connections between subscribers and their corresponding WCS–station. A short outlook at the end of this paper gives insight how such analysis can be performed, the results of which can be found in Gloaguen et al. (2005a, 2005b).

All programming work for the extensive simulation studies has been done using methods from the GeoStoch library. This JAVA–based library comprises software tools designated to analyze data with methods from stochastic geometry; see Mayer, Schmidt and Schweiggert (2004) and *http://www.geostoch.de.*

2   Mathematical background

In this section, the basic mathematical notation used in the present paper is introduced and a brief account of some relevant notions of stochastic geometry is given. Particularly, we put emphasis on the introduction of random (iterated) tessellations, which are used as models for the road system in the SSLM. For a detailed discussion of the mathematical background, it is referred to the literature, for example Schneider and Weil (2000) and Stoyan, Kendall and Mecke (1995). Further information about random (iterated) tessellations can also be found, e.g. in Maier and Schmidt (2003), Møller (1989), and Okabe et al. (2000).

2.1 Basic notations

The abbreviations int $B, \partial B$, and B^c are used to denote the interior, the boundary, and the complement of a set $B \subset \mathbb{R}^2$, respectively, where $\mathbb{R}^2$ denotes the 2–dimensional Euclidean space. Notice that by $|B|$ we denote the 2–dimensional Lebesgue measure for an arbitrary measurable set $B \in \mathbb{R}^2$, i.e. $|B|$ is the area of B.

The families of all closed sets, compact sets, and convex bodies (compact and convex sets) in $\mathbb{R}^2$ are denoted by $\mathcal{F}, \mathcal{K}$, and $\mathcal{C}$, respectively. Recall that a random closed set Ξ in $\mathbb{R}^2$ is a measurable mapping $\Xi : \Omega \to \mathcal{F}$ from some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ into the measurable space $(\mathcal{F}, \mathcal{B}(\mathcal{F}))$, where $\mathcal{B}(\mathcal{F})$ denotes the smallest σ–algebra of subsets of $\mathcal{F}$ that contains all sets $\{F \in \mathcal{F}, F \cap K \neq \emptyset\}$ for any $K \in \mathcal{K}$. Particularly, the random closed set Ξ is called a random compact set or a random convex body if $\mathbb{P}(\Xi \in \mathcal{K}) = 1$ or $\mathbb{P}(\Xi \in \mathcal{C}) = 1$, respectively.

2.2 Random tessellations

A *tessellation* in $\mathbb{R}^2$ is a countable family $\tau = \{C_n\}_{n \geq 1}$ of convex bodies $C_n \in \mathcal{C}$ such that int $C_n \neq \emptyset$ for all n, int $C_n \cap$ int $C_m = \emptyset$ for all $n \neq m$, $\bigcup_{n \geq 1} C_n = \mathbb{R}^2$, and $\sum_{n \geq 1} \mathbb{I}_{\{C_n \cap K \neq \emptyset\}} < \infty$ for any $K \in \mathcal{K}$. Notice that the sets C_n, called the *cells* of τ, are polygons in $\mathbb{R}^2$. The family of all tessellations in $\mathbb{R}^d$ is denoted by $\mathcal{T}$. A *random tessellation* $\{\Xi_n\}_{n \geq 1}$ in $\mathbb{R}^d$ is a sequence of random convex bodies Ξ_n such that $\mathbb{P}(\{\Xi_n\}_{n \geq 1} \in \mathcal{T}) = 1$. Notice that a random tessellation $\{\Xi_n\}_{n \geq 1}$ can also be considered as a marked point process $\sum_{n \geq 1} \delta_{[\alpha(\Xi_n), \Xi_n^0]}$, where $\alpha : \mathcal{C}' \to \mathbb{R}^d$, $\mathcal{C}' = \mathcal{C} \setminus \{\emptyset\}$, is a measurable mapping such that $\alpha(C) \in C$ and $\alpha(C + x) = \alpha(C) + x$ for any $C \in \mathcal{C}'$ and $x \in \mathbb{R}^d$, and where $\Xi_n^0 = \Xi_n - \alpha(\Xi_n)$ is the centered cell corresponding to Ξ_n which contains the origin. The point $\alpha(C) \in \mathbb{R}^d$ is called the associated point of C and can be chosen, for example, to be the lexicographically smallest point of C.

2.3 Examples of non–iterated random tessellations

Figure 4 shows realizations of our three basic non–iterated tessellation models, namely the PLT, the PVT, and the PDT.

The cells of a (deterministic) Voronoi tessellation are convex polygons in $\mathbb{R}^2$, namely the closure of all planar points which are closest (in the sense of the 2-dimensional Euclidean
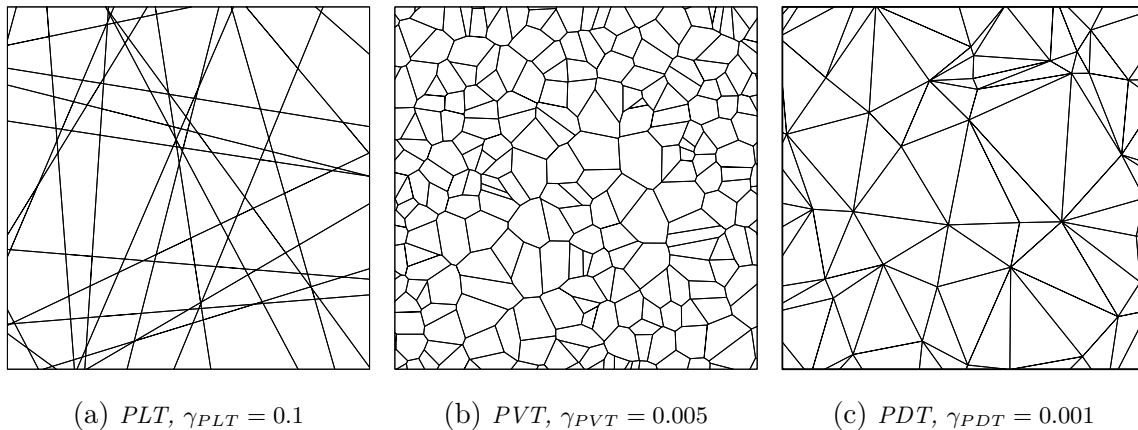
(a) *PLT*, $\gamma_{PLT} = 0.1$ (b) *PVT*, $\gamma_{PVT} = 0.005$ (c) *PDT*, $\gamma_{PDT} = 0.001$

Figure 4: *Realizations of basic tessellations and corresponding intensity γ*

metric) to the nucleus of this cell. If the set of nuclei is induced by a stationary random Poisson point process, we call the resulting tessellation a (random) PVT. The intensity γ_{PVT} of the PVT corresponds to the intensity of its generating Poisson point process and can be interpreted as the mean number of cells per unit area.

The PDT is closely related to the PVT. Indeed, consider a PVT, i.e., a Voronoi tessellation whose nuclei are induced by a stationary Poisson point process. The cells of its corresponding PDT are obtained by connecting the nuclei of neighboring cells, i.e., cells that share a common edge, of the PVT. Since in the case of a PVT the nuclei are not collinear with probability one, i.e., almost surely three pairwise different points do not lie on the same line, the cells of the corresponding PDT are triangles. The intensity γ_{PDT} can be interpreted as the mean number of vertices of the PDT per unit area.

The PLT is induced by a random Poisson line process in $\mathbb{R}^2$ and can be interpreted as a marked point process on $\mathbb{R} \times [0, \pi]$. This is due to the fact that each line is determined by the signed perpendicular distance between the line and the origin o and by the angle measured in anti–clockwise direction between the orientation vector of the line and the x–axis. Generally, the intensity γ_{PLT} is interpreted as the mean total length of edges per unit area.

Let $\lambda_1, \ldots, \lambda_4$ denote the mean number of vertices, the mean number of edges, the mean number of cells, and the mean total length of edges per unit area, respectively. Table 1 shows the relationship between these four intensities and a tessellation with intensity γ; see e.g. Stoyan, Kendall and Mecke (1995). In each case γ has to be interpreted differently.

Non–iterated tessellations serve as starting point to construct more refined tessellation mod-

Table 1: *Values of $\lambda_1, \ldots, \lambda_4$ for a given tessellation with intensity γ*

Tessellation	λ_1	λ_2	λ_3	λ_4
PLT	$\frac{1}{\pi}\gamma^2$	$\frac{2}{\pi}\gamma^2$	$\frac{1}{\pi}\gamma^2$	γ
PVT	2γ	3γ	γ	$2\sqrt{\gamma}$
PDT	γ	3γ	2γ	$\frac{32}{3\pi}\sqrt{\gamma}$

els, so–called iterated tessellations. In Section 2.4 we give a mathematical definition of such tessellations and in Section 2.5 we consider examples of so called 1–fold nestings. This means that each cell of some initial tessellation is further tessellated using a certain tessellation model, however not necessarily the same model as in the case of the initial tessellation.

2.4   Random iterated tessellations

A (deterministic) iterated tessellation $\tau = \{C_{n\nu} \cap C_n : \text{int } C_{n\nu} \cap \text{int } C_n \neq \emptyset\}$ in $\mathbb{R}^d$ consists of an initial tessellation $\tau = \{C_n\}_{n\geq 1}$ in $\mathbb{R}^d$ and a sequence $(\tau_n)_{n\geq 1}$ of component tessellations $\tau_n = \{C_{n\nu}\}_{\nu\geq 1}$. Hence, in order to define the notion of a random iterated tessellation, we can proceed as follows. Let Ξ be a random convex body in $\mathbb{R}^d$, where int $\Xi \neq \emptyset$, and let $X = \{\Xi_n\}_{n\geq 1}$ be a random tessellation in $\mathbb{R}^d$. Then, the mapping $Y(\cdot \mid \Xi) : \Omega \to N(\mathcal{F}')$ defined by $Y(B \mid \Xi) = \sum_{n\geq 1} \delta_{\Xi_n \cap \Xi}(B) \, \mathbb{1}_{\{\text{int } \Xi_n \cap \text{int } \Xi \neq \emptyset\}}$ for $B \in \mathcal{B}(\mathcal{F}')$ is a point process in $\mathcal{C}'$, where $\mathcal{F}' = \mathcal{F} \setminus \{\emptyset\}$. The space of all non–negative and integer–valued measures on $\mathcal{B}(\mathcal{F}')$ is denoted by $N(\mathcal{F}')$, where each $\eta \in N(\mathcal{F}')$ can be represented by a finite or countable sum of Dirac measures δ_F of sets $F \in \mathcal{F}'$, i.e., $\eta(B) = \sum_{n\geq 1} \eta(\{F_n\}) \delta_{F_n}(B)$ for any $B \in \mathcal{B}(\mathcal{F}')$, and that $\eta(\{F \in \mathcal{F} : F \cap K \neq \emptyset\}) < \infty$ for any $K \in \mathcal{K}$. Notice that $Y(\cdot \mid \Xi)$ can be seen as one possible way to describe a random tessellation in Ξ.

