Densities of shortest path lengths in spatial stochastic networks

Florian Voss, Catherine Gloaguen, Frank Fleischer, Volker Schmidt

October 7, 2010

Abstract We consider a spatial stochastic model for telecommunication networks, the stochastic subscriber line model, and we investigate the distribution of the typical shortest path length between network components. Therefore, we derive a representation formula for the probability density of this distribution which is based on functionals of the so-called typical serving zone. Using this formula, we construct an estimator for the density of the typical shortest path length and we analyze the statistical properties of this estimator. Moreover, we introduce new simulation algorithms for the typical serving zone which are used in a numerical study in order to estimate the density and moments of the typical shortest path length for different specific models.

Keywords Point Processes · Stochastic Geometry · Random Tessellations · Telecommunication Systems · Palm Calculus

1 Introduction

In the present paper, we consider models for hierarchical telecommunication networks. These networks consist of higher-level components (HLC) and lower-level components (LLC) located along the underlying road system of the network. Then to each HLC a domain is associated which we call serving zone. Now a LLC in this serving zone is connected to the corresponding HLC on the shortest path along the road system. Such telecommunication networks were recently investigated based on the stochastic subscriber line model (SSLM), a spatial stochastic model for the telecommunication access networks. The SSLM represents the road system of the network by random

F. Voss
Boehringer-Ingelheim Pharma GmbH & Co. KG, 55216 Ingelheim, Germany
Tel.: +49-6132-7793504
Fax: +49-6132-7293504
E-mail: florian.voss@boehringer-ingelheim.com

C. Gloaguen
Orange Labs, 92131 Issy les Moulineaux, France
E-mail: catherine.gloaguen@orange-ftgroup.com

F. Fleischer
Boehringer-Ingelheim Pharma GmbH & Co. KG, 88397 Biberach, Germany
E-mail: frank.fleischer@boehringer-ingelheim.com

V. Schmidt
Ulm University, Institute of Stochastics
E-mail: volker.schmidt@uni-ulm.de

segment systems and both the locations of LLC and HLC are modelled by random point processes. In this context, the typical shortest path connection length C^* from LLC to their associated HLC is an important performance characteristic in the analysis of telecommunication networks. Formally, the distribution of C^* is defined in terms of Palm distributions ([1,16]). However, it can be shown that under some extra conditions the empirical distribution of the shortest path lengths of the LLC inside a sampling window W converges to the distribution of C^* if W increases unboundedly. So in view of this convergence result C^* can be considered as the shortest path from a LLC chosen at random among all LLC. Estimators for the mean typical shortest path length as well as the mean typical subscriber line length were introduced in [6]. In the present paper, we generalize these results in different ways, i.e., we consider a more general model and we introduce estimators for the density of the typical shortest path length which can be computed based on Monte Carlo simulations of the segment system inside the typical serving zone.

The paper is organized as follows. First, we introduce the SSLM and give exact mathematical definitons of all considered model components. In particular, we define the typical shortest path length C^*. Then, using Palm theory for stationary marked point processes, we derive a representation formula for the density of C^* which is based on some functional of the typical segment system S_H^* inside the typical serving zone. By means of this formula, we construct estimators for the density of C^* based on i.i.d. observations of S_H^* which can be obtained by Monte Carlo simulation. It is then shown that these estimators possess excellent statistical properties. Simulation algorithms for S_H^* are known for specific cases of the model ([2,4,19]) and we introduce new simulation algorithms for the case that the HLC are modelled by thinnings of the nodes of the underlying segment system. Finally, the results are evaluated in a numerical study.

2 Stochastic modelling of hierarchical telecommunication networks

In this introductory section we briefly describe the stochastic network model which is considered in the present paper. This model is suitable e.g. for applications to hierarchical telecommunication access networks [3,5,6,19]. The model is based on (marked) point processes and random tessellations. We briefly explain some mathematical background on these topics and introduce the notation we are using. Further details can be found e.g. in [1,15,16,17]. For recent surveys on the application of point processes, random tessellations and random geometric graphs in order to model and analyze telecommunication networks, see [8,23].

2.1 Marked point processes

Marked point processes can be used e.g. to model the locations of network components and subscribers in telecommunication networks. We first recall some basic results regarding marked point processes and introduce the notions we are using. Let $\mathcal{B}^2$ denote the Borel-σ-algebra in $\mathbb{R}^2$ and N the family of locally finite counting measures on $\mathcal{B}^2$ which are simple. This means that they have no multiple atoms. Each $\varphi \in \mathsf{N}$ can then be represented by the set $\{x_n\}$ of its atoms, i.e., if δ_x denotes the Dirac measure with $\delta_x(B) = 1$ if $x \in B$ and $\delta_x(B) = 0$ if $x \notin B$, it holds that $\varphi = \sum_n \delta_{x_n}$. The σ-algebra on N generated by the sets $\{\varphi \in \mathsf{N} : \varphi(B) = j\}$ for $j \geq 0$ and $B \in \mathcal{B}^2$ is denoted by $\mathcal{N}$ and we define the shift operator $\mathsf{T}_x : \mathsf{N} \mapsto \mathsf{N}$ by $\mathsf{T}_x\varphi(B) = \varphi(B + x)$ for $x \in \mathbb{R}^2$ and $B \in \mathcal{B}^2$. Here we write $B + x = \{x + y : y \in B\}$ for the set B shifted by the vector x. A point process X is then defined as a random element of the measurable space $(\mathsf{N}, \mathcal{N})$ and we identify X with the set $\{X_n\}$ of its (random) atoms. We sometimes write $X = \{X_n\}$ for brevity.

Now let $\mathbb{M}$ be some Polish space and let $\mathcal{B}_{\mathbb{M}}$ denote its Borel-σ-algebra. The family of all counting measures on $\mathcal{B}^2 \otimes \mathcal{B}_{\mathbb{M}}$ which are simple and in addition locally finite in the first component is then denoted by $\mathsf{N}_{\mathbb{M}}$. Again, the counting measure $\psi = \sum_n \delta_{(x_n,m_n)} \in \mathsf{N}_{\mathbb{M}}$ can be identified with its atoms (x_n, m_n) which now have two components: the location $x_n \in \mathbb{R}^2$ and the mark $m_n \in \mathbb{M}$. We can define the σ-algebra $\mathcal{N}_{\mathbb{M}}$ on $\mathsf{N}_{\mathbb{M}}$ in the same way as above, but the shift operator $\mathsf{T}_x : \mathsf{N}_{\mathbb{M}} \mapsto \mathsf{N}_{\mathbb{M}}$ translates now only the first component of the atoms of $\psi \in \mathsf{N}_{\mathbb{M}}$ by $-x$, i.e., $\mathsf{T}_x(\psi) = \sum_n \delta_{(x_n-x,m_n)}$. We then call a random element $X = \{(X_n, M_n)\}$ of $(\mathsf{N}_{\mathbb{M}}, \mathcal{N}_{\mathbb{M}})$ a marked point process.

2.2 Palm distributions; Neveu's exchange formula

We define stationarity, isotropy and ergodicity of (marked) point processes in the usual way. Let $X = \{(X_n, M_n)\}$ be a stationary marked point process with intensity $\lambda > 0$, i.e., $\lambda = \mathbb{E}\#\{n : X_n \in [0, 1)^2\}$. Then its Palm mark distribution $P_X^o : \mathcal{B}_{\mathbb{M}} \to [0, 1]$ is defined by

$$P_X^o(G) = \frac{\mathbb{E}\#\{n : X_n \in [0, 1)^2, M_n \in G\}}{\lambda}\,, \qquad G \in \mathcal{B}_{\mathbb{M}}\,.$$

We call a random variable M^*, which is distributed according to P_X^o, the typical mark of X. Furthermore, we define the Palm distribution $P_X^* : \mathcal{N}_{\mathbb{M}} \otimes \mathcal{B}_{\mathbb{M}} \to [0, 1]$ of X by

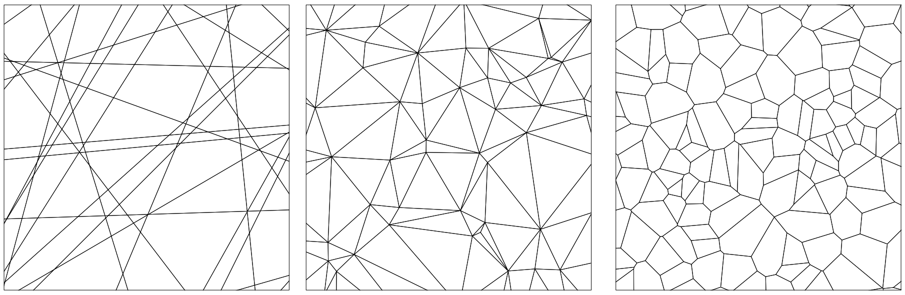
$$P_X^*(A \times G) = \frac{\mathbb{E}\#\{n : X_n \in [0, 1)^2, M_n \in G, \mathsf{T}_{X_n} X \in A\}}{\lambda}\,, \quad A \in \mathcal{N}_{\mathbb{M}}, G \in \mathcal{B}_{\mathbb{M}}\,.$$

Moreover, let $X^{(1)} = \{(X_n^{(1)}, M_n^{(1)})\}$ and $X^{(2)} = \{(X_n^{(2)}, M_n^{(2)})\}$ be two jointly stationary marked point processes with intensities λ_1 and λ_2 and mark spaces $\mathbb{M}_1$ and $\mathbb{M}_2$, respectively. Then we can regard $Y = (X^{(1)}, X^{(2)})$ as a random element of the product space $\mathsf{N}_{\mathbb{M}_1,\mathbb{M}_2} = \mathsf{N}_{\mathbb{M}_1} \times \mathsf{N}_{\mathbb{M}_2}$ and we can define the Palm distribution $P_Y^{(i)}$ of Y on $\mathcal{N}_{\mathbb{M}_1} \otimes \mathcal{N}_{\mathbb{M}_2} \otimes \mathcal{B}_{\mathbb{M}_i}$ with respect to the i-th component $X^{(i)}$ of Y, $i = 1, 2$, by

$$P_Y^{(i)}(A \times G) = \frac{\mathbb{E}\#\{n : X_n^{(i)} \in [0, 1)^2, M_n^{(i)} \in G, \mathsf{T}_{X_n^{(i)}} Y \in A\}}{\lambda_i}\,, \tag{1}$$

where $A \in \mathcal{N}_{\mathbb{M}_1} \otimes \mathcal{N}_{\mathbb{M}_2}$, $G \in \mathcal{B}_{\mathbb{M}_i}$ and the shift operator is defined by $\mathsf{T_x Y} = (\mathsf{t_x X^{(1)}}, \mathsf{t_x X^{(2)}})$. Notice that the Palm mark distribution $P_{X^{(i)}}^o$ of $X^{(i)}$ is a marginal distribution of $P_{X^{(i)}}^*$ and $P_Y^{(i)}$, respectively. Furthermore, note that an unmarked point process can be regarded as a marked point process whose marks are all identical, hence the Palm distribution given in (1) can also be defined for pairs of marked and unmarked point processes.

