

# Joint distributions for total lengths of shortest-path trees in telecommunication networks

David Neuhäuser · Christian Hirsch ·  
Catherine Gloaguen · Volker Schmidt

Received: date / Accepted: date

**Abstract** Shortest-path trees play an important role in the field of optimising fixed-access telecommunication networks with respect to costs and capacities. Distributional properties of the corresponding two half-trees originating from the root of such a tree are of special interest for engineers. In the present paper, we derive parametric approximation formulas for the marginal density functions of the total lengths of both half-trees. Besides, a parametric copula is used in order to combine the marginal distributions of these functionals to a bivariate joint distribution as, naturally, the total lengths of the half-trees are not independent random variables. Asymptotic results for infinitely sparse and infinitely dense networks are discussed as well.

**Keywords** Shortest-path tree · Palm calculus · parametric copula · tree length · network planning · pseudo-maximum-likelihood · stochastic geometry

**Mathematics Subject Classification (2000)** 60D05 · 65C99 · 62E17

## 1 Introduction

The *Stochastic Subscriber Line Model*, as described in [2], [4] and [14], is a powerful tool to model access networks in urban areas. In [9], it has recently been extended by so-called *shortest-path trees* which allow engineers and telecommunication companies such as *Orange Labs* not only for estimating cable lengths but

---

This work was supported by Orange Labs through Research grant No. 46146063-9241. Christian Hirsch was supported by a research grant from DFG Research Training Group 1100 at Ulm University. Parts of the numerical results were obtained by the help of Jan Sommer.

---

D. Neuhäuser, C. Hirsch and V. Schmidt  
Ulm University, Helmholtzstraße 18, 89069 Ulm, Germany  
Tel.: +49-(0)7315023555  
E-mail: david.neuhaeuser@uni-ulm.de

C. Gloaguen  
Orange Labs, 38-40 rue du Général Leclerc, 92794 Issy-Moulineaux Cedex 9, France  
Tel.: +33-(0)145296441

also for calculating capacity demands in telecommunication networks. Since the geometry of such trees can be extremely tricky and complex, the idea of attacking this problem is to build models of increasing elaborateness which enable to describe shortest-path trees in further and further details. A first step towards this goal has been done in [9], where we focused on the bivariate distribution of the lengths of the main branches of each half-tree. The present paper can be seen as a continuation and extension of this research process. Consequently, we study further important functionals of the shortest-path tree, namely the total lengths of the corresponding half-trees.

The paper is organised as follows. In Section 2, we introduce the mathematical tools which allow for modelling telecommunication networks by means of stochastic geometry. In particular, random tessellations, point processes and Palm calculus play an important role in this context. Sections 3 and 4 deal with modelling the total lengths of the half-trees in the shortest-path tree, where we proceed in a similar way as in [9], using parametric approximation formulas for the marginal distributions of the total lengths of the half-trees. However, in contrast to the situation considered in [9], it now turns out that the marginal distributions investigated in the present paper are much more difficult to handle and as a consequence, the fitting procedure is no longer just straightforward. Furthermore, we employ parametric copulas in order to combine these univariate distributions as the half-trees (and therefore their total lengths) cannot be assumed to be independent. Asymptotic results for infinitely sparse and infinitely dense networks are discussed as well. Then, in Section 5, we validate the model in two ways. On the one hand, we perform a visual validation which is comparable to the one considered in [9]. Additionally, on the other hand, we also provide a statistically rigorous distributional test for our model based on a multivariate version of the Wald-Wolfowitz test introduced in [3]. Section 6 finally concludes the paper and gives an outlook to possible future research.

## 2 Mathematical Preliminaries

As already mentioned in the introduction, the *Stochastic Subscriber Line Model* (SSLM) uses tools of stochastic geometry to model copper or optical fibre networks in urban but also rural regions. The SSLM consists of four layers, which will be -for the convenience of the reader- briefly discussed in the following.