Furthermore, if $X = \{\Xi_n\}_{n\geq 1}$ is an arbitrary random tessellation in $\mathbb{R}^d$ and if $\{X_n\}_{n\geq 1}$ is an independent sequence of independent and identically distributed random tessellations $X_n = \{\Xi_{n\nu}\}_{\nu\geq 1}$ in $\mathbb{R}^d$, then the mapping $Y : \Omega \to N(\mathcal{F}')$ defined by $Y(B) = \sum_n Y_n(B \mid \Xi_n)$ and $Y_n(B \mid \Xi_n) = \sum_{\nu\geq 1} \delta_{\Xi_{n\nu} \cap \Xi_n}(B) \, \mathbb{1}_{\{\text{int } \Xi_{n\nu} \cap \text{int } \Xi_n \neq \emptyset\}}$ for $B \in \mathcal{B}(\mathcal{F}')$ is called the point–process representation of an iterated random tessellation (or X/X_n–nesting) in $\mathbb{R}^d$ with initial tessellation X and component tessellations $X_1, X_2, \ldots$. Clearly, the point process Y is stationary and isotropic, respectively, provided that both the initial tessellation X and the component tessellations $X_1, X_2, \ldots$ possess these properties.

2.5 Examples of iterated random tessellations

An iterated random tessellation X can itself be an initial tessellation for a further tessellation. In particular, it is possible to construct random tessellations with $k \in \mathbb{N}_0$ iterations which are called k–fold iterated tessellations. For example, a X_0/X_1 tessellation denotes a 1-fold iterated tessellation, which can be described by the two corresponding intensity parameters γ_0 and γ_1, respectively. Trivially, each non–iterated tessellation can be regarded as a 0–fold iterated tessellation. In the case of 1–fold iterated tessellations, if both for X_0 and X_1 PLTs, PVTs, or PDTs are considered, we end up with nine possible models; see Figures 5 to 7.

(a) *PLT/PLT,*

$\gamma_0 = 0.05, \gamma_1 = 0.1$

(b) *PLT/PVT,*

$\gamma_0 = 0.05, \gamma_1 = 0.005$

(c) *PLT/PDT,*

$\gamma_0 = 0.05, \gamma_1 = 0.001$
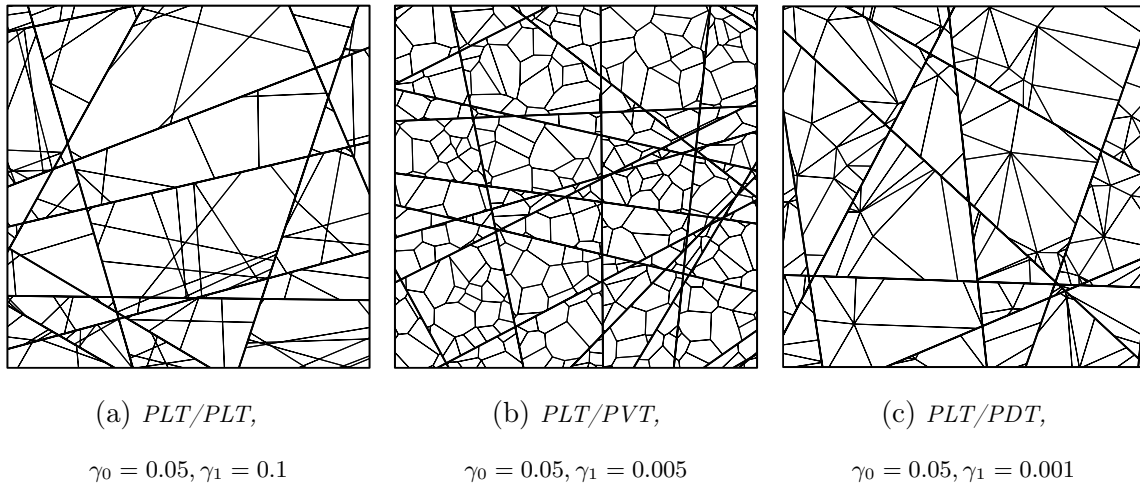
Figure 5: *Realizations of 1–fold X_0/X_1–tessellations with intensities γ_0 and γ_1 and initial PLT*

(a) *PVT/PLT,*

$\gamma_0 = 0.0005, \gamma_1 = 0.1$

(b) *PVT/PVT,*

$\gamma_0 = 0.0005, \gamma_1 = 0.001$

(c) *PVT/PDT,*

$\gamma_0 = 0.0005, \gamma_1 = 0.001$

Figure 6: *Realizations of 1–fold X_0/X_1–tessellations with intensities γ_0 and γ_1 and initial PVT*

The so–called Bernoulli thinning (see Figure 8) allows for variants of k–fold tessellations. For example in the context of infrastructure modelling in urban areas this can be used to

(a) *PDT/PLT,*

$\gamma_0 = 0.0001, \gamma_1 = 0.1$

(b) *PDT/PVT,*

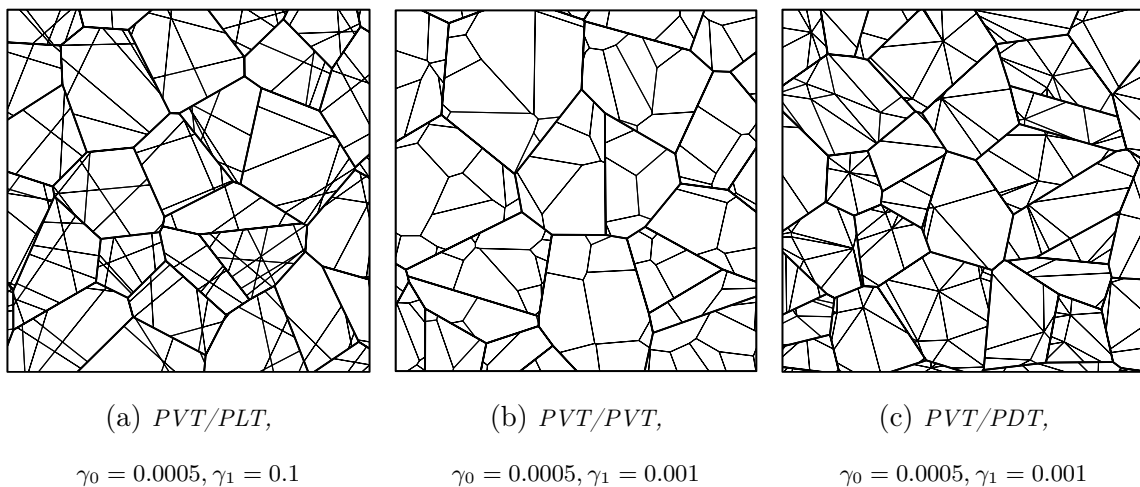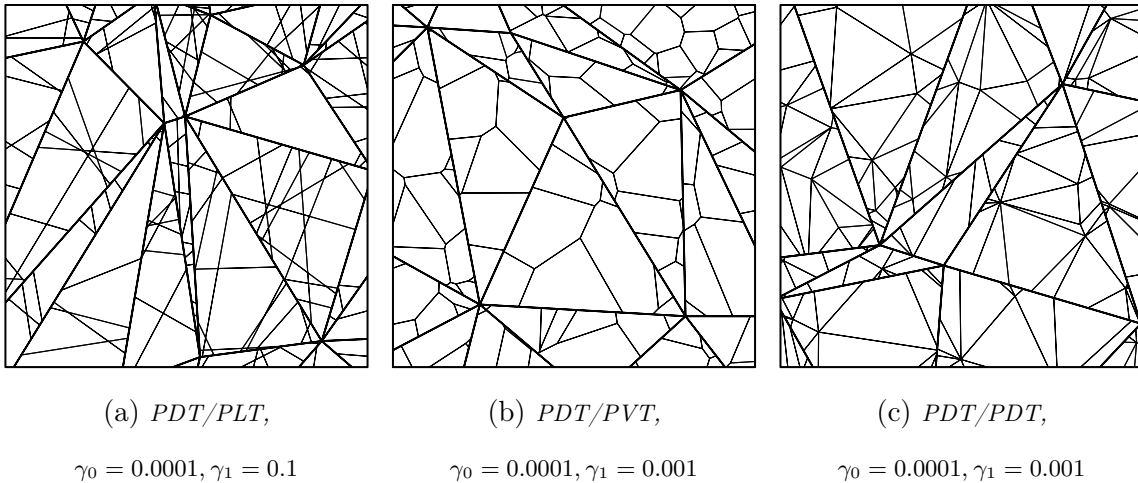$\gamma_0 = 0.0001, \gamma_1 = 0.001$

(c) *PDT/PDT,*

$\gamma_0 = 0.0001, \gamma_1 = 0.001$

Figure 7: *Realizations of 1–fold X_0/X_1 tessellations with intensities γ_0 and γ_1 and initial PDT*

consider graveyards or parks. In such a case there are cells of X_0 which are not iterated further by X_1. More generally, the nth cell of an initial tessellation X_0 is subdivided by a member of a given finite family $\{X_{1,n}, \ldots, X_{s,n}\}$ of component tessellations $X_{l,n}$, $1 \leq l \leq s$, where each of them has a certain probability of being selected. If no component tessellation is selected, the corresponding cell is not further iterated. Such an iterated random tessellation is called *clustered iterated random tessellation* or *multi type nesting* in $\mathbb{R}^2$. It is denoted by $X_0/(p_1 X_{1,1}, \ldots, p_s X_{1,s})$ with non–negative weights $p_1, \ldots, p_s$ satisfying $p_1 + \ldots + p_s \leq 1$ and random tessellations $X_{1,n}, \ldots, X_{s,n}$ which are independent for each $n \in \mathbb{N}$. Figure 8 displays realizations of 1–fold nestings, where the Bernoulli thinning technique has been applied in the case of a $X_0/p\,X_1$–nesting ($p \in [0, 1]$).