Later on, we use Neveu's exchange formula for jointly stationary marked point processes. Using the notation introduced above, and $\psi = (\psi^{(1)}, \psi^{(2)})$ for the elements of $\mathsf{N}_{\mathbb{M}_1,\mathbb{M}_2}$, this formula takes the following form (see e.g. [11, 14]).



(a) PLT (b) PDT (c) PVT

Fig. 1 Realizations of PLT, PDT and PVT.

Lemma 1 *For any measurable* $f : \mathbb{R}^2 \times \mathbb{M}_1 \times \mathbb{M}_2 \times \mathsf{N}_{\mathbb{M}_1,\mathbb{M}_2} \to [0,\infty)$*, it holds that*

$$\lambda_1 \int_{\mathsf{N}_{\mathbb{M}_1,\mathbb{M}_2}\times\mathbb{M}_1} \int_{\mathbb{R}^2\times\mathbb{M}_2} f(x, m^{(1)}, m^{(2)}, \mathsf{T}_x\psi)\ \psi^{(2)}(d(x, m^{(2)}))\, P_Y^{(1)}(d(\psi, m^{(1)}))$$
$$= \lambda_2 \int_{\mathsf{N}_{\mathbb{M}_1,\mathbb{M}_2}\times\mathbb{M}_2} \int_{\mathbb{R}^2\times\mathbb{M}_1} f(-x, m^{(1)}, m^{(2)}, \psi)\ \psi^{(1)}(d(x, m^{(1)}))\, P_Y^{(2)}(d(\psi, m^{(2)}))\,. \tag{2}$$

2.3 Random tessellations

We model the underlying infrastructure of a telecommunication network, e.g. road systems or railways, by the edge set of some random tessellation of $\mathbb{R}^2$. A random tessellation T is a partition $\{\Xi_n\}$ of $\mathbb{R}^2$ into random (compact and convex) polygons Ξ_n which are locally finite. The polygons Ξ_n are called the cells of T. Note that T can be regarded as a marked point processes $\{(\alpha(\Xi_n), \Xi_n^o)\}$, where $\alpha(\Xi_n)$ is a random vector such that $\alpha(\Xi_n) \in \Xi_n$ and $\alpha(\Xi_n + x) = \alpha(\Xi_n) + x$ for each $x \in \mathbb{R}^2$ which describes the location of the cell Ξ_n, and $\Xi_n^o = \Xi_n - \alpha(\Xi_n)$ is a copy of Ξ_n shifted by $-\alpha(\Xi_n)$, i.e., Ξ_n^o contains the origin o. We then call the random points $\alpha(\Xi_n)$ the nuclei of the cells Ξ_n of T. To each random tessellation we can define the point process $T^{(0)}$ which consists of the vertices of T. Furthermore, note that we can identify T with the edge set $T^{(1)} = \bigcup_n \partial\Xi_n$ of T.

The most frequently considered models for random tessellations are Poisson line tessellations (PLT), Poisson-Delaunay tessellations (PDT) and Poisson-Voronoi tessellations (PVT) which are all constructed based on Poisson processes, see Figure 1 for realizations. Details on these models can be found e.g. in [16].

Now suppose that T is stationary. Then we define the intensity γ of T as $\gamma = \mathbb{E}\nu_1(T^{(1)} \cap [0,1]^2)$, i.e., the mean length of $T^{(1)}$ per unit area, where we denote by ν_1 the 1-dimensional Hausdorff measure. Moreover, $T^{(0)}$ is a stationary point process if T is stationary. In the following the road model is assumed to be the edge set of a stationary random tessellation with intensity $\gamma \in (0,\infty)$.

2.4 Point processes on edges and vertices

For any T with intensity γ, we model the locations of HLC and LLC by point processes $H = \{H_n\}$ and $L = \{L_n\}$, respectively, which are concentrated on $T^{(1)}$ almost surely.

An important special case for H is given if H is a (conditional) Poisson process on $T^{(1)}$ with (linear) intensity λ_ℓ. Then, H is constructed by placing linear Poisson processes on the edges of T with (linear) intensity λ_ℓ. This means that H is a planar Cox process whose random driving measure $\Lambda : \mathcal{B}^2 \longrightarrow [0, \infty]$ is given by

$$\Lambda(B) = \lambda_\ell \nu_1(B \cap T^{(1)}), \qquad B \in \mathcal{B}^2 .$$

Note that in this case H is stationary, isotropic and ergodic if the tessellation T is stationary, isotropic and ergodic, respectively. Furthermore, if T is stationary, then the planar intensity λ of H is given by $\lambda = \lambda_\ell \gamma$.

Another special case for H is obtained from the point process $T^{(0)}$ of vertices of T by independent thinning according to a certain thinning probability $p \in (0, 1)$. If T is stationary and H a p-thinning of $T^{(0)}$, then H is stationary with planar intensity $\lambda = p\lambda^{(0)}$, where $\lambda^{(0)}$ is the intensity of $T^{(0)}$.

However, in general, we do not assume that H is a Poisson process on $T^{(1)}$, admitting that H is an arbitrary stationary point process on $T^{(1)}$ which is conditionally independent of L given T. For stationary H we define the quotient $\lambda_\ell = \lambda/\gamma$ as the planar intensity of H divided by the (length) intensity of $T^{(1)}$. Note that for a Poisson process H on $T^{(1)}$ the intensity quotient λ_ℓ is exactly the linear intensity of the Poisson processes on the edges of T.

In contrast to this, L is always assumed to be a Poisson process on $T^{(1)}$ with linear intensity λ'_ℓ, possibly different from λ_ℓ.

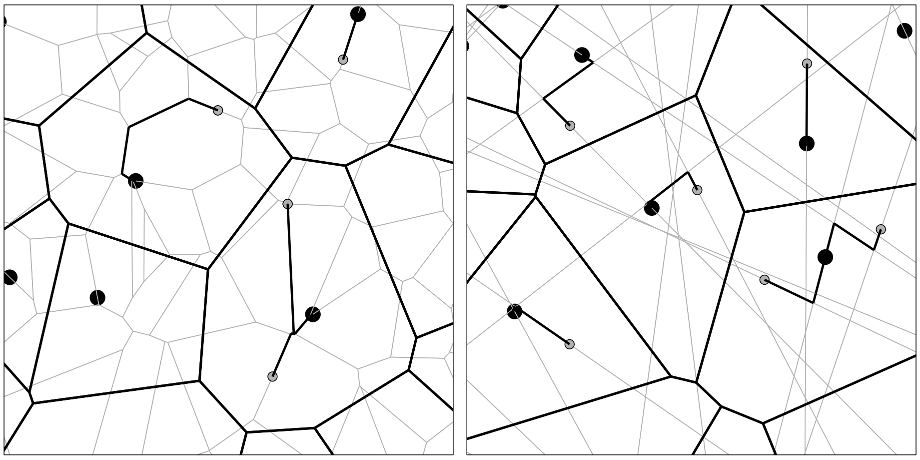
2.5 Service zones and shortest paths

Let $T_H = \{\Xi_{H,n}\}$ denote a random tessellation where the nuclei $\{\alpha(\Xi_{H,n})\}$ of the cells $\Xi_{H,n}$ are given by the points H_n of H. For instance, T_H can be the Voronoi tessellation induced by H, but further examples are possible like Laguerre tessellations ([9, 10]) or aggregated Voronoi tessellations ([18]). We then regard the cell $\Xi_{H,n}$ of T_H as the serving zone of the HLC located at the point H_n of H.

Using the four modelling components T, H, L and T_H introduced above, we can construct the marked point process $L_C = \{(L_n, C_n)\}$. Here the point L_n is marked with C_n, the length of the shortest path from L_n to H_j along the edge set $T^{(1)}$ provided that $L_n \in \Xi_{H,j}$. It can be shown that L_C is stationary if T and H are stationary. Realizations of serving zones and shortest paths are displayed in Figure 2(a) and (b) for a point process H on $T^{(1)}$, where T is a PVT and a PLT, respectively.

In the following we investigate the distribution of the typical mark C^* of L_C which we call the typical shortest path length, i.e., we are interested in the Palm mark distribution $P^o_{L_C}$ of L_C.

Note that we can construct the realizations of L_C from the corresponding realizations of L and H_S, where $H_S = \{(H_n, S^o_{H,n})\}$ denotes the stationary marked point process with marks given by $S^o_{H,n} = (T^{(1)} \cap \Xi_{H,n}) - H_n$. Thus, we can consider the vector $Y = (L, H_S)$ and its Palm distribution $P^{(1)}_Y$ with respect to L, which has been



(a) PVT as road model (b) PLT as road model

Fig. 2 Higher-level components with their serving zones (black) and lower-level components (grey with black boundary) with shortest paths (dashed) along the edge set of T (grey).

introduced in (1), instead of L_C. Let $(L^*, \widetilde{H}_S)$ denote a vector which is distributed according to $P_Y^{(1)}$, where we use the notation $\widetilde{H}_S = \{(\widetilde{H}_n, \widetilde{S}^o_{H,n})\}$, $\widetilde{H} = \{\widetilde{H}_n\}$, and

$$\widetilde{T}^{(1)} = \bigcup_{n\geq 1} \big(\widetilde{S}^o_{H,n} + \widetilde{H}_n\big)\,. \tag{3}$$

Then C^* can be regarded as the shortest path length from the origin o to its associated point of $\widetilde{H}$ along the edge set $\widetilde{T}^{(1)}$ given in (3).