### 2.1 Random Tessellations of Poisson Type

The first layer of the SSLM concerns the geometrical support. With the intention to reach as many customers as possible, cables and fibres are installed along the road system of cities and towns. Modelling road systems in the SSLM is carried out by the usage of connected, stationary random graphs  $T$ , such as various types of random tessellations or  $\beta$ -skeletons, see [10]. The present paper is restricted to random tessellations based upon a homogeneous Poisson point process in  $\mathbb{R}^2$ , namely Poisson-Voronoi tessellations (PVT), Poisson-Delaunay tessellations (PDT) as well as Poisson line tessellations (PLT). To get more information about these well-known random geometric graphs, the reader is referred to [12] and [13].

## 2.2 Cox Processes Representing Network Nodes

The second layer of the SSLM consists of the placement of the network components. Note that we consider two-hierarchy-level networks, i.e., investigated path properties are always between a higher-level component (HLC) and a lower-level component (LLC), such as paths between service area interfaces and subscribers. All network components are assumed to be located on the underlying infrastructure represented by a stationary random tessellation  $T$ . In particular, we make use of a stationary Cox process  $X_H = \{X_{H,i}\}$  in order to model the locations of HLC. The random intensity measure of  $X_H$  is concentrated on the edge set  $T^{(1)}$  of the tessellation  $T$  and is proportional to the one-dimensional Hausdorff measure  $\nu_1$  on  $T^{(1)}$ . More precisely, we have  $\mathbb{E}X_H(B) = \lambda_\ell \mathbb{E}\nu_1(B \cap T^{(1)})$  for each Borel set  $B \subset \mathbb{R}^2$  and linear intensity  $\lambda_\ell > 0$ . Note that the planar intensity of  $X_H$  is thus given by  $\lambda_H = \lambda_\ell \gamma$  where  $\gamma = \mathbb{E}\nu_1([0, 1]^2 \cap T^{(1)})$  denotes the total length of  $T^{(1)}$  per unit area. For additional information about point processes, the reader can consult e.g. [1] and [6]. Note that if the intensity of the LLC is sufficiently high, then the corresponding point process differs only negligibly from the underlying road system. Due to this approximation, we only consider certain sub-networks defined by the HLC.

## 2.3 Cox-Voronoi Cells as Serving Zones and Typical Cells

Although we have the underlying infrastructure and the network components modelled, we still have to specify how HLC and LLC – any point of the underlying infrastructure in the present setting – are connected to each other. The third layer of the SSLM, the topological part, defines these connection rules. We link each LLC to its nearest HLC in the Euclidean sense. More formally, we define the tessellation  $\{\Xi_{H,i}\}_{i \geq 1}$  as the Voronoi tessellation generated by the points of the Cox process  $X_H = \{X_{H,i}\}$  of higher-level components (also called Cox-Voronoi tessellation). Then, each LLC located inside a so-called *serving zone*  $\Xi_{H,i}$  is connected to its corresponding HLC  $X_{H,i}$  along the edges of the underlying tessellation  $T$ . Nodes are connected via shortest paths, i.e., the distance between LLC and corresponding HLC measured along  $T^{(1)}$  is the smallest possible one. For further details on connection rules, the reader is referred to [10].

Note that the supply of the network for one single user is often not as important as the expected network quality averaged over all users in a large area. Palm calculus and Palm distributions provide a convenient formalisation of this problem. Therefore, the Palm version  $X_H^*$  of  $X_H$  is of special interest in this context. Its distribution can be interpreted as conditional distribution of  $X_H$  given that there is a HLC located at the origin  $o = (0, 0)^\top \in \mathbb{R}^2$ . More formally, the distribution of  $X_H^*$  is determined by the representation formula

$$\mathbb{E}g(X_H^*) = \frac{1}{\lambda_H} \mathbb{E} \sum_{i: X_{H,i} \in [0, 1]^2} g(\{X_{H,n}\} - X_{H,i}),$$

where  $g : \mathbb{L} \rightarrow [0, \infty)$  is an arbitrary measurable function and  $\mathbb{L}$  denotes the family of all locally finite sets of  $\mathbb{R}^2$ . Furthermore, we define the *typical serving zone*  $\Xi_H^*$