(a) *PLT/PVT (p = 75%),*

$\gamma_0 = 0.05, \gamma_1 = 0.005$

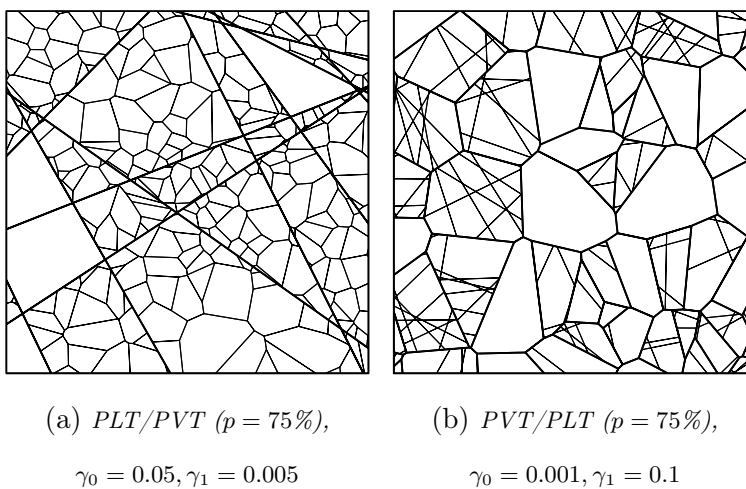(b) *PVT/PLT (p = 75%),*

$\gamma_0 = 0.001, \gamma_1 = 0.1$

Figure 8: *Realizations of 1–fold tessellations with Bernoulli thinning*

Similar to Section 2.3, mean value relationship can be obtained for iterated tessellations; see e.g. Maier (2003) and the references therein. Consider the case of a 1–fold $X_0/p\, X_1$–nesting ($p \in [0,1]$) and let again $\lambda_1, \ldots, \lambda_4$ denote the mean number of vertices, the mean number of edges, the mean number of cells, and the mean total length of edges per unit area, respectively, however with respect to the 1–fold tessellation. Let $\lambda_1^{(0)}, \ldots, \lambda_4^{(0)}$ and $\lambda_1^{(1)}, \ldots, \lambda_4^{(1)}$ denote the corresponding characteristics of X_0 and of X_1, respectively. Then,

$$
\begin{aligned}
\lambda_1 &= \lambda_1^{(0)} + p\lambda_1^{(1)} + \frac{4p}{\pi} \lambda_4^{(0)} \lambda_4^{(1)}, \\
\lambda_2 &= \lambda_2^{(0)} + p\lambda_2^{(1)} + \frac{6p}{\pi} \lambda_4^{(0)} \lambda_4^{(1)}, \\
\lambda_3 &= \lambda_3^{(0)} + p\lambda_3^{(1)} + \frac{2p}{\pi} \lambda_4^{(0)} \lambda_4^{(1)}, \\
\lambda_4 &= \lambda_4^{(0)} + p\lambda_4^{(1)}.
\end{aligned}
$$

Table 2 shows the dependence of the four characteristics $\lambda_1, \ldots, \lambda_4$ on p and on the intensities γ_0 and γ_1 of X_0 and X_1, respectively.

Notice that the case of a X_0/X_1–nesting (i.e., $p = 1$) is degenerate in the sense that a symmetry can be observed in the intensities γ_0 and γ_1 of X_0 and X_1, respectively. The four characteristics alone cannot be used to discriminate between PVT/PLT and PLT/PVT, between PLT/PDT and PDT/PLT, and between PVT/PDT and PDT/PVT.

3 Model choice based on comparison of distance measures

In the present section we introduce our model choice algorithm. After a description of the procedure itself, it is verified using simulated input data.

3.1 Characteristics of input data

We assume that we observe our input data through a rectangular sampling window W. The input data are either simulated realizations of tessellations or (possibly preprocessed) real infrastructure data. The observed input data are then used to estimate certain characteristics describing the spatial–geometric structure of the input data. In particular, we consider characteristics which are measured per unit area. Popular examples comprise the characteristics $\lambda_1, \ldots, \lambda_4$, which correspond to the mean number of vertices, the mean number of edges,

Table 2: *Mean–value formulae for X_0/pX_1–tessellations*

	PLT/PLT	PLT/PVT	PLT/PDT
λ_1	$\frac{1}{\pi}\gamma_0^2 + \frac{1}{\pi}p\gamma_1^2 + \frac{4}{\pi}p\gamma_0\gamma_1$	$\frac{1}{\pi}\gamma_0^2 + 2p\gamma_1 + \frac{8}{\pi}p\gamma_0\sqrt{\gamma_1}$	$\frac{1}{\pi}\gamma_0^2 + p\gamma_1 + \frac{128}{3\pi^2}p\gamma_0\sqrt{\gamma_1}$
λ_2	$\frac{2}{\pi}\gamma_0^2 + \frac{2}{\pi}p\gamma_1^2 + \frac{6}{\pi}p\gamma_0\gamma_1$	$\frac{2}{\pi}\gamma_0^2 + 3p\gamma_1 + \frac{12}{\pi}p\gamma_0\sqrt{\gamma_1}$	$\frac{2}{\pi}\gamma_0^2 + 3p\gamma_1 + \frac{64}{\pi^2}p\gamma_0\sqrt{\gamma_1}$
λ_3	$\frac{1}{\pi}\gamma_0^2 + \frac{1}{\pi}p\gamma_1^2 + \frac{2}{\pi}p\gamma_0\gamma_1$	$\frac{1}{\pi}\gamma_0^2 + p\gamma_1 + \frac{4}{\pi}p\gamma_0\sqrt{\gamma_1}$	$\frac{1}{\pi}\gamma_0^2 + 2p\gamma_1 + \frac{64}{3\pi^2}p\gamma_0\sqrt{\gamma_1}$
λ_4	$\gamma_0 + p\gamma_1$	$\gamma_0 + 2p\sqrt{\gamma_1}$	$\gamma_0 + \frac{32}{3\pi}p\sqrt{\gamma_1}$

	PVT/PLT	PVT/PVT	PVT/PDT
λ_1	$\frac{1}{\pi}p\gamma_1^2 + 2\gamma_0 + \frac{8}{\pi}p\gamma_1\sqrt{\gamma_0}$	$2(\gamma_0 + p\gamma_1) + \frac{16}{\pi}p\sqrt{\gamma_0\gamma_1}$	$2\gamma_0 + p\gamma_1 + \frac{256}{3\pi^2}p\sqrt{\gamma_0\gamma_1}$
λ_2	$\frac{2}{\pi}p\gamma_1^2 + 3\gamma_0 + \frac{12}{\pi}p\gamma_1\sqrt{\gamma_0}$	$3(\gamma_0 + p\gamma_1) + \frac{24}{\pi}p\sqrt{\gamma_0\gamma_1}$	$3(\gamma_0 + p\gamma_1) + \frac{128}{\pi^2}p\sqrt{\gamma_0\gamma_1}$
λ_3	$\frac{1}{\pi}p\gamma_1^2 + \gamma_0 + \frac{4}{\pi}p\gamma_1\sqrt{\gamma_0}$	$\gamma_0 + p\gamma_1 + \frac{8}{\pi}p\sqrt{\gamma_0\gamma_1}$	$\gamma_0 + 2p\gamma_1 + \frac{128}{3\pi^2}p\sqrt{\gamma_0\gamma_1}$
λ_4	$p\gamma_1 + 2\sqrt{\gamma_0}$	$2(\sqrt{\gamma_0} + p\sqrt{\gamma_1})$	$2\sqrt{\gamma_0} + \frac{32}{3\pi}p\sqrt{\gamma_1}$

	PDT/PLT	PDT/PVT	PDT/PDT
λ_1	$\frac{1}{\pi}p\gamma_1^2 + \gamma_0 + \frac{128}{3\pi^2}p\gamma_1\sqrt{\gamma_0}$	$2p\gamma_1 + \gamma_0 + \frac{256}{3\pi^2}p\sqrt{\gamma_1\gamma_0}$	$\gamma_0 + p\gamma_1 + \frac{4096}{9\pi^3}p\sqrt{\gamma_0\gamma_1}$
λ_2	$\frac{2}{\pi}p\gamma_1^2 + 3\gamma_0 + \frac{64}{\pi^2}p\gamma_1\sqrt{\gamma_0}$	$3(p\gamma_1 + \gamma_0) + \frac{128}{\pi^2}p\sqrt{\gamma_1\gamma_0}$	$3(\gamma_0 + p\gamma_1) + \frac{2048}{3\pi^3}p\sqrt{\gamma_0\gamma_1}$
λ_3	$\frac{1}{\pi}p\gamma_1^2 + 2\gamma_0 + \frac{64}{3\pi^2}p\gamma_1\sqrt{\gamma_0}$	$p\gamma_1 + 2\gamma_0 + \frac{128}{3\pi^2}p\sqrt{\gamma_1\gamma_0}$	$2(\gamma_0 + p\gamma_1) + \frac{2048}{9\pi^3}p\sqrt{\gamma_1\gamma_0}$
λ_4	$p\gamma_1 + \frac{32}{3\pi}\sqrt{\gamma_0}$	$2p\sqrt{\gamma_1} + \frac{32}{3\pi}\sqrt{\gamma_0}$	$\frac{32}{3\pi}\left(\sqrt{\gamma_0} + p\sqrt{\gamma_1}\right)$

the mean number of cells, and the mean total length of edges with respect to the unit area, respectively.