In the same way we can consider $(\widetilde{L}, H^*_S)$ distributed according to $P_Y^{(2)}$ and we use the notion $T^{*(1)}$ for the edge set of H^*_S. Note that $\widetilde{L}$ is then a Poisson process on $T^{*(1)}$ with linear intensity λ'_ℓ which is independent of H^*_S given $T^{*(1)}$. Furthermore, it holds that

$$\mathbb{E}\nu_1(S^*_H) = \gamma/\lambda = 1/\lambda_\ell\,, \tag{4}$$

where S^*_H denotes the typical mark of H_S, i.e., the typical segment system inside the typical serving zone. This can be easily obtained from Campbell's theorem for stationary marked point processes which gives that

$$\begin{aligned}
\gamma &= \mathbb{E}\sum_{n=1}^{\infty} \nu_1\big((S^o_{H,n} + H_n) \cap [0,1)^2\big) \\
&= \lambda \int_{\mathbb{R}^2} \mathbb{E}\nu_1(S^*_H \cap ([0,1)^2 - x))\,\nu_2(dx) \\
&= \lambda\mathbb{E}\big[\int_{S^*_H}\int_{\mathbb{R}^2} \mathbb{1}_{[0,1)^2-x}(y)\,\nu_2(dy)\,\nu_1(dx)\big] \\
&= \lambda\mathbb{E}\nu_1(S^*_H)\,.
\end{aligned}$$

3 Distribution of typical shortest path length and its estimation

We now analyze the distribution of the typical shortest path length C^*. First we derive a representation formula for the density f_{C^*} of C^* which expresses the distribution of C^* in terms of the Palm distribution $P^*_{H_S}$ of H_S by the help of Neveu's exchange formula stated in Lemma 1. Using this representation formula we first show that the density is piecewise continuous and construct estimators for f_{C^*} based on i.i.d. samples of the typical segment system S^*_H. Such samples can be obtained from Monte Carlo simulation of S^*_H which is e.g. possible for Voronoi tessellations T_H if T is a PLT, PVT and a PDT, respectively, and H is a Poisson process on $T^{(1)}$ ([2,4,19]) or if H is a p-thinning of $T^{(0)}$, see Section 4. Finally, we discuss some useful properties of the proposed estimators for f_{C^*}.

3.1 Probability density of typical shortest path length

In this section we show a theorem which can be used to compute the probability density of the typical shortest path length C^*. In order to proof this theorem we first state an auxiliary result which allows us to represent the quantity $\mathbb{E}h(C^*)$ for each non-negative measurable function $h : \mathbb{R} \longmapsto [0,\infty)$ in terms of the Palm distribution of H_S.

Lemma 2 *Let $h : \mathbb{R} \longmapsto [0,\infty)$ be an arbitrary measurable function. Then,*

$$\mathbb{E}\, h(C^*) = \lambda_\ell \mathbb{E} \int_{S^*_H} h(c(y))\, \nu_1(dy)\,, \tag{5}$$

where H^*_S is distributed according to $P^*_{H_S}$, S^*_H denotes the typical segment system centred at o of H^*_S and $c(y)$ denotes the shortest path length from y to o along the segments of H^*_S.

Proof Let $\mathcal{L}$ be the family of all locally finite segment systems in $\mathbb{R}^2$ containing the origin. We apply Neveu's exchange formula (see Lemma 1) in order to show (5). Consider the function $f : \mathbb{R}^2 \times [0,\infty) \times \mathcal{L} \times \mathsf{N}_{[0,\infty),\mathcal{L}} \to [0,\infty)$ given by

$$f(x,c,\xi,\psi) = \begin{cases} h(c) & \text{if } x \in \xi\,, \\ 0 & \text{otherwise}\,. \end{cases}$$

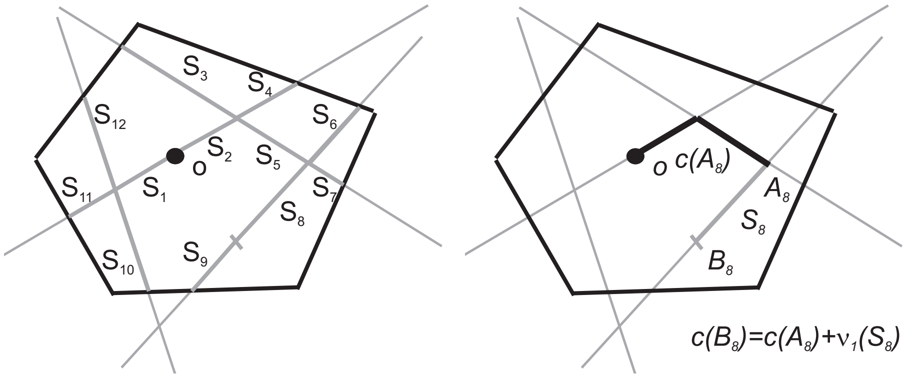
Then, using (2) with the function f defined above, we get that

$$\begin{aligned} \mathbb{E}h(C^*) &= \frac{\lambda}{\gamma\lambda'_\ell}\mathbb{E}\Big(\int_{S^*_H} h(c(x))\, \widetilde{L}(dx)\Big) \\ &= \frac{\lambda_\ell}{\lambda'_\ell}\mathbb{E}\Big(\mathbb{E}\Big(\int_{S^*_H} h(c(x))\, \widetilde{L}(dx) \;\Big|\; T^{*(1)}\Big)\Big)\,, \end{aligned}$$

where $\widetilde{L}$ is a (conditional) Poisson process, given $T^{*(1)}$, with linear intensity λ'_ℓ on $T^{*(1)}$. Thus Campbell's theorem yields

$$\mathbb{E}\Big(\int_{S^*_H} h(c(x))\, \widetilde{L}(dx) \;\Big|\; T^{*(1)}\Big) = \lambda'_\ell \int_{S^*_H} h(c(y))\nu_1(dy)\,,$$

which completes the proof of (5). □



(a) Segments of S_H^* (b) Split segment

Fig. 3 S_H^* divided into segments $S_1, \dots, S_N$

As can be seen in the proof of Lemma 2, the assumption that L is a Poisson process on $T^{(1)}$ is actually not essential for our results. In fact, it is sufficient that L is stationary, independent of H given T with $\mathbb{E}(L(B)|T^{(1)}) = \lambda'_\ell \nu_1(T^{(1)} \cap B)$ for all $B \in \mathcal{B}^2$. Furthermore, note that using (4) the representation of $\mathbb{E}h(C^*)$ given in (5) can also be written as

$$\mathbb{E}\, h(C^*) = \frac{1}{\mathbb{E}\nu_1(S_H^*)}\, \mathbb{E} \int_{S_H^*} h(c(y))\, \nu_1(dy)\,.$$

We also emphasize the important fact that the expectation $\mathbb{E}h(C^*)$ does not depend on the (linear) intensity λ'_ℓ of the conditional Poisson process L. Moreover, we can rewrite (5) as

$$\mathbb{E}\, h(C^*) = \lambda_\ell\, \mathbb{E} \sum_{i=1}^{N} \int_{c(A_i)}^{c(B_i)} h(u)\, du\,, \tag{6}$$

where the segment system S_H^* is divided into N line segments $S_1, \dots, S_N$ with endpoints $A_1, B_1, \dots, A_N, B_N$ such that $S_H^* = \bigcup_{i=1}^N S_i$ and $\nu_1(S_i \cap S_j) = 0$ for $i \neq j$, where $c(A_i) < c(B_i) = c(A_i) + \nu_1(S_i)$, see Figure 3. Some segments of S_H^* are split in this way at so-called distance peaks, where a point z on S_H^* is called a distance peak if there are two different shortest paths from z to o. Note that the random variable N appearing in (6) is the random number of line segments of S_H^* obtained after the splitting. It can be shown that

$$\mathbb{E}N \le a\, \mathbb{E}K^* < \infty\,, \tag{7}$$

where $a > 0$ is some constant and K^* denotes the number of segments emanating from the typical vertex of T. With the notation introduced above we can derive a representation formula for the probability density f_{C^*} of typical shortest path length which can be used in order to estimate this density.

Theorem 1 *The distribution of the typical shortest path length C^* is absolutely continuous, where the density is given by*

$$f_{C^*}(x) = \begin{cases} \lambda_\ell \mathbb{E} \sum_{i=1}^{N} \mathbb{1}_{[c(A_i),c(B_i))}(x) & \text{if } x \geq 0\,, \\ \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

Proof The distribution P_{C^*} of C^* can be written as $P_{C^*}(B) = \mathbb{E}\mathbb{1}_B(C^*)$ for $B \in \mathcal{B}(\mathbb{R})$. Thus, using (6) with $h(x) = \mathbb{1}_B(x)$, we get that for each $B \in \mathcal{B}(\mathbb{R})$

$$\begin{aligned} P_{C^*}(B) &= \lambda_\ell \mathbb{E} \sum_{i=1}^{N} \int_{c(A_i)}^{c(B_i)} \mathbb{1}_B(u)\, du \\ &= \lambda_\ell \mathbb{E} \sum_{i=1}^{N} \int_B \mathbb{1}_{[c(A_i),c(B_i))}(u)\, du \\ &= \int_B \lambda_\ell \mathbb{E} \sum_{i=1}^{N} \mathbb{1}_{[c(A_i),c(B_i))}(u)\, du\,, \end{aligned}$$

which proves the theorem. □

We now show that f_{C^*} given in (8) is right-continuous and has left-hand limits, i.e., f_{C^*} is a cadlag function on $[0, \infty)$.

Lemma 3 *The density f_{C^*} of C^* is a cadlag function on $[0,\infty)$. Furthermore, f_{C^*} is bounded and if H has no points in the vertex set $T^{(0)}$ of T, then $f_{C^*}(0) = 2\lambda_\ell$. If H is a p-thinning of $T^{(0)}$, then $f_{C^*}(0) = \lambda_\ell \mathbb{E}K^*$.*

Proof For each $x_0 \in \mathbb{R}$, it holds with probability 1 that

$$\lim_{x \searrow x_0} \sum_{i=1}^{N} \mathbb{1}_{[c(A_i),c(B_i))}(x) = \sum_{i=1}^{N} \mathbb{1}_{[c(A_i),c(B_i))}(x_0)\,.$$

Besides this, we have $\mathbb{E}N < \infty$, see (7). Thus, using (8), the dominated convergence theorem yields that f_{C^*} is right-continuous. Furthermore, it is easy to see that the function $g_{C^*} : [0,\infty) \to [0,\infty)$ with

$$g_{C^*}(x) = \lambda_\ell \mathbb{E} \sum_{i=1}^{N} \mathbb{1}_{(c(A_i),c(B_i)]}(x)\,, \qquad x \geq 0$$

is a version of the density f_{C^*} which is left-continuous with $f_{C^*}(x) = g_{C^*}(x)$ for each continuity point of f_{C^*}, i.e., the left-hand limits of f_{C^*} exist. Furthermore, $f_{C^*}(x) \leq \lambda_\ell \mathbb{E}N < \infty$ for each $x \in \mathbb{R}$, which shows that f_{C^*} is bounded. If H has almost surely no points at the vertices of T, then the origin o is almost surely located in the relative interior of a line segment of S_H^*. Thus there are almost surely 2 segments emanating from o and, consequently, $f_{C^*}(0) = 2\lambda_\ell$. On the other hand, if H coincides with an independently thinned vertex set of T, then the expected number of segments emanating from o under $P_{H_S}^*$ is equal to $\mathbb{E}K^*$ by definition, which completes the proof. □

Using the representation formula for $f_{C^*}(x)$ stated in Theorem 1, see (8), we are able to derive an estimator $\widehat{f}_{C^*}(x)$ for $f_{C^*}(x)$ which is based on data obtained by simulations of the segment system S_H^* within the typical serving zone Ξ_H^*. This is explained in more detail in the following section. Anyhow, we emphasize that by means of this representation formula we can estimate $f_{C^*}(x)$ without looking at the points of $\widetilde{L}$, i.e., the locations of lower–level components.