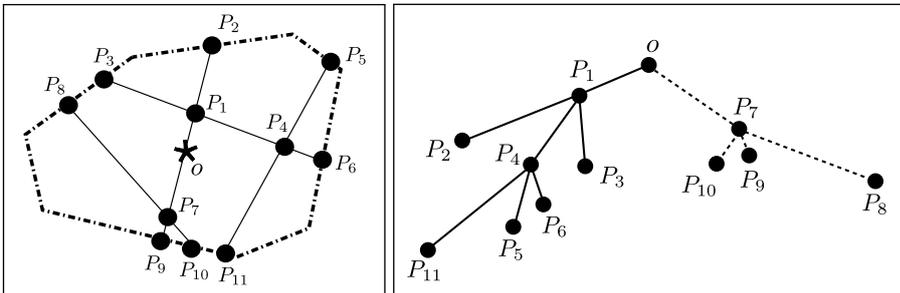
of  $X_H$  as the Voronoi cell whose cell centre is located at the origin  $o$  in the Voronoi tessellation constructed from  $X_H^*$ , i.e.,

$$\Xi_H^* = \{x \in \mathbb{R}^2 : \|x\| \leq \|x - X_{H,j}^*\| \text{ for all } j \geq 1\}.$$

Besides, in order to construct the shortest-path tree, we are interested in a further characteristic of the Voronoi tessellation constructed from  $X_H^*$ , the so-called *typical segment system*  $S_H^*$ . To define this, let  $S_{H,i} = \Xi_{H,i} \cap T^{(1)}$  denote the segment system of the serving zone  $\Xi_{H,i}$  which belongs to the corresponding HLC  $X_{H,i}$ . Then, the typical segment system  $S_H^*$  is defined as the typical mark of the point process of higher-level nodes  $X_{H,i}$  marked with their corresponding segment systems  $S_{H,i}$ . For definitions and details on marked point processes, Palm mark distribution and typical marks, the reader is referred to [1] and [13].

## 2.4 Shortest-Path Tree Representing the Inner Life of a Serving Zone

Under our assumption on the intensity of LLC, the fourth layer of the SSLM, the so-called shortest-path tree  $G$ , is roughly speaking just a rearrangement of the typical segment system  $S_H^*$ . Following the paths from all points in  $S_H^*$  to the origin  $o$  in the typical serving zone  $\Xi_H^*$  induces a natural tree structure. Note that these paths can sometimes leave  $\Xi_H^*$ . For further details on this tree, see [9]. Besides, one should bear in mind that  $G$  can be subdivided into two subtrees, say  $G_1^h$  and  $G_2^h$ , as the origin  $o$  (representing the root of the tree) lies with probability 1 in the interior of a segment of  $S_H^*$ . Note that  $G_1^h$  is characterised as the half-tree containing the longest branch of the shortest-path tree. For an illustration of the situation, see Figure 2.1.



**Fig. 2.1** Typical cell  $\Xi_H^*$  (dot-dashed) and typical segment system  $S_H^*$  (solid) on the left-hand side and corresponding shortest path tree  $G$  with its half-trees  $G_1^h$  (solid) and  $G_2^h$  (dashed) on the right-hand side

## 3 Modelling Approach

One specific functional of  $G$  which has already been discussed in our previous work [9] is the so-called skeletal backbone, i.e., the main branches in the half-tree  $G_1^h$

and  $G_2^h$ . In the present paper, we want to go one step further and examine the total lengths of  $G_1^h$  and  $G_2^h$ . The goal is to derive a general approach to model the joint bivariate distribution of the functional  $I = (I_1, I_2)$ , where

$$I_i = \nu_1(G_i^h), \quad i \in \{1, 2\}. \quad (3.1)$$

### 3.1 Bivariate Copulas

Note that, analogously to the skeletal backbone, the random functionals considered in (3.1) are certainly not independent. Thus, they prevent the usage of a bivariate distribution function given by

$$F_{(I_1, I_2)}(x_1, x_2) = F_{I_1}(x_1) \cdot F_{I_2}(x_2).$$

In this context, copulas provide a powerful framework to model the correlation structure of the functionals  $I_1, I_2$ , and this methodology has already been successfully applied in [9]. For the convenience of the reader, we briefly recall the definition of a bivariate copula. A function  $C : [0, 1]^2 \rightarrow [0, 1]$  is called *bivariate copula* if there exists a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  supporting a random vector  $U = (U_1, U_2)$  such that