These characteristics can be interpreted as global characteristics and they are chosen both because they are a good representation of the underlying tessellation model and because of their relative simplicity regarding theoretical formulae; see Maier and Schmidt (2003) and Tables 1 and 2. Beyond that, one could also consider local characteristics which refer to single cells, like the mean edge–length per cell, the mean perimeter per cell, and the mean area per cell. However, it turns out that these characteristics are less useful, because unbiased estimators for them are not obvious. Therefore, we concentrate on the global characteristics

in the following descriptions and consider the vector

$$\lambda = (\lambda_1, \ldots, \lambda_4)^\top . \tag{3.1}$$

3.2 Unbiased estimators

The intensities of the vector λ given by (3.1) have to be estimated from the input data. Therefore, we need a vector of (intensity) estimators

$$\widehat{\lambda} = (\widehat{\lambda}_1, \ldots, \widehat{\lambda}_4)^\top , \tag{3.2}$$

where each entry of this vector is an estimator for the corresponding entry in (3.1). Further information about estimation of such characteristics can be found in literature, for example Baddeley and Jensen (2004) and Ohser and Mücklich (2000) as well as in the references therein. The vector estimators used in the course of this paper are

$$\widehat{\lambda} = \frac{1}{|W|} \, (n_v \, , n_e \, , n_c \, , l_e \,)^\top , \tag{3.3}$$

Clearly, with n_v denoting the number of vertices contained within the sampling window W, the estimator $\widehat{\lambda}_1$ is an unbiased estimator for λ_1. In order to get estimates for λ_2, it is often suggested to consider the number n_e of edges whose lexicographically smaller end point is contained in W. Alternatively, in case of a rectangular sampling window W, a similar estimator is obtained if n_e counts all edges completely within W and the edges which intersect with the upper and right boundary of W. Similarly, in the formula for the estimator $\widehat{\lambda}_3$, n_c denotes the number of cells obtained by counting an associated point of the cells, for example the lexicographically smallest vertice of each cell. Alternatively, again in case of a rectangular sampling window W, n_c may count the cells completely within W and the cells which intersect exclusively the upper and/or right boundary of W. Finally, $\widehat{\lambda}_4$ is an unbiased estimator for λ_4 if l_e measures the total length of the edge–set contained in W.

3.3 Distance measures

In order to compare the estimated vector of characteristics of the input data with the corresponding vector of calculated values for the tessellation models under comparison, we have to consider different distance measures.

However, good choices for such measures are far from being obvious, hence several possibilities have to be examined. Particularly, if $x = (x_1, \ldots, x_n)^\top$ and $y = (y_1, \ldots, y_n)^\top$ denote two vectors with n entries, the following metrics have been taken into account.

Euclidean distance

$$d_e(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad \text{(absolute)} \qquad d'_e(x, y) = \sqrt{\sum_{i=1}^{n}\left(\frac{x_i - y_i}{x_i}\right)^2} \quad \text{(relative)}$$

absolute–value distance

$$d_a(x, y) = \sum_{i=1}^{n}|x_i - y_i| \quad \text{(absolute)} \qquad d'_a(x, y) = \sum_{i=1}^{n}\left|\frac{x_i - y_i}{x_i}\right| \quad \text{(relative)}$$

maximum–norm distance

$$d_m(x, y) = \max_{i=1,\ldots,n}|x_i - y_i| \quad \text{(absolute)} \qquad d'_m(x, y) = \max_{i=1,\ldots,n}\frac{|x_i - y_i|}{x_i} \quad \text{(relative)}$$

Notice that the absolute distance measures d_e, d_a, and d_m can be influenced strongly by single components with possibly extreme values, whereas relative measures like d'_e, d'_a, and d'_m should be preferable for our purposes since a certain effect of averaging occurs and since they are scale–invariant. Furthermore notice that the relative distance measures are not symmetric in their arguments x and y anymore. Therefore, it is necessary to handle distance measures of this kind with care in the subsequent examinations. This means that if we use a certain relative distance measure, the scaling always needs to be done with respect to the same reference argument. This applies throughout the whole paper both in case of the minimization procedure and in the case of Monte–Carlo tests.

3.4 Optimal model choice

In this section we describe how an optimal tessellation model τ^* (and corresponding optimal intensities) is obtained with respect to a chosen distance measure. In particular, we consider k–fold tessellations with k being either 0 (i.e., non–iterated tessellations) or 1 (i.e., tessellation consisting of an initial tessellation X_0 and a component tessellation X_1).

Let $k = 0$. Then we are dealing with a PLT, PVT, or PDT as competing models for τ^*. We know that these models can be described by one intensity parameter $\gamma > 0$, which of course has a different meaning for each of the three models as explained in Section 2.3. First of all, we estimate the relevant characteristics of the input data using 3.2. By stepwise going

through a range of intensities for γ and by calculating each time the (theoretical) vector of characteristics (see Table 1), we finally determine an optimal vector λ_{min}, in the sense that the distance of the calculated vector of characteristics to the vector of the input characteristics is minimized. Finally, τ^* (and also the corresponding optimal intensity $\gamma^* > 0$) is obtained by minimizing between all tessellation models with respect to the distance $d(\widehat{\lambda}, \lambda_{min})$.

Now, consider the case of $k = 1$. This means that we consider a 1–fold $X_0/p\,X_1$ Poisson-type tessellation as described in Section 2.5. In particular, PLT, PVT, and PDT are considered as models both for X_0 and X_1. Figures 5 to 7 display examples of all possible choices. Each of these 1–fold tessellations can be described by two intensity parameters $\gamma_0 > 0$ and $\gamma_1 > 0$ as well as the probability p of the Bernoulli–thinning. Again, in the case $p = 1$, by stepwise going through a range of intensities for γ_0 and for γ_1 and by calculating the vector

$$\lambda = (\lambda_1, \ldots, \lambda_4)$$

in each step using theoretical formulae (see Table 2), an optimal vector λ_{min} of characteristics is determined as described above. When we have obtained a vector λ_{min} for each tessellation model (and therefore a corresponding optimal intensity pair (γ_0^*, γ_1^*)), the overall minimal value λ_{min}^* is obtained once again by minimizing between all tessellation models with respect to the distance $d(\widehat{\lambda}, \lambda_{min})$. Hence, the result is a vector λ_{min}^* of characteristics and its corresponding optimal tessellation model τ^* with intensity parameters γ_0^* and γ_1^*. One further dimension of minimization is introduced if we additionally consider the case $p < 1$, which eventually leads to an optimal Bernoulli–thinning parameter p^*.

3.5 Extensions of the decision procedure

To obtain the optimal tessellation model, several extensions of the decision procedure described in Section 3.4 are possible. In particular, one can choose $\varepsilon \geq 0$ in order to obtain the interval $[d^*, d^*(1 + \varepsilon)]$. Then we would consider every tessellation model to be a candidate for a possible description of the road system if we obtain an optimal distance value situated within the interval for this model. Assume we consider a model with obtained optimal distance, d_{min} say, where $d_{min} \in [d^*, d^*(1 + \varepsilon)]$. The error for such a model with respect to the optimal model with distance d^* is then given by $(\lambda_{min} - \lambda_{min}^*)/\lambda_{min}^*$.

3.6 Verification of the model choice procedure

The following so–called Monte–Carlo test technique is a general test principle based on simulations and is widely used in different fields of applications.

We start by establishing a null hypothesis H_0 which we want to test. This hypothesis states that the input data can be described by a certain tessellation model $\tau(H_0)$, where this model depends on an intensity parameter $\gamma(H_0)$ in the case of non–iterated tessellations and on intensity parameters $\gamma_0(H_0)$ and $\gamma_1(H_0)$ (and possibly a Bernoulli parameter $p(H_0)$) in the case of nested tessellations.

In order to validate the optimal choice of a tessellation model by the minimization procedure in Section 3.4, i.e. in order to validate τ^*, we choose $\tau(H_0) = \tau^*$ and $\gamma(H_0) = \gamma^*$ (in case of non–iterated tessellation models) or $\gamma_0(H_0) = \gamma_0^*$ and $\gamma_1(H_0) = \gamma_1^*$ (in case of nested tessellations).

The alternative hypothesis H_1 of such a test states that the input data can be described by the other models that are under consideration. For example if we consider non–iterated tessellations and if H_0 states that $\tau(H_0) = \tau^*$, where τ^* is a PLT say, then H_1 would state that the input data can be described by a PVT or a PDT (both with some fixed intensity parameter which can be obtained through the minimization procedure).

Notice that in case of real input data it is useful not only to test H_0 with $\tau(H_0) = \tau^*$, but for $\tau(H_0)$ to go through all tessellation models under consideration. Hence, if τ^* is a PLT for example, we also do tests with the second–best and third–best model, i.e. $\tau(H_0)$ is chosen to be a a PVT and a PDT for example with some fixed intensity parameters $\gamma(H_0)$.

In what follows, both for the evaluation of the method with simulated data and later on with real data, we will only state the null hypothesis H_0 for short.

Having stated H_0, a significance level α has to be chosen, which can be interpreted as the maximal error to reject H_0 despite its correctness. Popular choices are $\alpha = 0.05$ or $\alpha = 0.01$. Subsequently, the tessellation model $\tau(H_0)$ is simulated n times. Notice that for example Stoyan and Stoyan (1994) suggest to use $n = 99$ if $\alpha = 0.05$ or $n = 999$ if $\alpha = 0.01$.

Then, we choose a distance measure and for each simulation we compute the distance between the estimated vector of the realization of $\tau(H_0)$ and the vector of characteristics that is obtained for $\tau(H_0)$ via theoretical calculation using the intensity parameter $\gamma(H_0)$ (or $\gamma_0(H_0)$

and $\gamma_1(H_0)$). Eventually, we obtain n distance values $d_1, \ldots, d_n$. One further value $d_{n+1} = d^*$ is obtained as distance between the vector of characteristics for $\tau(H_0)$ and the (estimated) vector of characteristics of input data.