3.2 Estimation of the density of typical shortest path length

In this section we construct estimators for the density of typical shortest path length using the representation formula derived in Theorem 1. In particular, we are interested in estimators based on independent samples of the (typical) segment system S_H^* which can be obtained from Monte Carlo simulation for various models ([2,4,19]). Note that the shortest path can partly run outside Ξ_H^*, so the line segment system has to be simulated also outside Ξ_H^* up to a certain distance. Suppose that for some sample size $n > 0$ we have n i.i.d. copies $S_{H,1}^*, \ldots, S_{H,n}^*$ of S_H^*. For each $j = 1, \ldots, n$, we can then calculate the shortest path length for all nodes of $S_{H,j}^*$. If there are line segments with a distance peak in their interior, they are divided into two segments as mentioned in Section 3.1. In this way we get the segments $S_1^{(j)}, \ldots, S_{N_j}^{(j)}$ and the shortest path lengths from the endpoints of the line segments to o, i.e., $c(A_1^{(j)}), c(B_1^{(j)}), \ldots, c(A_{N_j}^{(j)}), c(B_{N_j}^{(j)})$. Based on this data, we can then construct two slightly different estimators. The first one is given by

$$\widehat{f}_{C^*}(x;n) = \lambda_\ell \, \frac{1}{n} \sum_{j=1}^{n} \sum_{i=1}^{N_j} \mathbb{1}_{[c(A_i^{(j)}), c(B_i^{(j)}))}(x) \; . \tag{9}$$

Furthermore, using (4), still another estimator can be constructed if the expected length of the line segment system in the typical cell is also estimated from simulated data. This leads to

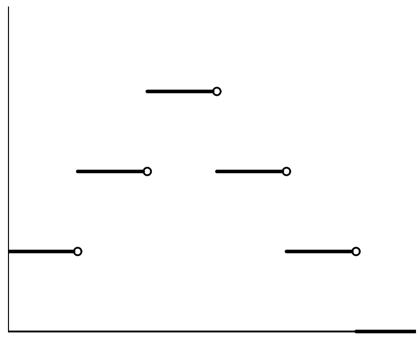
$$\widetilde{f}_{C^*}(x;n) = \frac{1}{\sum\limits_{j=1}^{n} \nu_1(S_{H,j}^*)} \sum_{j=1}^{n} \sum_{i=1}^{N_j} \mathbb{1}_{[c(A_i^{(j)}), c(B_i^{(j)}))}(x) \; . \tag{10}$$

Note that for practical computation of $\widehat{f}_{C^*}(x;n)$ and $\widetilde{f}_{C^*}(x;n)$ we can use the fact that $\widehat{f}_{C^*}(x;n)$ and $\widetilde{f}_{C^*}(x;n)$ are step functions with respect to x. For example, in order to compute $\widehat{f}_{C^*}(x;n)$, for each line segment $S_i^{(j)}$ with endpoints $A_i^{(j)}$ and $B_i^{(j)}$ we add the value λ_ℓ / n if $x \in [c(A_i^{(j)}), c(B_i^{(j)}))$, see Figure 4.

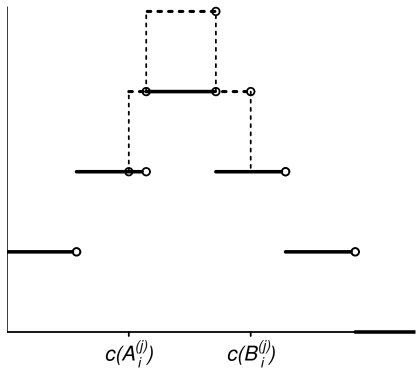
We now discuss some useful properties of the estimators $\widehat{f}_{C^*}(x;n)$ and $\widetilde{f}_{C^*}(x;n)$ introduced in (9) and (10), respectively.

Theorem 2 *For any $n > 0$ and $x \in \mathbb{R}$, $\widehat{f}_{C^*}(x;n)$ is an unbiased estimator of $f_{C^*}(x)$. Moreover, $\widehat{f}_{C^*}(x;n)$ and $\widetilde{f}_{C^*}(x;n)$ are strongly consistent for each $x \in \mathbb{R}$ provided that $n \to \infty$. For each measurable function $h : \mathbb{R} \longmapsto [0, \infty)$, it holds that*

$$\mathbb{E}\left[\int_{\mathbb{R}} h(x) \widehat{f}_{C^*}(x;n) \, dx\right] = \mathbb{E} h(C^*) \tag{11}$$



(a) $\widehat{f}_{C^*}(\cdot\,;n)$ before adding a segment



(b) Adding the contribution of a segment with endpoints $A_i^{(j)}$ and $B_i^{(j)}$

Fig. 4 Computation of the density $\widehat{f}(\cdot\,;n)$

and

$$\mathbb{P}\left(\lim_{n\to\infty}\int_{\mathbb{R}} h(x)\widehat{f}_{C^*}(x;n)\,dx = \mathbb{E}h(C^*)\right) = \mathbb{P}\left(\lim_{n\to\infty}\int_{\mathbb{R}} h(x)\widetilde{f}_{C^*}(x;n)\,dx = \mathbb{E}h(C^*)\right) = 1\,. \tag{12}$$

Proof The unbiasedness of $\widehat{f}_{C^*}(x;n)$ immediately follows from (8) and (9). Furthermore, the strong law of large numbers yields that $\lim_{n\longrightarrow\infty}\widehat{f}_{C^*}(x;n) = f_{C^*}(x)$ with probability 1 for each $x\in\mathbb{R}$. Similarly, in view of (4), we get that almost surely

$$\lim_{n\longrightarrow\infty}\frac{1}{n}\sum_{j=1}^{n}\nu_1(S^*_{H,j}) = \frac{1}{\lambda_\ell}\,.$$

Thus, altogether, we have $\lim_{n\longrightarrow\infty}\widetilde{f}_{C^*}(x;n) = f_{C^*}(x)$ with probability 1. Using the unbiasedness of $\widehat{f}_{C^*}(x;n)$, Fubini's theorem yields that

$$\mathbb{E}\int_{\mathbb{R}} h(x)\widehat{f}_{C^*}(x;n)\,dx = \int_{\mathbb{R}} h(x)\mathbb{E}\widehat{f}_{C^*}(x;n)\,dx \;=\; \int_{\mathbb{R}} h(x)f_{C^*}(x)\,dx \;=\; \mathbb{E}h(C^*)\,,$$

i.e., (11) holds. Furthermore, since

$$\int_{\mathbb{R}} h(x)\widehat{f}_{C^*}(x;n)\,dx = \frac{\lambda_\ell}{n}\sum_{j=1}^{n}\int_{\mathbb{R}}\sum_{i=1}^{N_j}\mathbb{1}_{[c(A_i^{(j)}),c(B_i^{(j)}))}(x)h(x)\,dx\,,$$

we get by the strong law of large numbers that the first probability in (12) is equal to 1. The second equality in (12) follows in the same way. □

Remark 1 Note that the estimator $\widetilde{f}_{C^*}(x;n)$ has some advantages, although it is not unbiased. For example, for each $n\geq 1$, we have

$$\int_{\mathbb{R}}\widetilde{f}_{C^*}(x;n)\,dx = 1\,,$$

whereas the integral over $\widehat{f}_{C^*}(x;n)$ is not equal to 1 in general.

Remark 2 Alternatively, for any bounded Borel set $W \subset \mathbb{R}^2$ with $\nu_2(W) > 0$, we can define estimators based on data in large sampling windows by

$$\widehat{f}_W(x;n) = \frac{\lambda_\ell}{\#\{j : H_j \in nW\}} \sum_{j \geq 1} \mathbb{1}_{nW}(H_j) \sum_{i=1}^{N_j} \mathbb{1}_{[c(A_i^{(j)}), c(B_i^{(j)}))}(x) , \qquad (13)$$

where $n \geq 1$ is any sufficiently large integer and N_j denotes the number of line segments $S_1^{(j)}, \ldots, S_{N_j}^{(j)}$ with endpoints $A_1^{(j)}, B_1^{(j)}, \ldots, A_{N_j}^{(j)}, B_{N_j}^{(j)}$ in the serving zone $\Xi_{H,j}$ of H_j. If $n \to \infty$, consistency results similar to those considered in Theorem 2 are valid for the estimator defined in (13) provided that $B(o, r) \subset W$ for some $r > 0$. This can be shown by means of the individual ergodic theorem for stationary point processes, see e.g. [1], Corollary 12.2.V. Nevertheless, we focus on the estimators introduced in (9) and (10) which are based on samples of the typical segment system S_H^*. These estimators are less computer-intensive. In addition, the estimator defined in (13) might suffer from the fact that neighboring serving zones are highly correlated and that there are always edge effects at the boundary of the sampling window. On the other hand, these problems can be avoided by regarding i.i.d. samples of S_H^*.

Remark 3 In [20], so-called capacities are investigated which are required at the points of Poisson processes on $T^{(1)}$. It is shown that the distribution of the typical capacity only depends on the length of a certain subtree within the shortest-path tree of the typical point of these conditional Poisson processes. This is the reason why an estimator for density $f_{\nu_1(T_{sub}^*)}$ of the typical subtree length $\nu_1(T_{sub}^*)$ is constructed in [20]. The estimator is defined by

$$\widehat{f}_{\nu_1(T_{sub}^*)}(x;n) = \lambda_\ell \, \frac{1}{n} \sum_{j=1}^{n} \sum_{i=1}^{N_j} \mathbb{1}_{[l(B_i^{(j)}), l(A_i^{(j)}))}(x) , \qquad (14)$$

where the segment system S_H^* is subdivided in the same way as in (9) and (10). Besides, $l(B_i^{(j)})$ denotes the subtree length at the endpoint $B_i^{(j)}$ of $S_i^{(j)}$, and $l(A_i^{(j)}) = \nu_1(S_i^{(j)}) + l(B_i^{(j)})$. Note that also the estimator given in (14) has useful statistical properties which are similar to those considered in Theorem 2.