$$C(u_1, u_2) = \mathbb{P}(U_1 \leq u_1, U_2 \leq u_2), \quad u_1, u_2 \in [0, 1],$$

and  $U_i \sim U[0, 1]$  for  $i \in \{1, 2\}$ . For further details on copulas, the reader is referred to [7] and [8]. Now, the bivariate joint distribution function of the random vector  $I = (I_1, I_2)$  can be written as

$$F_I(x) = C_I(F_{I_1}(x_1), F_{I_2}(x_2)),$$

where  $x = (x_1, x_2)$ ,  $x_1, x_2 > 0$ . In order to obtain a parsimonious parametric model applicable for any values of the linear intensity  $\lambda_\ell$  of HLC and the length intensity  $\gamma$  of  $T$ , we utilise a family of parametric copulas. One possible tool for fitting parametric models to given datasets is the maximum-likelihood method. Suppose we consider parametric models  $F_{I_1}(\cdot; \eta_1)$ ,  $F_{I_2}(\cdot; \eta_2)$  and  $C_I(\cdot; \eta)$  with parameter vectors  $\eta_1$ ,  $\eta_2$  and  $\eta$  for the marginals as well as for the copula, respectively, and assume an i.i.d. sample  $I^{(j)} = (I_1^{(j)}, I_2^{(j)})$ ,  $j = 1, \dots, n$  where  $n$  denotes the sample size. Then, the log-likelihood function  $\text{loglik}_1$  has representation

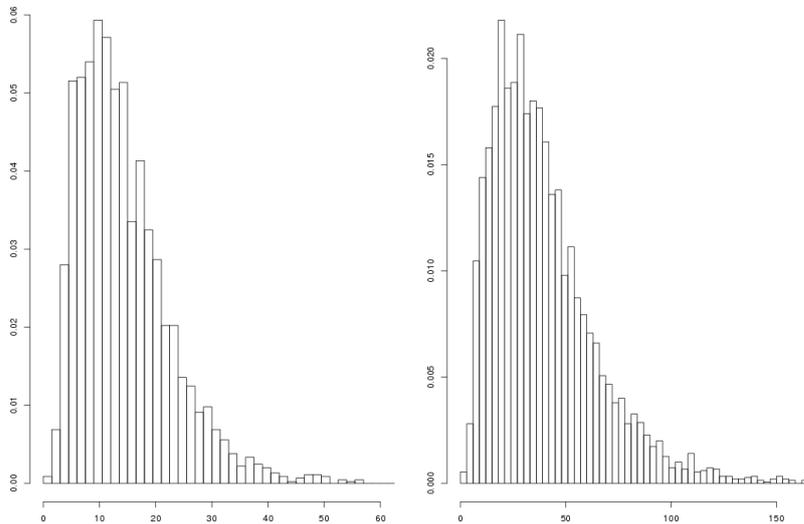
$$\begin{aligned} & \text{loglik}_1(\eta_1, \eta_2, \eta) \\ &= \sum_{j=1}^n \left( \log f_{I_1}(I_1^{(j)}; \eta_1) + \log f_{I_2}(I_2^{(j)}; \eta_2) \right. \\ & \quad \left. + \log \left[ c_I \left( F_{I_1}(I_1^{(j)}; \eta_1), F_{I_2}(I_2^{(j)}; \eta_2); \eta \right) \right] \right), \end{aligned} \quad (3.2)$$

where  $c_I(\cdot; \eta)$  denotes the density function of  $C_I(\cdot; \eta)$ . An uncomfortable disadvantage of this tool is that while computing maximum-likelihood estimates of the parameters of the bivariate distributions describing  $I_1$  and  $I_2$ , numerical problems can occur due to the inherent optimisation problem that involves several parameters. In the following, we therefore make use of a *parametric pseudo-maximum-likelihood approach* to avoid such difficulties. In particular, the optimisation problem can be subdivided into two steps which are considered separately. First, we

determine the univariate marginal distributions of  $I_1$  and  $I_2$ . Second, a parametric copula is fitted in order to model their correlation structure. For further details on the pseudo-maximum-likelihood approach, see [9] and especially [11].