Notice that the distance measure can be chosen independently from the distance measure used for the minimization procedure. Again, it can be expected that relative distance measures perform better than absolute ones. However, it has to be pointed out that the order of the arguments of relative distance measures has to be kept in mind, i.e., complying to the definition in Section 3.3, the vector of characteristics calculated for $\gamma(H_0)$ (or $\gamma_0(H_0)$ and $\gamma_1(H_0)$) would be the first argument.

Subsequently, the $n + 1$ values $d_1, \ldots, d_n$ and d^* are ordered in ascending order, which leads to a sequence $d_{(1)}, \ldots, d_{(n+1)}$, where $d_{(i)}$ denotes the ith smallest distance for $1 \leq i \leq n + 1$. The null hypothesis H_0 is rejected if the position i^* of d^* in this ordered sequence is contained in the rejection region $R_\alpha = [n - \alpha(n + 1) + 2, \ldots, n + 1]$. Alternatively, we may consider the p–value, which can be expressed as $1 - (i^* - 1)/(n + 1)$. Notice that this value is always in the range of $[1, 1/(n + 1)]$. As always, H_0 should be rejected if the obtained p–values are smaller than the chosen significance level α.

In case of simulated input data the power P_{MC} of Monte–Carlo tests can be estimated and together with this the probability of the error to accept H_0 despite it is not true.

The procedure is in principle analogous to the proceeding described above, except that we replace $\tau = \tau(H_0)$ by one of the tessellations stated in the alternative hypothesis H_1, where an intensity parameter of τ can be obtained via the minimization procedure. This can be interpreted in the sense that we examine the performance under H_1.

Finally, we repeat the procedure k times, $k \geq 1$, and for each $\ell = 1, \ldots, k$ we are able to report whether the position i_ℓ^* of the distance d_ℓ^* within the ordered sequence $d_{\ell,(1)}, \ldots, d_{\ell,(n+1)}$ of distances is in R_α or not. The distance d_ℓ^* is calculated between the estimated vector of characteristics of the input data and the vector of characteristics of the model, which can be calculated using the theoretical intensity value stated in H_1.

Hence, an estimate of the power P_{MC} of the Monte–Carlo test is obtained by regarding the estimator

$$\widehat{P}_{MC} = \frac{1}{k} \#\{i_\ell^* \in R_\alpha, \ell = 1, \ldots, k\}. \tag{3.4}$$

The power of the Monte–Carlo test can be considered to be high if $\widehat{P}_{MC}$ takes values which are close to one.

3.7 Numerical examples using simulated data

In the following, we present numerical results, where input data are derived by simulations of the competing tessellation models. We concentrate on relative distance measures since simulation studies showed that such distance measures do indeed perform better than absolute distance measures.

Assume that the input data are realizations of a non–iterated PLT with parameter $\gamma = 0.1$ and assume further that we want to verify if our procedure can correctly decide between a PLT, a PVT, and a PDT. Particularly, the vector $\lambda = (\lambda_1, \ldots, \lambda_4)$ of characteristics as introduced in (3.1) is considered. Using the vector $\widehat{\lambda}$ in (3.2) as given by (3.3) as estimator for the vector of these characteristics, Table 3 shows estimates based on one and on 1000 realizations, respectively, of the input data in a quadratic sampling window of side length 300 (and area 9×10^4). Notice however that in case of the 1000 realizations for example, a similar quality of estimation can be obtained by only one single realization of the tessellation, but then in a quadratic sampling window of area 9×10^7.

Table 3: *Estimation of characteristics based on n realizations for PLT–input*

	$n = 1$	$n = 1000$	Theoretical
$\widehat{\lambda}_1$	0.00333	0.00318	0.00318
$\widehat{\lambda}_2$	0.00611	0.00630	0.00637
$\widehat{\lambda}_3$	0.00333	0.00318	0.00318
$\widehat{\lambda}_4$	0.10291	0.09995	0.10000

For the minimization procedure we try to find optimal values for γ within the range $[0.0001, 0.5]$, where the step width is chosen to be 0.00001. Table 4 displays the numerical values for all three relative distance measures d' and the corresponding optimized intensity parameter γ.

For example in the case of the relative Euclidean distance measure d'_e, the numerical values displayed in Table 4 suggest a decision in favor of a PLT as optimal tessellation model τ^*

Table 4: *Optimal values of d'_e, d'_a, and d'_m and corresponding optimized parameter γ*

	$d'_{e,min}$	γ	$d'_{a,min}$	γ	$d'_{m,min}$	γ
PLT	0.07154	0.10070	0.10112	0.10180	0.04382	0.10010
PVT	0.47164	0.00200	0.76115	0.00200	0.34000	0.00220
PDT	0.62160	0.00180	1.03066	0.00170	0.43816	0.00190

with intensity parameter $\gamma^* = 0.10070$. However, it can be seen that the optimal γ–value is relatively stable and does not depend strongly on the chosen distance measure.

Table 5: *Monte-Carlo test for PLT–input, where $\tau(H_0)$ is a PLT with $\gamma = 0.10070$*

α	n	R_α	d^*	$d_{(1)}$	$d_{(n+1)}$	i^*	p–value	reject
0.05	99	$[96, 100]$	0.00223	0.00011	0.04312	8	0.93	no
0.01	999	$[991, 1000]$	0.00223	0.00015	0.07141	108	0.893	no

Table 5 displays the results of a Monte–Carlo test, where the null hypothesis H_0 states that $\tau(H_0) = \tau^*$, i.e. H_0 states that the input data can be represented by a PLT with intensity $\gamma(H_0) = \gamma^* = 0.10070$. The distances are calculated using the relative Euclidean distance. To get an impression of the range of the ascending ordered sequence of distances $d_{(1)}, \ldots, d_{(n+1)}$, the values $d_{(1)}$ and $d_{(n+1)}$ are displayed in Table 5. Furthermore, the position i^* of d^* within this ordered sequence of distances is given. The decision to not reject the null hypothesis can be obtained via two approaches. First, we see that $i^* \notin R_\alpha$. Second, the p–values are very large overall and also compared to the significance levels $\alpha = 5\%$ and $\alpha = 1\%$. Therefore, H_0 is not rejected and hence we may say that the input data can be represented by a PLT with intensity parameter 0.1007.

Finally, we examine the power P_{MC} of this Monte–Carlo test. Hence, we proceed as described in Section 3.6 and replace in the simulations the PLT model by the models stated in the alternative hypothesis H_1, namely by a PVT (with some optimal intensity parameter) and by a PDT (with some optimal intensity parameter). We choose $k = 1000$, i.e. the whole procedure of estimating the power using the estimator $\widehat{P}_{MC}$ in (3.4) is repeated 1000 times.

In case of a PVT and a significance level $\alpha = 0.05$, Table 6 shows the results of one of the 1000 repetitions, and as estimated power we obtain the value $\widehat{P}_{MC} = 1$. The same result is

Table 6: *Power examination for PLT–input with simulated PVT*

α	n	R_α	d^*	$d_{(1)}$	$d_{(n+1)}$	i^*
0.05	99	$[96, 100]$	0.01355	0.00004	0.01355	100
0.01	999	$[991, 1000]$	0.01355	0.00005	0.01355	1000

obtained for the case of a PDT with intensity $\gamma = 0.00180$ (again for $\alpha = 0.05$).

Notice that this rather high estimated value for power can be explained by the fact that PLT on one side and both PVT and PDT on the other side are quite different models with regard to their geometrical structure.

As a second example the input data are now derived from a $X_0/p\,X_1$–tessellation, where X_0 and X_1 are chosen to be a PLT with parameter $\gamma_1 = 0.08$ and a PDT with parameter $\gamma_2 = 0.0008$, respectively. In the case of $p = 1$ we obtain 9 competing models. Additionally we can consider the case where p is any arbitrary number with $0 \leq p < 1$. Notice however, that such a minimization increases the computational complexity remarkably. Therefore, we first concentrate on the case $p = 1$, which is interesting in its own right due to a certain symmetry inherent in the intensity formulae shown in Table 2.

Again the same vector $\lambda_1, \ldots, \lambda_4$ of characteristics given in (3.1) is considered. Table 7, which can be understood completely analogously to Table 3, shows the performance of our estimators, based on one and on 1000 sample realizations of the PLT/PDT–tessellation, respectively.

For the minimization procedure we try to find optimal values for γ_0 and γ_1 within the range $[0.00001, 0.15]$, where the step width is chosen to be 0.00001. As distance measures the relative Euclidean metric d'_e and the relative absolute–value metric d'_a are considered. For the alternative measure d'_m similar results are obtained. Table 8 displays the minimal distance values $d'_{e,min}$ and $d'_{a,min}$ of d'_e and d'_a, respectively, for each tessellation model and the corresponding optimized intensity parameters γ_0 and γ_1. Minimizing the obtained distance values over all nine possible tessellation models, the decision is in favor of a PLT/PDT or a PDT/PLT and hence, due the symmetry explained above, is not unique. However, in order

Table 7: *Estimation of characteristics based on n realizations of PLT/PDT–input*

	$n = 1$	$n = 1000$	Theoretical
$\widehat{\lambda}_1$	0.01172	0.01231	0.012619
$\widehat{\lambda}_2$	0.01951	0.02165	0.021147
$\widehat{\lambda}_3$	0.00779	0.00829	0.008528
$\widehat{\lambda}_4$	0.17271	0.17433	0.176034

to get a unique decision, a possible solution is to consider additionally characteristics of the initial tessellation. Therefore, we also estimate the vector of characteristics λ separately for the initial tessellation X_0, which in this example is a PLT. These estimates and their theoretical values are given in Table 9.

The model fitting procedure for X_0 is carried through completely analogous to the one for non–iterated random tessellations. The obtained minimal distance values of the distance measures d'_e and d'_a are shown in Table 10 along with the corresponding intensity values. According to these distance values, the decision for an optimal initial tessellation is in favor of a PLT. Hence, we would decide in favor of a PLT/PDT as optimal model with intensities $\gamma_0^* = 0.07200$ and $\gamma_1^* = 0.00085$ as obtained before by using the metric d'_e, or $\gamma_0^* = 0.07100$ and $\gamma_1^* = 0.00088$ for the metric d'_a. As can be seen, using d'_e or d'_a makes no big difference in the resulting numerical values for the intensity parameters.