3.3 Almost sure convergence of the maximum error

In Theorem 2 we proved that the estimator $\widehat{f}_{C^*}(x;n)$ of $f_{C^*}(x)$ is unbiased and consistent for each fixed $x \in \mathbb{R}$. We now show that $\widehat{f}_{C^*}(x;n)$ converges to $f_{C^*}(x)$ uniformly in x. In order to prove this result we need some further notation. For any $B \in \mathcal{B}(\mathbb{R})$ and $j = 1, \ldots, n$, let $Z_j(B) = \sum_{i=1}^{N_j} \big(\mathbb{1}_B(c(A_i^{(j)})) + \mathbb{1}_B(c(B_i^{(j)}))\big)$. Then $Z_1, \ldots, Z_n$ are i.i.d. (non-simple) point processes on $[0, \infty)$ with finite intensity measure since $\mathbb{E}Z_j(\mathbb{R}) \leq 2\mathbb{E}N_j < \infty$. Furthermore, let $\mathbb{Q} \subset \mathbb{R}$ denote the set of rational numbers.

Theorem 3 *It holds that*

$$\mathbb{P}\left(\lim_{n \longrightarrow \infty} \sup_{x \in \mathbb{R}} |\widehat{f}_{C^*}(x;n) - f_{C^*}(x)| = 0 \right) = 1 \,.$$

Proof It is sufficient to show that for almost all $\omega \in \Omega$ and $\varepsilon > 0$ there exists an integer $N(\varepsilon,\omega) \geq 1$ such that $|\widehat{f}_{C^*}(x;n,\omega) - f_{C^*}(x)| \leq \varepsilon$ for all $n > N(\varepsilon,\omega)$ and $x \in \mathbb{R}$. For each $q \in \mathbb{Q}$, we have

$$\begin{aligned}&|\widehat{f}_{C^*}(x;n,\omega) - f_{C^*}(x)| \\ &\quad \leq |\widehat{f}_{C^*}(x;n,\omega) - \widehat{f}_{C^*}(q;n,\omega)| + |\widehat{f}_{C^*}(q;n,\omega) - f_{C^*}(q)| + |f_{C^*}(q) - f_{C^*}(x)| \,.\end{aligned} \tag{15}$$

Now let $\varepsilon > 0$ and $q, q' \in \mathbb{Q}$ such that $q < q'$ and $x \in [q,q')$. Then we get that

$$|\widehat{f}_{C^*}(x;n,\omega) - \widehat{f}_{C^*}(q;n,\omega)| \leq \frac{\lambda_\ell}{n} \sum_{j=1}^{n} Z_j((q,q'),\omega)\,. \tag{16}$$

Furthermore, since the point processes $Z_1,\ldots,Z_n$ are i.i.d., we get from the strong law of large numbers that with probability 1

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} Z_j((q,q')) = \mathbb{E} Z_1((q,q'))\,.$$

On the other hand, since $\mathbb{E}Z_1([0,\infty)) < \infty$, $\lim_{x\longrightarrow\infty} f_{C^*}(x) = 0$ and f_{C^*} is a cadlag function on $[0,\infty)$, we can choose $q_0,\ldots,q_m \in \mathbb{Q}$ with $q_0 = 0 < q_1 < \cdots < q_m < q_{m+1} = \infty$ such that for all $i = 0,\ldots,m$ and $x \in [q_i, q_{i+1})$ we get that

$$\lambda_\ell \mathbb{E} Z_1((q_i,q_{i+1})) < \frac{\varepsilon}{3}\,, \quad |f_{C^*}(x) - f_{C^*}(q_i)| < \frac{\varepsilon}{3}\,. \tag{17}$$

Moreover, we then have almost surely

$$\lim_{n\longrightarrow\infty} \frac{\lambda_\ell}{n} \sum_{j=1}^{n} Z_j((q_i,q_{i+1})) = \lambda_\ell \mathbb{E} Z_1((q_i,q_{i+1})) < \frac{\varepsilon}{3}$$

and $\lim_{n\longrightarrow\infty} \widehat{f}_{C^*}(q_i;n,\omega) = f_{C^*}(q_i)$ for all $i = 0,\ldots,m$. This means that we can choose $N(\varepsilon,\omega) \geq 1$ such that

$$\frac{\lambda_\ell}{n} \sum_{j=1}^{n} Z_j((q_i,q_{i+1}),\omega) < \frac{\varepsilon}{3} \qquad \text{and} \qquad |\widehat{f}_{C^*}(q_i;n,\omega) - f_{C^*}(q_i)| < \frac{\varepsilon}{3} \tag{18}$$

for all $i = 0,\ldots,m$ and $n > N(\varepsilon,\omega)$. Hence, combining the inequalities in (15)–(18) we get for each $\varepsilon > 0$ and almost all $\omega \in \Omega$ that there exists $N(\varepsilon,\omega) \geq 1$ such that

$$|\widehat{f}_{C^*}(x;n,\omega) - f_{C^*}(x)| < \varepsilon$$

for all $x \in \mathbb{R}$ and $n \geq N(\varepsilon,\omega)$ which completes the proof. □

Remark 4 The result of Theorem 3 is also valid for the estimator $\widehat{f}_W(x;n)$ introduced in (13), provided that $B(o,r) \subset W$ for some $r > 0$. Then,

$$\mathbb{P}\left(\lim_{n\longrightarrow\infty} \sup_{x\in\mathbb{R}} |\widehat{f}_W(x;n) - f_{C^*}(x)| = 0\right) = 1\,.$$

Moreover, in a similar way, it can be shown that

$$\mathbb{P}\left(\lim_{n\longrightarrow\infty} \sup_{x\in\mathbb{R}} |\widehat{f}_{\nu_1(T^*_{sub})}(x;n) - f_{\nu_1(T^*_{sub})}(x)| = 0\right) = 1\,,$$

where $\widehat{f}_{\nu_1(T^*_{sub})}$ is the estimator defined in (14).

3.4 Rates of convergence and estimation variance

Using our approach we can avoid the simulation of LLC when estimating the probability density of C^*. This has some advantages with respect to performance of the estimator given in (9), in comparison to estimators based on explicit simulation of LLC and their shortest path lengths.

A common measure for the deviation of an estimator $\widehat{f}$ from a probability density f to be estimated is the mean integrated squared error of $\widehat{f}$ which is denoted by $MISE(\widehat{f})$, where

$$MISE(\widehat{f}) = \mathbb{E}\int_{\mathbb{R}} (\widehat{f}(x) - f(x))^2\, dx\,.$$

Note that for the estimator $\widehat{f}_{C^*}(x;n)$ of $f_{C^*}(x)$ introduced in (9) we have

$$\begin{aligned}
MISE(\widehat{f}_{C^*}) &= \mathbb{E}\int_{\mathbb{R}} (\widehat{f}_{C^*}(x;n) - f_{C^*}(x))^2\, dx \\
&= \frac{\lambda_\ell^2}{n}\int_{\mathbb{R}} \mathrm{Var}\Big(\sum_{i=1}^N \mathbb{1}_{[c(A_i),c(B_i))}(x)\Big)\, dx \\
&\le \frac{\lambda_\ell^2}{n}\int_{\mathbb{R}} \mathbb{E}\Big(\sum_{i,j=1}^N \mathbb{1}_{[c(A_i),c(B_i))}(x))\mathbb{1}_{[c(A_j),c(B_j))}(x)\Big)\, dx \\
&= \frac{\lambda_\ell^2}{n}\mathbb{E}\sum_{i,j=1}^N \nu_1([c(A_i),c(B_i))\cap[c(A_j),c(B_j))) \le \frac{\lambda_\ell^2}{n}\mathbb{E}(N\nu_1(S_H^*))\,.
\end{aligned}$$

Thus, the rate of convergence of $MISE(\widehat{f}(x;n))$ is of order $1/n$ if $\mathbb{E}(N\nu_1(S_H^*)) < \infty$. Suppose that we simulate the points of $\widetilde{L}$ on S_H^*, calculate the shortest path length $\widetilde{C}_i$ of each point $\widetilde{L}_i \in S_H^*$ and then construct a kernel estimator for f_{C^*}. Note that the computational effort for the kernel estimator is similar to the effort required for $\widehat{f}_{C^*}$ because for every point $\widetilde{L}_i$ we have to compute the shortest path lengths to the endpoints of the segment S_j with $\widetilde{L}_i \in S_j$. However, the rate of convergence of the mean integrated squared error of kernel estimators is of order $n^{-4/5}$ or slower, see [22]. Having this in mind, it becomes clear that the estimator $\widehat{f}_{C^*}$ is superior to kernel estimators.

We now consider an estimator $\widehat{F}_{C^*}(x)$ for the value $F_{C^*}(x)$ of the distribution function $F_{C^*}:\mathbb{R}\to[0,1]$ of C^* which is given by

$$\widehat{F}_{C^*}(x;n) = \frac{\lambda_\ell}{n}\sum_{i=1}^n \nu_1(S_{H,i}^*(x))\,, \tag{19}$$

where $S_{H,i}^*(x) = \{y\in S_{H,i}^* : c(y)\le x\}$ for $x\in\mathbb{R}$. Note that $\widehat{F}_{C^*}(x;n) = \int_0^x \widehat{f}_{C^*}(t;n)\,dt$, where $\widehat{f}_{C^*}(t;n)$ is the density estimator introduced in (9). Thus, Theorem 2 yields that the estimator $\widehat{F}_{C^*}(x)$ is unbiased and strongly consistent for $F_{C^*}(x)$. On the other hand, suppose that we simulate the LLC-points $\widetilde{L}_j^{(i)}$ of $\widetilde{L}^{(i)}$ on $S_{H,i}^*$ and compute the shortest path lengths $\widetilde{C}_j^{(i)}$ from the LLC-points to the origin. Then the natural estimator

$$\widetilde{F}_{C^*}(x;n) = \frac{\lambda_\ell}{\lambda'_\ell}\frac{1}{n}\sum_{i=1}^n \#\{\widetilde{L}_j^{(i)}\in S_{H,i}^* : \widetilde{C}_j^{(i)}\le x\}$$

is another unbiased estimator for $F_{C^*}(x)$, see the proof of Lemma 2. Furthermore, it holds that

$$\mathbb{E}(\widetilde{F}_{C^*}(x;n) \mid S^*_{H,i}, i=1,\dots,n) = \frac{\lambda_\ell}{\lambda'_\ell}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(\#\{\widetilde{L}^{(i)}_j \in S^*_{H,i} : \widetilde{C}^{(i)}_j \le x\} \mid S^*_{H,i}) .$$

Since $\widetilde{L}^{(i)}$ is a Poisson process with intensity measure $\lambda'_\ell \nu_1(\,\cdot\, \cap S^*_{H,i})$ given $S^*_{H,i}$, we can apply Campbell's theorem to get

$$\mathbb{E}(\#\{\widetilde{L}_{i,j} \in S^*_{H,i} : \widetilde{L}_{i,j} \le x\} \mid S^*_{H,i}) = \lambda'_\ell \nu_1(S^*_{H,i}(x)) .$$

Thus we have

$$\widehat{F}_{C^*}(x;n) = \mathbb{E}(\widetilde{F}_{C^*}(x;n) \mid S^*_{H,i}, i=1,\dots,n) ,$$

and hence $\operatorname{Var}\widehat{F}_{C^*}(x;n) \le \operatorname{Var}\widetilde{F}_{C^*}(x;n)$, which shows that the estimator $\widehat{F}_{C^*}(x;n)$ introduced in (19) is again superior to the natural estimator based on explicit measuring the shortest path lengths for $\widetilde{L}$.