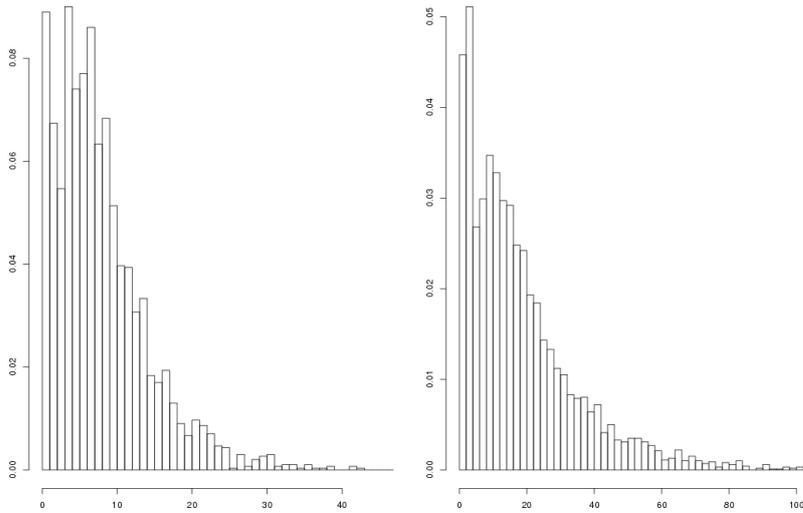
### 3.2 Fitting Procedure for the Marginal Densities

The fitting procedure for the marginal density of  $I_1$  is quite different from the procedure for  $I_2$ . On the one hand, in order to fit an approximative parametric representation formula for the density of  $I_1$ , we can proceed by using ordinary maximum-likelihood estimation and by manual choice of an eligible class of commonly used distributions. Looking at the histograms of the datasets, we can use e.g. gamma distributions, Weibull distributions, etc., see Figure 3.1. Results are provided in Section 4. Note that similar to [15], a certain scaling invariance in our model can be shown for underlying PVT, PDT or PLT and therefore it suffices to investigate the dependence of the distribution of  $I_i$  on the scaling parameter  $\kappa = \frac{\gamma}{\lambda_\ell}$ .



**Fig. 3.1** Histograms of the functional  $I_1$  with underlying PVT for  $\kappa = 20$  (left) and with underlying PDT for  $\kappa = 50$  (right)

Finding an approximative parametric representation formula for the density of  $I_2$  on the other hand is a much more challenging task. In Figure 3.2, the mindful reader detects a very high peak at the left hand side of each histogram which makes a density fitting by commonly used methods and parametric distributions (as we did in [9]) nearly impossible. This effect is caused by the fact that if the origin, representing the root of  $G$ , is located close to the boundary of the typical serving zone  $\Xi_H^*$ , the number and lengths of segments in  $G_2^h$  are very small. This, in turn, causes a very small value of  $I_2$ . The effect decreases for increasing values of



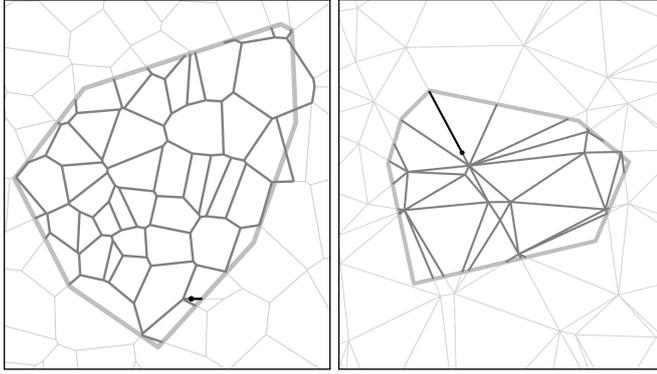
**Fig. 3.2** Histograms of the functional  $I_2$  with underlying PVT for  $\kappa = 20$  (left) and with underlying PDT and  $\kappa = 50$  (right)