Alternatively, a slight modification of the original model fitting procedure can be applied to get also a unique decision. Doing so, the optimal intensity from the model fitting for the initial tessellation is kept as intensity γ_0^* for the iterated tessellation. Thereafter, the model choice algorithm is applied for 1–fold nestings, where the model decision has now to be made only between PLT/PLT, PLT/PVT, and PLT/PDT knowing already that the PLT is the initial tessellation. Table 11 displays the results, which are obtained in this way. Here, the numerical values of $d'_{e,min}$ suggest a decision in favor of a PLT/PDT with parameters $\gamma_0^* = 0.08557$ and $\gamma_1^* = 0.00080$. The values obtained for different metrics seem not to differ too much since the intensity values for $d'_{a,min}$ are the same.

In analogy to the examples for non–iterated tessellations as input data, the correctness of the

Table 8: *PLT/PDT–input: Minimal distance values $d'_{e,min}$ and $d'_{a,min}$ with the corresponding optimized parameters γ_0 and γ_1 for all 9 types of X_0/X_1–tessellations*

	$d'_{e,min}$	γ_0	γ_1	$d'_{a,min}$	γ_0	γ_1
PLT/PLT	0.08676	0.10500	0.05400	0.10132	0.08700	0.07000
PLT/PVT	0.08469	0.10900	0.00063	0.10204	0.10800	0.00063
PLT/PDT	0.02055	0.07200	0.00085	0.03465	0.07100	0.00088
PVT/PLT	0.08469	0.00063	0.10900	0.10204	0.00063	0.10800
PVT/PVT	0.21228	0.00003	0.00555	0.35795	0.00003	0.00555
PVT/PDT	0.02643	0.00090	0.00117	0.04174	0.00084	0.00125
PDT/PLT	0.02055	0.00085	0.07200	0.03465	0.00088	0.07100
PDT/PVT	0.02643	0.00117	0.00090	0.04174	0.00125	0.00084
PDT/PDT	0.03995	0.00069	0.00070	0.06165	0.00069	0.00070

decision is verified using the Monte–Carlo test technique, where H_0 states that $\tau(H_0)$ is a PLT/PDT with $\gamma_0(H_0) = 0.08557$ and $\gamma_1(H_0) = 0.00080$. Furthermore, a power examination can be conducted for the tessellation models (PLT/PLT and PLT/PVT, respectively) contained in the alternative hypothesis leading to estimated powers but, as one surely expects, with lower values as in the non–iterated case.

We close this section by a short discussion of X_0/pX_1–tessellations where $0 \leq p < 1$ and we consider the case where both X_0 and X_1 are PLTs with intensities $\gamma_0 = 0.08$ and $\gamma_1 = 0.05$, respectively. Here, the symmetry in the intensity formulae that caused more detailed examinations in the case $p = 1$ does not occur and we get a decision in favor of a PLT/PLT with fixed Bernoulli–thinning parameter $p = 0.5$ and optimal intensity parameters $\gamma_0^* = 0.07796, \gamma_1^* = 0.05549$; see Table 12.

3.8   Systematic examination

In the preceding section we explained the fitting procedure for one chosen value of γ in the case of 0–fold nestings and for two chosen values γ_0 and γ_1 in the case of 1–fold nestings.

Table 9: *Estimation of characteristics of n realization of the initial tessellation of the PLT/PDT–tessellation*

	$n = 1$	$n = 1000$	Theoretical
$\widehat{\lambda}_1$	0.00232	0.00206	0.00204
$\widehat{\lambda}_2$	0.00464	0.00411	0.00407
$\widehat{\lambda}_3$	0.00232	0.00206	0.00204
$\widehat{\lambda}_4$	0.08754	0.08033	0.08000

Table 10: *Initial tessellation of PLT/PDT: Distance values $d'_{e,min}$ and $d'_{a,min}$ with corresponding optimized parameter γ for PLT, PVT, and PDT*

	$d'_{e,min}$	γ	$d'_{a,min}$	γ
PLT	0.02338	0.08557	0.02458	0.08541
PVT	0.47263	0.00147	0.76779	0.00155
PDT	0.62622	0.00128	1.07166	0.00116

The intention of this section is to show that our procedure of course works correctly not only for these choices but also for other numerical values of intensity parameters. We constraint ourselves to present a systematic examination in the case of simple tessellations.

We consider input data derived from a PLT with intensity parameter $\gamma \in [0.01, 0.3]$. The step width is 0.001 for $0.01 \leq \gamma \leq 0.05$ and step width 0.01 for $0.05 \leq \gamma \leq 0.3$. Notice that the sampling window for the simulations is again a rectangle with side length 300.

Figure 9 shows the results of the examination of our fitting procedure where the absolute Euclidean distance d_e was used. We observe that overall the PLT is indeed recognized as best model, however the distinction between the models gets better and better with increasing intensity γ. In contrast to that we see the same result with the relative Euclidean distance d'_e in Figure 10. There, the distances between the optimal model (PLT) and its two alternatives is quite larger than in case of the absolute distance. Moreover, the quality of distinction is the same for each value of γ.

Table 11: *PLT/PDT-input: Distance values $d'_{e,min}$ and $d'_{a,min}$ with corresponding optimized parameters γ_0 and γ_1 for PLT/PLT, PLT/PVT, and PLT/PDT with fixed X_0*

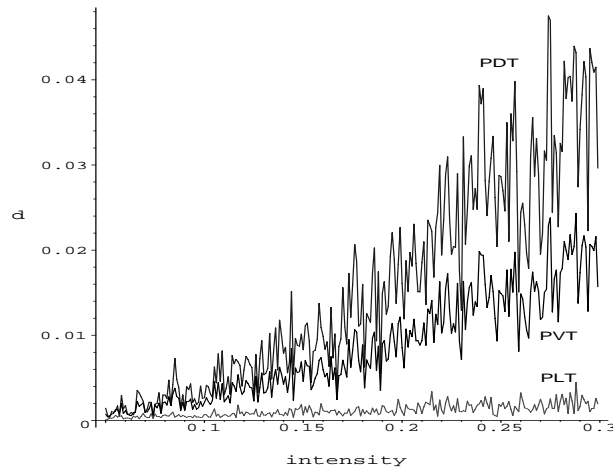|	$d'_{e,min}$	γ_0	γ_1	$d'_{a,min}$	γ_0	γ_1
PLT/PLT	0.13808	0.08557	0.09000	0.26178	0.08557	0.09000
PLT/PVT	0.21450	0.08557	0.00203	0.34765	0.08557	0.00203
PLT/PDT	0.02028	0.08557	0.00080	0.03325	0.08557	0.00080

Figure 9: *Fitting procedure for PLT input with intensity γ using the absolute Euclidean distance d_e*

Notice that, referring also to the remark on the similiar quality of estimation in differently sized sampling windows at the beginning of Section 3.7, it is of course possible to do such systematic examinations within sampling windows of side lengths other than 300. For example one could take a sampling window where the side length is ten times as large, i.e. a sampling window of side length 3000 as is the case with real data considered in the upcoming Section 4. In order to compare results with respect to the quality of estimation and minimization to the results obtained within the sampling window of side length 300 as considered in this section, the intensity parameter would also have to be adapted. Therefore, the situation of input data in a larger window with rather small intensity parameter for example can be compared to the situation of input data in smaller sampling windows but with larger intensity.

Table 12: *PLT/PLT–input with Bernoulli–thinning parameter $p = 0.5$ and intensity parameters $\gamma_0 = 0.08, \gamma_1 = 0.05$: Distance values $d'_{e,min}$ with corresponding optimized parameters γ_0 and γ_1*

	$d'_{e,min}$	γ_0	γ_1
PLT/PLT	0.07625	0.07796	0.05549
PLT/PDT	0.08132	0.05507	0.00097
PLT/PVT	0.08110	0.08622	0.00046
PDT/PLT	0.07821	0.00028	0.09615
PDT/PDT	0.11929	0.00023	0.00134
PDT/PVT	0.15293	0.00052	0.00131
PVT/PLT	0.15603	0.00052	0.10000
PVT/PDT	0.16862	0.00182	0.00010
PVT/PVT	0.28715	0.00208	0.00019

4 Analysis of real infrastructure data

4.1 Preprocessing of raw data

Most often, infrastructure data of certain urban (or rural) areas are given as raw data and need to be preprocessed. Here, we briefly describe the measures that were necessary to get data which could be used for our model fitting. For further information on general image retrieval techniques see, for example, Serra (1982) and Soille (2003).

Figure 3 shows the infrastructural geometry of Paris, where coordinates are given as geodesic data. In this case, the Lambert2–projection methodology was used to obtain these data in the form of (locally) Cartesian–like treatable (x, y)–coordinates. Another quite widely used methodology of this form is called Gauss–Krüger methodology.

Regarding Figure 11 (a), it is quite obvious that certain preprocessing steps are necessary in order to use the raw data for fitting models. As can be seen, roads are given as series of line segments, where each line segment consists of a start point and an end point. Furthermore, a mark is attached to each line segment, which describes the type of the road. These road
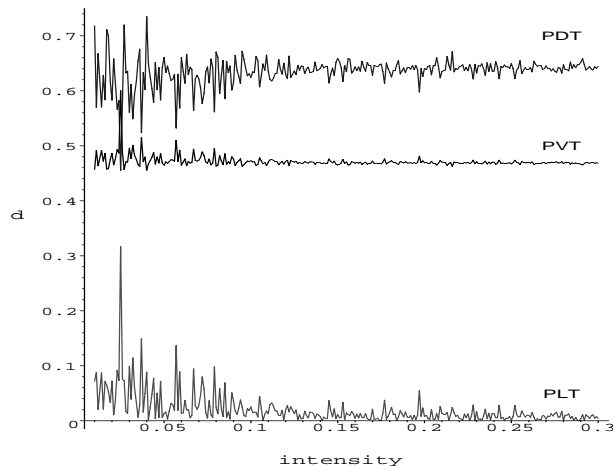
Figure 10: *Fitting procedure for PLT input with intensity γ using the relative Euclidean distance d'_e*

types rank from highways and national routes down to small side streets. In our analysis, we concentrate on only two types, which we identified to be the most common road types in Paris, namely intercity main roads and side streets. Finally, we remove dead end streets and by traversing through the line segments it is possible to reconstruct a tessellation which consists of polygonal cells; Figure 11 (b). Still however, points may exist, which do not belong to the set of vertices of the obtained tessellation. Clearly, these are points where only two line segments emanate from and therefore it is easy to not account for such points.