3.5 Scaling invariance

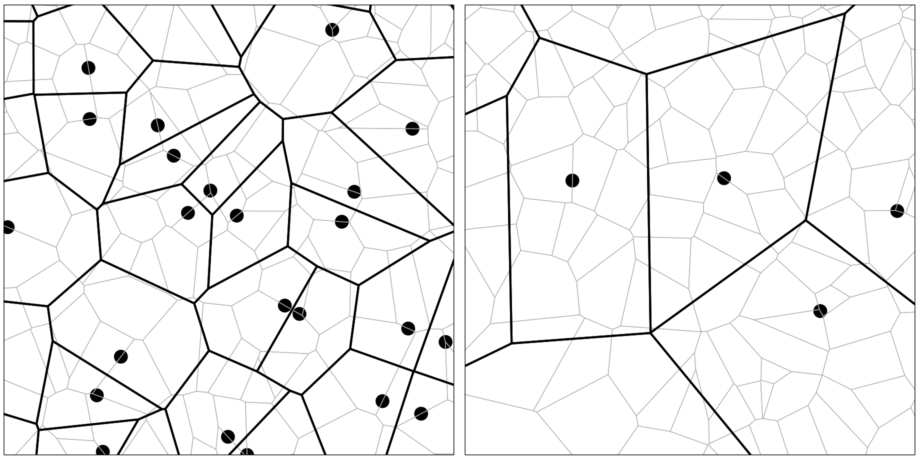
Suppose from now on that T is a PLT, PVT and PDT, respectively, and H is a Poisson process on $T^{(1)}$ or H is a p-thinning of the vertices $T^{(0)}$. Furthermore, assume that T_H is the Voronoi tessellation induced by H. For all these models, a scaling invariance property holds with respect to the two parameters $\lambda_\ell = \lambda/\gamma$ and γ if the value of the quotient $\kappa = \gamma/\lambda_\ell$ is constant. Note that κ determines the density of segments inside the serving zones. Small values of κ yield a sparse segment system, whereas large values of κ result in a dense segment system inside the serving zones. In Figure 5 realizations of serving zones are displayed for $\kappa = 20$ and 60 for a conditional Poisson process H on PVT.

Assume that $\gamma = a\widetilde{\gamma}$ and $\lambda_\ell = a\widetilde{\lambda}_\ell$ for some fixed values of $\widetilde{\gamma} > 0, \widetilde{\lambda}_\ell > 0$ and $a > 0$. Note that the scaling of the parameters by some $a > 0$ corresponds to a scaling of the model with parameters $\widetilde{\gamma}, \widetilde{\lambda}_\ell$ by the factor $1/a$. For instance, $S^*_H(\gamma, \lambda_\ell) \stackrel{d}{=} 1/a\, S^*_H(\widetilde{\gamma}, \widetilde{\lambda}_\ell)$, where $S^*_H(\gamma, \lambda_\ell)$ denotes the typical segment system with parameters γ and λ_ℓ. If we then e.g. regard the typical segment system S^*_H and the typical cell Ξ^*_H, we get that the expected number of segments of $S^*_H(\gamma, \lambda_\ell)$ stays constant, whereas $\mathbb{E}\nu_1(S^*_H(\gamma, \lambda_\ell))$ and $\sqrt{\mathbb{E}\nu_2(\Xi^*_H(\gamma, \lambda_\ell))}$ grow linearly proportional to $1/a$ for $\gamma = a\widetilde{\gamma}$ and $\lambda_\ell = a\widetilde{\lambda}_\ell$, see e.g. [2,4,6]. This scaling invariance can also be used to calculate the probability density $f_{C^*}(x;\gamma,\lambda_\ell)$ from the knowledge of the density $f_{C^*}(x;\widetilde{\gamma},\widetilde{\lambda}_\ell)$.

Theorem 4 *For any pair (γ, λ_ℓ) of parameters $\gamma, \lambda_\ell > 0$, consider the density of the shortest path length $f_{C^*}(x;\gamma,\lambda_\ell)$ given in (8). Then*

$$f_{C^*}(x;\gamma,\lambda_\ell) = a f_{C^*}(ax;\widetilde{\gamma},\widetilde{\lambda}_\ell)$$

if $\gamma/\lambda_\ell = \widetilde{\gamma}/\widetilde{\lambda}_\ell$ and $a > 0$ with $\lambda_\ell = a\widetilde{\lambda}_\ell$.



(a) $\kappa = 20$ (b) $\kappa = 60$

Fig. 5 Realizations of serving zones for different values of κ and PVT as road model.

Proof Recall that $S^*_H(\gamma, \lambda_\ell) \stackrel{d}{=} 1/a\, S^*_H(\widetilde{\gamma}, \widetilde{\lambda}_\ell)$. Thus, for each $x > 0$, we get that

$$\begin{aligned} f_{C^*}(x;\gamma,\lambda_\ell) &= \lambda_\ell \mathbb{E} \sum_{i=1}^{N} 1\!\!\mathrm{I}_{[c(A_i),c(B_i))}(x) \\ &= a\widetilde{\lambda}_\ell \mathbb{E} \sum_{i=1}^{\widetilde{N}} 1\!\!\mathrm{I}_{[c(\widetilde{A}_i),c(\widetilde{B}_i))}(a\,x) = a f_{C^*}(ax;\widetilde{\gamma},\widetilde{\lambda}_\ell)\,, \end{aligned}$$

where we used that the shortest path does not change if the model is scaled, but the shortest path length grows linearly as a 1-dimensional quantity. □

Because of the scaling invariance property discussed in this section, we can now concentrate our investigation on certain parameter pairs only. In the following we always assume that $\gamma = 1$ and consider different values of κ which cover realistic network scenarios. Numerical results corresponding to values of $\gamma \neq 1$ are then obtained by an appropriate scaling. Note that for conditional Poisson processes all values of $\kappa \in (0, \infty)$ are achievable. However, for p-thinnings of the vertices $T^{(0)}$ it is not possible to achieve arbitrary small values of κ since $\lambda_\ell = \lambda/\gamma$ which yields $\kappa = \gamma^2/(p\lambda^{(0)})$, where $\lambda^{(0)}$ denotes the intensity of $T^{(0)}$. Thus, $p = \gamma^2/(\kappa\lambda^{(0)})$ and hence $\kappa \geq \gamma^2/\lambda^{(0)}$. In particular, using known relationships between the intensities γ and $\lambda^{(0)}$ for PLT, PVT and PDT, see e.g. [17], Chapter 10, we have that $\gamma^2/\lambda^{(0)} = \pi$ for PLT, $\gamma^2/\lambda^{(0)} = 2$ for PVT and $\gamma^2/\lambda^{(0)} = (32/(3\pi))^2 \approx 11.53$ for PDT, respectively. Note that for $\gamma = 1$ we have $\kappa = 1/\lambda_\ell = \mathbb{E}\nu_1(S^*_H)$ if H is a Poisson process on $T^{(1)}$ or a thinning of $T^{(0)}$, respectively.

4 Simulation of the typical serving zone and its segment system

We focus on the two special cases that H is a Poisson process on the edges $T^{(1)}$ and a p-thinning of the vertices $T^{(0)}$ of T, respectively. Furthermore, we assume that the serving zones T_H are constructed as the Voronoi tessellation induced by H.

Note that simulation algorithms for S_H^* are then available if H is a conditional Poisson process and T is either a PDT, PLT or PVT ([2,4,19]). These algorithms were used in order to obtain the numerical results discussed below. Now let H be a p-thinning of $T^{(0)}$, then S_H^* can be simulated using the ideas explained in the following.

If T is a PDT, then $T^{(0)}$ is a Poisson process. Thus, due to Slyvniak's theorem which states that the Palm distribution of a Poisson process X is obtained by adding the origin o to X, we have that T^* is the Delaunay tessellation with respect to the point process $X^* = X \cup \{o\}$ and S_H^* can be simulated as follows.

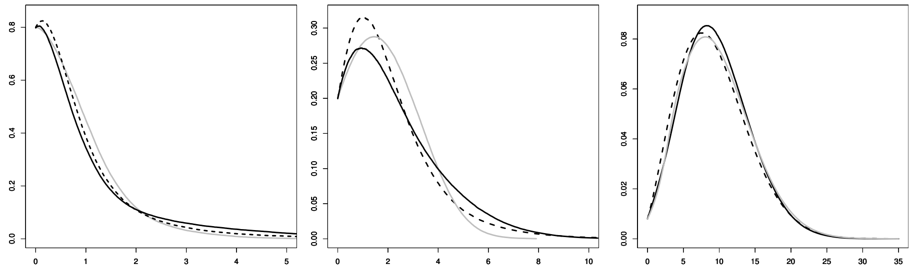
1. Simulate a stationary Poisson process $X = \{X_n\}$ radially, construct $X^* = X \cup \{o\}$ and construct the Delaunay tessellation T^* with respect to X^*.
2. Define $H^* = \{o\}$ and add X_i to H^* if $U_i < p$, where $U_1, U_2, \ldots$ are i.i.d. with $U_i \sim U[0,1)$.
3. Construct the Voronoi cell $\varXi_H^*$ at o with respect to H^*.
4. Construct $S_H^* = T^{*(1)} \cap \varXi_H^*$.

Now let T be a PLT, then we have to simulate T^* with respect to the Palm distribution of $T^{(0)}$, i.e., under the condition that there are two lines which contain the origin. It is known that the angle $\varPhi$ between the two lines ℓ_1 and ℓ_2 which contain the typical point of $T^{(0)}$ is distributed according to the density $f_\varPhi(x) = \sin(x)/2$ for $x \in [0, \pi)$, see e.g. [13]. Furthermore, ℓ_1 and ℓ_2 are isotropic. Thus, S_H^* can be simulated in the following way.

1. Simulate angles $\varPhi_1 \sim U[0, \pi)$ and $\varPhi \sim f_\varPhi$ and generate lines ℓ_1 and ℓ_2 through o with angles $\varPhi_1$ and $\varPhi_2 = \varPhi_1 + \varPhi$.
2. Simulate a stationary and isotropic Poisson line process $\{\ell_3, \ell_4, \ldots\}$ independent of ℓ_1 and ℓ_2. Define $T^* = \{\ell_1, \ell_2, \ell_3, \ldots\}$.
3. Construct $\{X_i\}$ as the union of all intersection points in $\mathbb{R}^2 \backslash \{o\}$ of pairs of lines $\ell_j, \ell_k \in T^*$ with $j \neq k$.
4. Define $H^* = \{o\}$ and add X_i to H^* if $U_i < p$, where $U_1, U_2, \ldots$ are i.i.d. with $U_i \sim U[0,1)$.
5. Construct the Voronoi cell $\varXi_H^*$ at o with respect to H^*.
6. Construct $S_H^* = T^{*(1)} \cap \varXi_H^*$.