$\kappa$  and vanishes for  $\kappa$  sufficiently large as the network becomes denser and denser. Figures 3.3 and 3.4 show two constellations, where the origin is located close to the boundary and deeply in the interior of the typical serving zone  $\Xi_H^*$ , respectively. Furthermore, this effect depends on the type of the underlying tessellation  $T$ . Therefore, we decided to subdivide the range  $R$  of the scaling parameter  $\kappa$  into two disjoint intervals  $R_{\text{peak}}$ , where this effect can be observed, and  $R \setminus R_{\text{peak}}$ , where it does not appear. Then, we fit an approximative parametric density for  $I_2$  for each of these two intervals separately. In particular, we have the following possible constellations which are shown in Table 3.1.

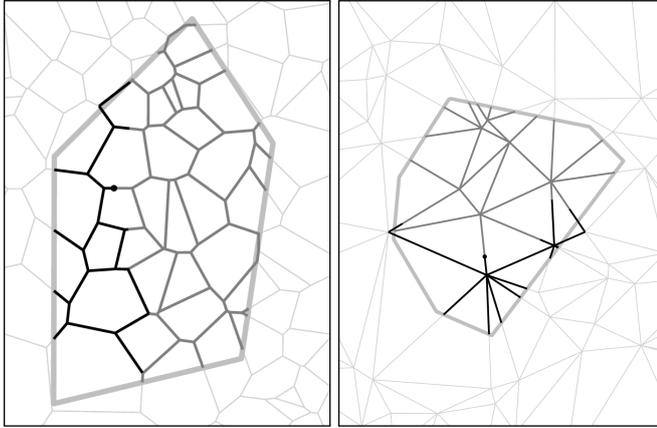
**Table 3.1** Subintervals  $R_{\text{peak}}$  and  $R \setminus R_{\text{peak}}$  of  $\kappa$ , depending on the type of the underlying tessellation  $T$

$T$	$R_{\text{peak}}$	$R \setminus R_{\text{peak}}$
PVT	$[10, 40]$	$(40, 1000]$
PDT	$[10, 90]$	$(90, 1000]$
PLT	$\emptyset$	$[10, 1000]$

So as to manage the difficulties caused by scenarios where the origin is located closely to the boundary of  $\Xi_H^*$  in the range  $R_{\text{peak}}$  of  $\kappa$ , we subdivide the data into two parts: scenarios where  $G_2^h$  consists of one single segment and scenarios where  $G_2^h$  consists of several segments. For each of the two cases, we fit parametric densities via the ordinary maximum-likelihood method. The parametric density function in  $R_{\text{peak}}$  is therefore a mixture of these two densities with a mixing parameter  $\rho \in [0, 1]$  which denotes the probability that  $G_2^h$  consists of several segments. It can be easily estimated by taking the ratio of the number of observed scenarios



**Fig. 3.3** Scenario 1:  $G_2^h$  (black) consisting of a single segment in PVT case (left) and PDT case (right)



**Fig. 3.4** Scenario 2:  $G_2^h$  (black) consisting of several segments in PVT case (left) and PDT case (right)

where  $G_2^h$  consists of several segments and the total number of realisations in the dataset. This kind of attacking the 'origin is close to boundary problem' turns out to provide very good fitting results, see Section 5. In the range  $R \setminus R_{\text{peak}}$ , where  $\kappa$  is sufficiently large, it is quite easy to fit a parametric density via the maximum-likelihood method. Afterwards, in order to obtain the parametric density for the whole range  $R$  of  $\kappa$ , we add the two densities multiplied by the indicator functions of their corresponding ranges  $R_{\text{peak}}$  and  $R \setminus R_{\text{peak}}$  (see Section 4.1.2). Note that in the PLT case, we have no subdivision of  $R$ , i.e.,  $R = R \setminus R_{\text{peak}}$ .