4.2   Numerical results

We examine preprocessed data of Paris within a rectangular sampling window W; see Figure 12. Since in our case we know how to differentiate between main roads and side streets, our philosophy is to use 1–fold tessellation models for data fitting. Hence we expect to get a realistic and adequate model for our data, since we additionally use the information about their hierarchical structure.

We restrict our description to the case of $X_0/p\,X_1$–tessellations where $p = 1$. As pointed out earlier, this is a rather complicated case since the decision is not unique from the first and we have to use some more information to get a unique decision. Notice that in the case of $p < 1$ the decision is unique. Minimization of an additional parameter p however is little more challenging regarding the computational run times.

We start by considering the main roads contained in Figure 12. Table 13 shows results of
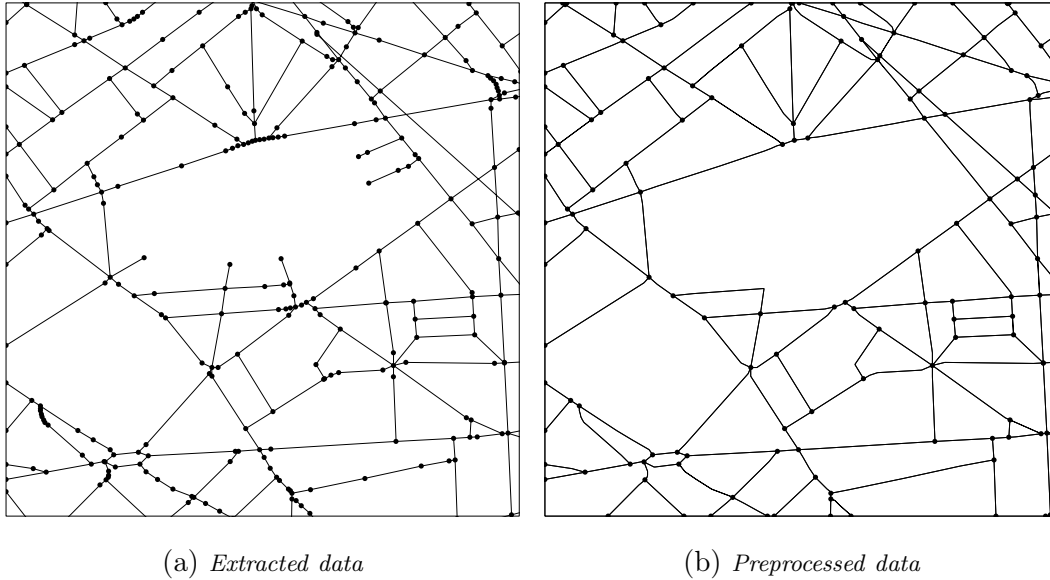
(a) *Extracted data* (b) *Preprocessed data*

Figure 11: *Infrastructure of Paris. From raw data to preprocessed data*

the minimization procedure applied to the main road data if we use the relative Euclidean metric d'_e as a measure of distance. Here the decision would be in favor of a PLT, i.e. τ^* is a PLT with optimal intensity parameter $\gamma^* = 0.002384$. However, regarding the distance values and using the alternative decision rule of Section 3.5 with $\varepsilon = 0.5$ for example, we cannot rule out a PVT with intensity parameter $\gamma = 0.000001$.

Table 13: *Main road input data of Figure 12: Distance values $d'_{e,min}$ and corresponding optimized parameter γ for PLT, PVT, and PDT*

X	$d'_{e,min}$	γ^*
PLT	0.21101	0.002384
PVT	0.29749	0.000001
PDT	0.73378	0.000001

As far as Monte–Carlo tests are concerned for real data input, the idea is to do these tests for all considered tessellation models and to check whether a certain model can be taken as representation of the real data. In the case of X_0 for example, we consider null hypotheses H_0 for non–iterated tessellations, i.e. $\tau(H_0)$ is chosen to be, one at a time, PLT, PVT, and PDT with intensity parameter $\gamma(H_0)$. Hence, Table 14, Table 15, and Table 16 show the results

(a) *Raw data*   (b) *Preprocessed data*

Figure 12: *Infrastructure data of Paris in a sampling window W (location: lower left vertex* $[5000, 3000]$, *upper right vertex* $[8000, 6000]$*)*

Table 14: *Monte-Carlo test for main road data of Figure 12,* $\tau(H_0)$ *is a PLT with* $\gamma(H_0) = 0.002384$

α	n	R_α	d^*	$d_{(1)}$	$d_{(n+1)}$	i^*	p–value	rejected
0.05	99	$[96, 100]$	0.24648	0.03774	5.39621	14	0.86	no
0.01	999	$[991, 1000]$	0.24648	0.03050	12.03780	126	0.875	no

of these tests, where the tables can be read in complete analogy to Table 5. Notice that in Table 14 the null hypothesis H_0 states that $\tau(H_0) = \tau^*$, i.e. H_0 states that the main road data can be described by a PLT with intensity $\gamma(H_0) = \gamma^* = 0.002384$, Here we test for the optimal tessellation model τ^* that has been identified through our minimization procedure. As we can see, for both significance levels we cannot reject H_0.

Table 15 and Table 16 present the results of the test where H_0 states that the main road data can be represented by a PVT with intensity parameter $\gamma(H_0) = 0.000001$ and a PDT with intensity parameter $\gamma(H_0) = 0.000001$, respectively. In view of Table 13 this means that we test for the second–best and third–best model obtained through the minimization procedure. As we can see, we cannot reject these null hypotheses, except for the case where $\tau(H_0)$ is a PDT for significance levels α with $\alpha \geq 0.02$.

Table 15: *Monte-Carlo test for real data of Figure* 12, $\tau(H_0)$ *is a PVT with* $\gamma(H_0) = 0.000001$

α	n	R_α	d^*	$d_{(1)}$	$d_{(n+1)}$	i^*	p–value	rejected
0.05	99	$[96, 100]$	0.37171	0.02496	1.01796	64	0.37	no
0.01	999	$[991, 1000]$	0.37171	0.02488	1.68198	583	0.418	no

Table 16: *Monte-Carlo test for real data of Figure* 12, $\tau(H_0)$ *is a PDT with* $\gamma(H_0) = 0.000001$

α	n	R_α	d^*	$d_{(1)}$	$d_{(n+1)}$	i^*	p–value	rejected
0.05	99	$[96, 100]$	1.44039	0.04723	1.66671	99	0.02	yes
0.01	999	$[991, 1000]$	1.44039	0.04710	2.03734	982	0.019	no

To get an optimal 1–fold nesting model we consider the following. We fix the optimal initial tessellation as given by Table 13. Since the distance is minimal for a PLT with parameter $\gamma_0 = 0.002384$ we concentrate on this model in the following to describe X_0. Hence, we finally only need to distinguish between PLT/PLT, PLT/PVT, and PLT/PDT.

Table 17 shows the results if we apply our minimization procedure with fixed initial tessellation type PLT and intensity $\gamma_0 = 0.002384$ to these three models. The results indicate that a PLT/PLT model would be optimal, where for the nested PLT we have the intensity parameter $\gamma_1 = 0.013906$.

Table 17: *Input data of Figure* 12: *Distance values* $d'_{e,min}$ *with fixed tessellation* X_0 *as PLT with* $\gamma_0 = 0.002384$ *and corresponding* γ_1 *for PLT/PLT, PLT/PVT, and PLT/PDT*

X	$d'_{e,min}$	γ_0	γ_1
PLT/PLT	0.15224	0.002384	0.013906
PLT/PVT	0.20455	0.002384	0.000044
PLT/PDT	0.36649	0.002384	0.000028

Finally, we consider again Monte–Carlo tests, where the idea is to test all tessellation models

given in Table 17, i.e. for H_0 we choose $\tau(H_0)$ to be a PLT/PLT, PLT/PVT, or PLT/PDT with intensity parameters $\gamma_0(H_0) = \gamma_0$ and $\gamma_1(H_0) = \gamma_1$, respectively, as given in Table 17.

We restrict ourselves to two examples. First we consider the case where H_0 states that $\tau(H_0)$ is a PLT/PDT with intensity parameters $\gamma_0(H_0) = 0.002384$ and $\gamma_1(H_0) = 0.000028$. Table 18 and Table 19 show the results both for the case of the relative Euclidean distance (Table 18) and for the case of the absolute Euclidean distance (Table 19), respectively. Regarding Table 18, we see that with the relative Euclidean distance we would not reject the

Table 18: *Monte-Carlo test for real data of Figure* 12, $\tau(H_0)$ *is a PLT/PDT with* $\gamma_0(H_0) = 0.002384$ *and* $\gamma_1(H_0) = 0.000028$ *(relative Euclidean distance)*

α	n	R_α	d^*	$d_{(1)}$	$d_{(n+1)}$	i^*	p–value	rejected
0.05	99	[96, 100]	0.43747	0.01339	1.28781	83	0.18	no
0.01	999	[991, 1000]	0.43747	0.00460	1.16786	816	0.185	no

null hypothesis, however regarding the rank i^* of d^* in the ordered sequence of distances (and hence regarding the p–value), this decision is relatively tight. Using however the absolute Euclidean distance for the same test, we conclude that the null hypothesis has to be rejected. Similar tests can also be done for the case where H_0 states that $\gamma(H_0)$ is a PLT/PVT type

Table 19: *Monte-Carlo test for real data of Figure* 12, $\tau(H_0)$ *is a PLT/PDT with* $\gamma_0(H_0) = 0.002384$ *and* $\gamma_1(H_0) = 0.000028$ *(absolute Euclidean distance)*

α	n	R_α	d^*	$d_{(1)}$	$d_{(n+1)}$	i^*	p–value	rejected
0.05	99	[96, 100]	0.00313	0.00003	0.00313	100	0.00	yes
0.01	999	[991, 1000]	0.00313	1.15251	0.00355	997	0.004	yes

tessellation.