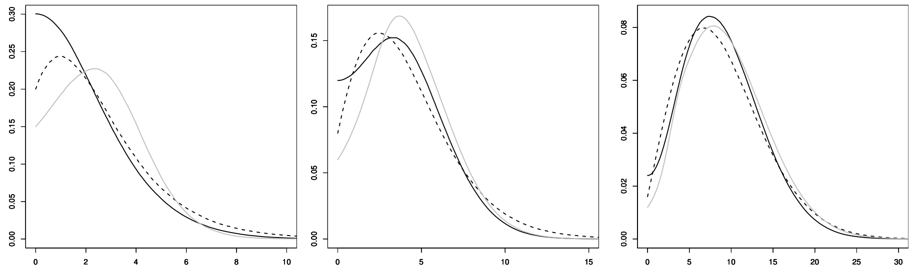
Finally, if T is a PVT, then we can proceed as follows. The duality of PVT and PDT yields that the dual tessellation of T^* is a PDT with respect to the Palm distribution of its nuclei. Thus, we can simulate a PDT starting from its typical cell and then construct the dual tessellation in order to obtain T^*. Note that the distribution of the typical cell $\varXi_D^*$ of a PDT is known. Let $Y_1 = RZ_1, Y_2 = RZ_2$ and $Y_3 = RZ_3$ denote the three vertices of $\varXi_D^*$, where $Z_i, i = 1, 2, 3$ are unit vectors which we identify with their polar angles. Then R is distributed according to the density $f_R(x) = 2\lambda^2\pi^2x^3\exp(-\lambda\pi x^2)$ for $x \geq 0$ and the joint density of (Z_1, Z_2, Z_3) is given by $\nu_2(\mathrm{conv}\{z_1, z_2, z_3\})/(12\pi^2)$ for $z_1, z_2, z_3 \in [0, 2\pi)^3$, see Theorem 10.4.4 in [16]. This yields the following simulation algorithm for S_H^*.

1. Simulate Y_1, Y_2 and Y_3 and an independent and stationary Poisson process $Y = \{Y_4, Y_5, \ldots\}$ in $\mathbb{R}^2 \backslash B(o, R)$, where $R = \| Y_1 \|$.
2. Construct T^* as the Voronoi tessellation with respect to $\{Y_i; i \geq 1\}$.
3. Define $H^* = \{o\}$ and add $X_i \in T^{*(0)} \backslash \{o\}$ to H^* if $U_i < p$, where $U_1, U_2, \ldots$ are i.i.d. with $U_i \sim U[0,1)$.
4. Construct the Voronoi cell $\varXi_H^*$ at o with respect to H^*.



(a) $\kappa = 2.5$ (b) $\kappa = 10$ (c) $\kappa = 250$

Fig. 6 Density for $\gamma = 1, \kappa = 2.5, 10, 250$ and Poisson process on PVT (grey), PDT (black), PLT (broken)



(a) $\kappa = 20$ (b) $\kappa = 50$ (c) $\kappa = 250$

Fig. 7 Density for $\gamma = 1, \kappa = 20, 50, 250$ and thinned vertices of PVT (grey), PDT (black), PLT (broken)

5. Construct $S_H^* = T^{*(1)} \cap \Xi_H^*$.

Note that for all three simulation algorithms introduced above some technical details have to be considered. For instance, we have to simulate the points radially and we have to find good stopping criteria for the simulation. We omit such details here, but similar problems are discussed e.g. in [2,4,19] for simulation algorithms of typical serving zones if H is a conditional Poisson process.

5 Numerical results

In this section we estimate the probability density $f_{C^*}(x)$ of the typical shortest path length for different models. We regard a random tessellation T which is a PDT, PLT and PVT, respectively, with (length) intensity γ. Furthermore, as already mentioned above, we distinguish two different cases for H. On the one hand, H is assumed to be a Poisson process on $T^{(1)}$ with linear intensity λ_ℓ and, on the other hand, H is the point process constructed from the vertices $T^{(0)}$ of T by independent thinning according to a probability $p \in (0,1)$. Finally, we regard the Voronoi tessellation T_H induced by H. Before we present the numerical results, we first recall that the considered models are scaling invariant, i.e., for any $\lambda_\ell, \gamma > 0$ with fixed quotient $\kappa = \gamma/\lambda_\ell$ we get the same structure of the model, only on a different scale.

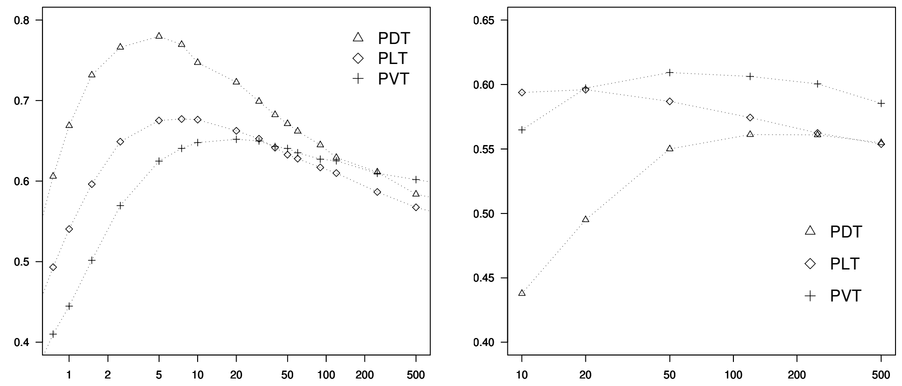
In order to estimate the density $f_{C^*}(x)$ of typical shortest path length we simulated $n = 50000$ segment systems within the typical Voronoi cell for different values of κ and conditional Poisson processes as well as p-thinnings of $T^{(0)}$ for PDT, PLT and PVT as road models. Based on these simulations, we estimated the density $\widehat{f}_{C^*}(x;n)$ as explained in Section 3.2. Some empirical densities which were estimated in this way are displayed in Figures 6 and 7. One can see that there is a clear difference between the shapes of the densities for small and large κ as well as for the different models considered in this paper. The difference between the models seems to decrease with increasing κ, but it is still noticeable. In [21] it is shown that for conditional Poisson processes H the typical shortest path length converges in distribution to ξZ for $\kappa \to \infty$, where $Z \sim Wei(\lambda\pi, 2)$ and $\xi \geq 1$ is some constant depending on the tessellation model, e.g., $\xi = 1$ for PLT and $\xi > 1$ for PDT and PVT. So there will always remain some differences between the densities as $\kappa \to \infty$ if H is a conditional Poisson process. We expect the same behavior for p-thinnings.

Based on the estimated densities we computed the means and coefficients of variation (where cv $Z = 100 \cdot \sqrt{\mathrm{Var} Z}/\mathbb{E}Z$). In Figure 8(a) the means are displayed for a conditional Poisson process H. First it is interesting that for small values of κ the mean typical shortest path length $\mathbb{E}C^*$ is the smallest for PVT and the largest for PDT. Then, for increasing κ, things change and $\mathbb{E}C^*$ is smaller for PLT than for PVT. Finally, for large κ, $\mathbb{E}C^*$ is the smallest for PLT and the largest for PVT. This is the intuitively expected behavior since, compared to PDT and PLT, the edges of PVT are shorter and there are more nodes where the shortest path has to change its direction. For PLT the direct Euclidean distance is obtained in the limit for $\kappa \to \infty$ ([21]), so the shortest path length has to be the shortest for PLT and PDT is somewhere in the middle.

In Figure 8(b) the means are displayed if H is a thinning of $T^{(0)}$. This time, $\mathbb{E}C^*$ is the smallest for PDT and the largest for PLT if κ is small. For increasing κ things change again and we have that $\mathbb{E}C^*$ is the smallest for PDT and the largest for PVT. Finally, for $\kappa > 250$, $\mathbb{E}C^*$ is the smallest for PLT and the largest for PVT as it is the case for large κ if H is a conditional Poisson process. If we compare the means $\mathbb{E}C^*$ for H being a conditional Poisson process and a thinning, respectively, we see that $\mathbb{E}C^*$ is smaller for thinnings. This difference is the largest for PDT which can be explained by the fact that in the average there are 6 segments emanating from each vertex and hence all LLC on these segments have the optimal Euclidean distance as the shortest path length. This difference is less for PVT and PLT since only 3 and 4 segments are emanating from each vertex, but it is still observable.

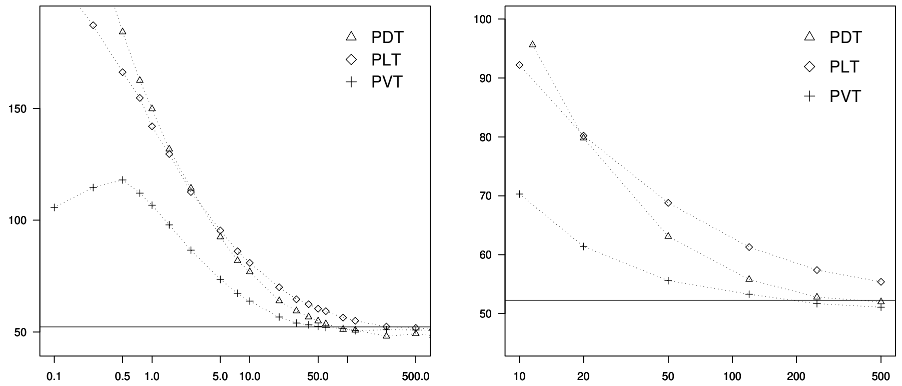
The cv is displayed in Figure 9(a) for a conditional Poisson process H on PVT, PDT and PLT, respectively. One can see that cv C^* approaches cv $Z = 52.27$ for large κ for all three models as expected. On the other hand, the behavior for small κ is quite different for the three models. Note that C^* converges to Z for all three models if $\kappa \to 0$ and λ_ℓ is fixed, where $Z \sim Exp(2\lambda_\ell)$, see [21]. So cv C^* should converge to cv $Z = 100$ for $\kappa \to 0$. However, this is only the case for PVT, for the values of κ considered in our study. The reason might be that the convergence for PDT and PLT is much slower, e.g., for PLT it is likely that even for small κ a line not containing the origin intersects the typical serving zone. Then points on this additional intersecting line can have very long shortest path lengths. With regard to equation (8) we then get heavier tails of the distribution yielding a higher cv compared to e.g. PVT.

For HLC resulting from thinnings of $T^{(0)}$ the cv is shown in Figure 9(b). The behavior is quite similar as for conditional Poisson processes. However, note that the



(a) Conditional Poisson process with $\gamma\lambda_\ell = 1$ (b) Thinned vertices of T with intensity 1

Fig. 8 $\mathbb{E}C^*$ for Voronoi tessellation T_H based on different models of H.

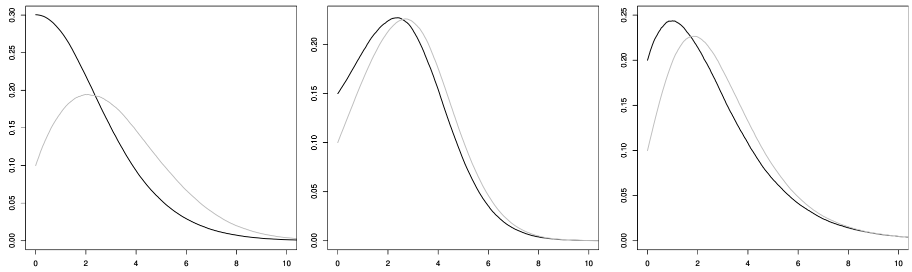


(a) Conditional Poisson process with $\gamma\lambda_\ell = 1$ (b) Thinned vertices of T with intensity 1

Fig. 9 cv C^* (horizontal axis) plotted versus κ for Voronoi tessellation T_H and different models of H. The horizontal lines are at $52.27 = \text{cv } Z$, where $Z \sim Wei(\lambda\pi, 2)$.

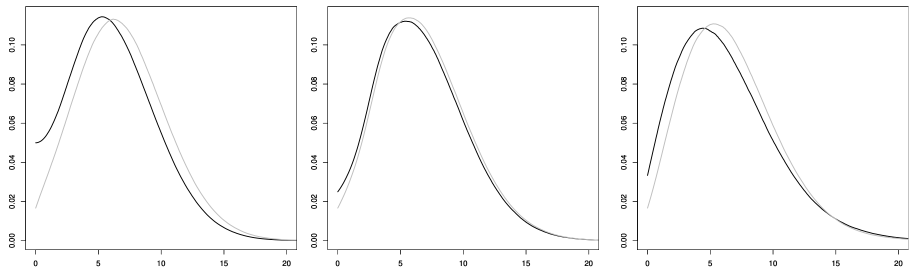
limit for $\kappa \to 0$ cannot be considered here. Compared to conditional Poisson processes, the cv is larger for thinnings which might be caused by the fact that the densities for thinnings seem to have heavier tails, see e.g. Figures 10 and 11.

In Figures 10 and 11 the densities are shown for different values of κ and for PDT, PLT and PVT, where H is a conditional Poisson process and a thinning of $T^{(0)}$, respectively. One can see that there is a large difference between the densities corresponding to conditional Poisson processes and thinnings which is decreasing with increasing κ. However, except for PVT, the difference is still noticeable even for larger values of κ. This difference is the largest for PDT, where e.g. at 0 the value of the density for thinnings is three times the value of the density for conditional Poisson processes, see also Lemma 3.



(a) PDT (b) PVT (c) PLT

Fig. 10 Density for $\gamma = 1, \kappa = 20$ for thinned vertices (black) and conditional Poisson process (grey)



(a) PDT (b) PVT (c) PLT

Fig. 11 Density for $\gamma = 1, \kappa = 120$ for thinned vertices (black) and conditional Poisson process (grey)

6 Discussion and Outlook

In this paper we introduced an estimator for the density of the typical shortest path length based on the Monte Carlo simulation of the typical serving zone. This estimator generalizes estimators for the mean typical shortest path length introduced in [6]. We have shown that this estimator possesses good statistical properties. For instance, we have shown that the maximum difference between the estimated density and the true density converges to zero almost surely and that we can estimate functionals of the typical shortest path length unbiasedly. Furthermore, we have shown that our estimator has lower variance and better convergence rates than kernel estimators.

We then used the estimator to compare the distribution of the typical shortest path length for different models. In order to estimate the density, we used simulation algorithms for the typical serving zone of conditional Poisson processes introduced in [2,4,19] and we introduced new simulation algorithms for the typical serving zone and its segment system if the locations of HLC are modelled by the thinned set of vertices of the underlying tessellation. We compared the densities, means and coefficients of variation and demonstrated that there are clear differences between the distribution of typical shortest path length for the different models considered in the present paper. Thus, we have shown that our model is flexible and can describe the distribution of connection distances for various settings.

With the techniques developed in this paper we can obtain empirical densities for specific models by extensive simulations. However, for practical applications it is a great advantage to have parametric densities which are directly available in an analytical form, where the parameters only depend on the scaling factor κ and the type of the underlying tessellation T. Then, for given road data, an optimal model can be chosen together with its parameters by the fitting techniques introduced in [5]. Thus, in [7] we applied the methods developed in this paper in order to construct a whole library of parametric distributions for the typical shortest path lengths for PDT, PLT and PVT as road models and a large range of κ. Therefore, for various values of κ, the density was estimated for PDT, PLT and PVT as road models and for conditional Poisson processes as model for the locations of HLC. In the next step, parametric densities were fitted to the estimated ones. The parametric densities f used for that purpose were chosen such that f fulfilled $f(0) = 2\lambda_\ell$ because of Lemma 3. A good choice seemed to be the densities of specific truncated Weibull distributions. These densities have two parameters which were fitted by a weighted least squares method. As a result parametric densities for shortest path lengths are now available whose parameters can be chosen based on the considered road model and the value of κ. In order to validate these parametric distributions with real data, the optimal road model was chosen for given data and the parametric densities corresponding to this model were then compared to histograms of real connection distances. The comparison revealed a very good fit of the parametric distributions, although so far only simple tessellations (PVT, PDT, PLT) were considered. Further details can be found in [7]. Note that in the same way parametric densities for C^* can be obtained if the locations of HLC are modelled by p-thinnings of $T^{(0)}$.

In our opinion, the above summarized techniques provide an efficient method for the analysis of telecommunication access networks. In future work, simulation algorithms for the typical cell together with the typical segment system have to be developed for more sophisticated tessellation models like iterated tessellations ([12]) based on PDT, PVT and PLT in order to obtain parametric densities for these models. With these models we expect to achieve even better fits for distance distributions observed in real networks than those obtained in [7] with simple tessellations.

Acknowledgements This research was supported by Orange Labs through research grant No. 46143714.

References

1. D. J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes*, vol. I and II. Springer, New York, 2005/2007.
2. F. Fleischer, C. Gloaguen, V. Schmidt, and F. Voss. Simulation of the typical Poisson-Voronoi-Cox-Voronoi cell. *Journal of Statistical Computation and Simulation*, 79(7):939–957, 2009.
3. C. Gloaguen, P. Coupé, R. Maier, and V. Schmidt. Stochastic modelling of urban access networks. In *Proc. 10th Internat. Telecommun. Network Strategy Planning Symp.*, pages 99–104. (Munich, June 2002), VDE, Berlin, 2002.
4. C. Gloaguen, F. Fleischer, H. Schmidt, and V. Schmidt. Simulation of typical Cox-Voronoi cells, with a special regard to implementation tests. *Mathematical Methods of Operations Research*, 62:357–373, 2005.
5. C. Gloaguen, F. Fleischer, H. Schmidt, and V. Schmidt. Fitting of stochastic telecommunication network models, via distance measures and Monte-Carlo tests. *Telecommunication Systems*, 31:353–377, 2006.

6. C. Gloaguen, H. Schmidt, R. Thiedmann, J.-P. Lanquetin, and V. Schmidt. Comparison of network trees in deterministic and random settings using different connection rules. In *Proceedings of the 5th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks* (WiOpt'07), Limassol, Cyprus 2007.
7. C. Gloaguen, F. Voss, and V. Schmidt. Parametric distance distributions for fixed access network analysis and planning. In *Proceedings of the 21st International Teletraffic Congress*, Paris, France, 2009.
8. M. Haenggi, J. G. Andrews, F. Baccelli, O. Dousse, and M. Franceschetti. Stochastic geometry and random graphs for the analysis and design of wireless networks. *IEEE Journal on Selected Areas in Communications*, 27:1029–1046, 2009.
9. C. Lautensack. *Random Laguerre Tessellations.* PhD thesis, Universität Karlsruhe (TH), 2007.
10. C. Lautensack and S. Zuyev. Random Laguerre tessellations. *Advances in Applied Probability*, 40:630–650, 2008.
11. R. Maier, J. Mayer, and V. Schmidt. Distributional properties of the typical cell of stationary iterated tessellations. *Mathematical Methods of Operations Research*, 59:287–302, 2004.
12. R. Maier and V. Schmidt. Stationary iterated random tessellations. *Advances in Applied Probability*, 35:337–353, 2003.
13. R. E. Miles. Random polygons determined by random lines in a plane. *Proc. Natl. Acad. Sci. (USA)*, 52:901–907, 1964.
14. J. Neveu. Sur les mesures de palm de deux processus ponctuels stationnaires. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 34:199–203, 1976.
15. A. Okabe, B. Boots, K. Sugihara, and S. N. Chiu. *Spatial Tessellations.* J. Wiley & Sons, Chichester, second edition, 2000.
16. R. Schneider and W. Weil. *Stochastic and Integral Geometry.* Springer, Berlin, 2008.
17. D. Stoyan, W. S. Kendall, and J. Mecke. *Stochastic Geometry and its Applications.* J. Wiley & Sons, Chichester, second edition, 1995.
18. K. Tchoumatchenko and S. Zuyev. Aggregated and fractal tessellations. *Probability Theory Related Fields*, 121:198–218, 2001.
19. F. Voss, C. Gloaguen, F. Fleischer, and V. Schmidt. Distributional properties of Euclidean distances in wireless networks involving road systems. *IEEE JSAC Issue on Stochastic Geometry and Random Graphs for Wireless Networks*, 27:1047–1055, 2009.
20. F. Voss, C. Gloaguen, and V. Schmidt. Capacity distributions in spatial stochastic models for telecommunication networks. *Preprint*, 2009 (to appear in *Image Analysis and Stereology*).
21. F. Voss, C. Gloaguen, and V. Schmidt. Scaling limits for shortest path lengths along the edges of stationary tessellations. *Preprint*, 2009 (submitted).
22. M. P. Wand and M. C. Jones. *Kernel Smoothing.* Chapman & Hall, London, 1995.
23. S. Zuyev. Stochastic geometry and telecommunications networks. In W. S. Kendall and I. Molchanov, editors, *New Perspectives in Stochastic Geometry.* Oxford University Press, Oxford, 2009.