Finally, Table 20 shows the results of the Monte–Carlo test for the null hypothesis H_0 with $\tau(H_0) = \tau^*$ and τ^* being a PLT/PLT as obtained according to the minimization procedure with optimal intensity parameters $\gamma_0^* = 0.002384$ and $\gamma_1^* = 0.013906$. In this case, using the relative Euclidean distance, we cannot reject this null hypothesis. We also considered this

null hypothesis with a Monte–Carlo test where the absolute Euclidean distance measure was used, obtaining similar results.

Table 20: *Monte-Carlo test for real data of Figure* 12, $\tau(H_0)$ *is a PLT/PLT with* $\gamma_0(H_0) = 0.002384$ *and* $\gamma_1(H_0) = 0.013906$ *(relative Euclidean distance)*

α	n	R_α	d^*	$d_{(1)}$	$d_{(n+1)}$	i^*	p–value	rejected
0.05	99	$[96, 100]$	0.15327	0.00968	1.22830	30	0.71	no
0.01	999	$[991, 1000]$	0.15327	0.00966	1.04893	306	0.695	no

5 Discussion and Outlook

One of the key necessities of any models like the SSLM for cost analysis and strategic planning of telecommunication networks is to accurately represent the underlying geometrical structure of the network. Therefore, the modelling task can be split into two steps, which are closely connected to each other.

A first step is to incorporate the spatial–geometric structure of the infrastructure along which in most cases, but especially in urban areas, the cable trench system is located. In the SSLM the road system is modelled using the concept of random tessellations. In this paper we propose a procedure which decides in favor of an optimal road system model within a class of given random tessellation models. The procedure has been tested with simulated characteristics as input data, which have been estimated from realizations of the tessellation models under consideration. Particulary, the comparison of input characteristics and theoretical tessellation models described by a certain intensity parameter is possible since we used some characteristics which are related to the intensity parameters through theoretically known formulae. The results of our method are quite impressive in the sense that relatively simple mathematical methods have been combined. In particular, it allows for a general classification of the different models regarding their intensities. Symmetries, for example in the case of PLT/PVT–nestings and PVT/PLT–nestings without Bernoulli–thinning, can be overcome by a slightly modified version of the model choice algorithm, depending on the separate knowledge of initial and nested tessellation data. If this information is not available,

i.e., if we cannot distinguish between initial and nested tessellation in a 1–fold nesting say, then it might be a good idea to choose some small $\varepsilon > 0$ and fit an $X_0/p\,X_1$–nesting with $p = 1 - \varepsilon$. Hence, the decision is unique and letting $\varepsilon \to 0$, i.e. executing a sequence of fitting steps with ε getting smaller and smaller, the hope is that also the decision for the limiting tessellation is in favor of that same $X_0/p\,X_1$–nesting.

Finally, the model choice procedure has been confronted with a set of (preprocessed) infrastructure data of Paris. Owing to the structure of the data, which clearly do not follow any of the proposed tessellation models, the fit is worse, but still relatively impressive regarding our numerical results. Naturally, we can only hope to identify one model among the theoretically proposed which comes closest to the given data.

Clearly, it is necessary to refine the fitting procedure. One possibility can be the application of central limit theorems, like in Heinrich, Schmidt and Schmidt (2005). There, asymptotical studies of the distribution of certain functionals of both Poisson line tessellations as well as Poisson–Voronoi tessellations are shown, where the asymptotic comes in through an unboundedly growing sampling window. Such results lead to central limit theorems and hence to (asymptotic) confidence intervals and tests.

In a second step, the chosen geometric model representation of the infrastructure has to be used for evaluation of the network. Therefore, the network equipment is placed onto the chosen tessellation model for the road system. Realizations of certain types of point processes are used to represent these nodes. In particular, one is interested in the tree connecting subscribers of a certain serving area to the corresponding WCS–station via intermediate stations of lower level along the road system. Routing techniques can be applied to analyze shortest paths between subscribers and equipment of any hierarchy level in the network. For example, the expected mean of shortest path lengths between a WCS–station and a SAI–station can be examined for random tessellation models. This will lead to simulated results or even theoretical formulae for the whole tree connecting subscribers of a certain serving zone to the corresponding WCS–station. Further information can be found in Gloaguen et al. (2005a, 2005b), where we present simulation techniques and results using simulation of typical cells, corresponding typical trees, and reduction of parameters through parametric scaling.

Acknowledgement

References

[1] F. Baccelli and B. Blaszczyszyn. (2001). "On a coverage process ranging from the Boolean model to the Poisson-Voronoi tessellation." *Advances in Applied Probability* **33**, 293–323.

[2] F. Baccelli, C. Gloaguen, and S. Zuyev. (2000). "Superposition of Planar Voronoi Tessellations." *Communications in Statistics, Series Stochastic Models* **16**, 69–98.

[3] F. Baccelli, M. Klein, M. Lebourges, and S. Zuyev. (1996). "Géomètrie aléatoire et architecture de réseaux." *Annales des Télécommunication* **51**, 158–179.

[4] F. Baccelli, D. Kofman, and J.L. Rougier. (1999). "Self organizing hierarchical multicast trees and their optimization." *Proceedings of IEEE Infocom '99*, 1081–1089, New York.

[5] F. Baccelli and S. Zuyev. (1996). "Poisson-Voronoi spanning trees with applications to the optimization of communication networks." *Operations Research* **47**, 619–631.

[6] A.J. Baddeley and E.B. Vedel Jensen. (2004). *Stereology for Statisticians.* Chapman & Hall.

[7] C. Gloaguen, P. Coupé, R. Maier and V. Schmidt. (2002). "Stochastic modelling of urban access networks." *Proc. 10th Internat. Telecommun. Network Strategy Planning Symp.*, (Munich, June 2002), VDE, Berlin, pp. 99-104.

[8] C. Gloaguen, F. Fleischer, H. Schmidt and V. Schmidt. (2005a). "Simulation of typical Cox-Voronoi cells, with a special regard to implementation tests." *Mathematical Methods of Operations Research* **62**, to appear.

[9] C. Gloaguen, F. Fleischer, H. Schmidt and V. Schmidt. (2005b). "Analysis of shortest paths and subscriber line lengths in telecommunication access networks." *Working paper*, under preparation.

[10] L. Heinrich, H. Schmidt and V. Schmidt (2005). "Central Limit Theorems for Poisson Hyperplane Tessellations". *Preprint*, submitted.

[11] R. Maier. (2003). *Iterated Random Tessellations with Applications in Spatial Modelling of Telecommunication Networks.* Doctoral Dissertation, University of Ulm.

[12] R. Maier, J. Mayer and V. Schmidt. (2004). "Distributional properties of the typical cell of stationary iterated tessellations." *Mathematical Methods of Operations Research* **59**, 287–302.

[13] R. Maier and V. Schmidt. (2003). "Stationary iterated tessellations." *Advances in Applied Probability* **35**, 337–353.

[14] J. Mayer, V. Schmidt and F. Schweiggert. (2004). "A unified simulation framework for spatial stochastic models." *Simulation Modelling Practice and Theory* **12**, 307–326.

[15] J. Møller. (1989). "Random tessellations in $\mathbb{R}^d$." *Advances in Applied Probability* **21**, 37–73.

[16] J. Ohser and F. Mücklich. (2000). *Statistical Analysis of Microstructures in Materials Science.* J.Wiley & Sons, Chichester.

[17] A. Okabe, B. Boots, K. Sugihara and S.N. Chiu. (2000). *Spatial Tessellations.* 2nd ed., J.Wiley & Sons, Chichester.

[18] R. Schneider and W. Weil. (2000). *Stochastische Geometrie.* Teubner, Stuttgart.

[19] J. Serra. (1982). *Image Analysis and Mathematical Morphology.* Academic Press, London.

[20] P. Soille. (2003). *Morphological Image Analysis.* Springer, Berlin.

[21] D. Stoyan, W.S. Kendall and J. Mecke. (1995). *Stochastic Geometry and its Applications.* 2nd ed., J. Wiley & Sons, Chichester.

[22] D. Stoyan and H. Stoyan. (1994). *Fractals, Random Shapes and Point Fields. Methods of Geometrical Statistics.* J.Wiley & Sons, Chichester.

[23] K. Tchoumatchenko and S. Zuyev. (2001). "Aggregate and fractal tessellations." *Probability Theory Related Fields* **121**, 198–218.

FOOTNOTES

AFFILIATION OF AUTHORS

Dr. Catherine GLOAGUEN
France Telecom R&D Division RESA/NET/NSO, 92794 Issy Moulineaux Cedex 9, France

Dipl.-Math. oec. Frank FLEISCHER M.Sc.
Department of Applied Information Processing and Department of Stochastics, University of Ulm, 89069 Ulm, Germany

Dipl.-Math. oec. Hendrik SCHMIDT M.Sc.
Department of Stochastics, University of Ulm, 89069 Ulm, Germany

Professor Volker SCHMIDT
Department of Stochastics, University of Ulm, 89069 Ulm, Germany

Contact author

Catherine GLOAGUEN

France Télécom R&D RESA/NET/NSO

38-40 Rue du Général Leclerc

92794 Issy Moulineaux Cedex 9, France

E-mail : catherine.gloaguen@francetelecom.com

Tel : + 33 1 45 29 64 41

Fax : + 33 1 45 29 63 07

Manuscript Dates

Submission : December, 20th, 2004

First revision : June, 24th, 2005

Second revision : October, 7th, 2005