We also remark that the methodology of separately considering the scenario where the origin is closely located to the boundary can also be useful in the analysis of further important quantities of  $G$ . For instance, in [9], it is observed that the distribution of the longest branch of  $G_2^h$  is of bimodal type where the location of the first mode is mainly determined by this 'boundary scenario'. However, in [9],

standard statistical methods were sufficient to obtain parametric density functions which provide a satisfactory approximation to the data.

### 3.3 Fitting Procedure for the Copula

The copula types considered in the present paper are the t-copula, the Gaussian copula and the Archimedean copulas Clayton, Gumbel, Joe, Ali-Mikhail-Haq as well as Frank (for further explanations, see [7] and [8]). The fitting procedure for the copula is done by means of Akaike's information criterion defined as

$$AIC = 2(p - \text{loglik}_2(\hat{\eta})), \quad (3.3)$$

where  $p$  denotes the number of parameters in the model and  $\text{loglik}_2(\hat{\eta})$  is the maximised log-likelihood where the log-likelihood function is given by

$$\text{loglik}_2(\eta) = \sum_{j=1}^n \log \left[ c_I \left( \hat{F}_{I_1}(I_1^{(j)}), \hat{F}_{I_2}(I_2^{(j)}); \eta \right) \right]. \quad (3.4)$$

Note that (3.4) is obtained from (3.2) by omitting summands not depending on  $\eta$  and by replacing the marginal distribution functions  $F_{I_i}$  by their empirical distribution functions  $\hat{F}_{I_i}$ . The copula type which minimises (3.3) is chosen to work with.

## 4 Results

We now present results for the joint bivariate distribution of  $I = (I_1, I_2)$  which is obtained by the parametric pseudo-maximum-likelihood approach discussed in Section 3.

### 4.1 Parametric Approximation Formulas for the Marginal Density Functions

Due to the different approaches when fitting parametric formulas to the density functions  $f_{I_1}$  and  $f_{I_2}$ , we obtain quite different results for the half-trees  $G_1^h$  and  $G_2^h$ . For  $f_{I_1}$  we can use the same type of distribution for all three types of the underlying tessellation  $T$  whereas for  $f_{I_2}$ , we have to distinguish between the types of  $T$ . In particular, we obtain the following results.

#### 4.1.1 Parametric Approximation Formula for $f_{I_1}$

For all three types of the underlying tessellation  $T$ , namely PVT, PDT and PLT, it approximately holds that  $I_1 \sim \text{Gamma}(k, \lambda)$  with some shape parameter  $k > 0$  and some scale parameter  $\lambda > 0$ . In particular, the density function of  $I_1$  is given by

$$f_{I_1}(x) = \frac{1}{\lambda^k} \frac{1}{\Gamma(k)} x^{k-1} e^{-\frac{x}{\lambda}},$$

where  $\Gamma(\cdot)$  denotes the gamma function and  $x > 0$ . Note that the parameters  $k$  and  $\lambda$  obtained by the fitting procedure both depend on the scaling parameter  $\kappa$  and on the type of  $T$ .

#### 4.1.2 Parametric Approximation Formula for $f_{I_2}$

Concerning approximation formulas for  $f_{I_2}$ , we have to distinguish between the different types of tessellation. The approximation formulas for  $f_{I_2}$  are based on the following three types of distributions which are given by

1. the lognormal distribution with density  $f_{\log N}(x) = \frac{1}{\sqrt{2\pi x\sigma}} \exp\left(-\frac{(\log x - \nu)^2}{2\sigma^2}\right)$  for some  $\nu \in \mathbb{R}$  and  $\sigma, x > 0$  (short:  $\log N(\nu, \sigma)$ ),
2. the Nakagami distribution with density  $f_{Naka}(x) = \frac{2\mu^\mu x^{2\mu-1}}{\Gamma(\mu)\Omega^\mu} \exp\left(-\frac{\mu}{\Omega}x^2\right)$  for some  $\mu \geq \frac{1}{2}$  and  $\Omega, x > 0$  (short:  $Naka(\mu, \Omega)$ ),
3. the Weibull distribution with density  $f_{Wei}(x) = \frac{\ell}{\theta} \left(\frac{x}{\theta}\right)^{\ell-1} \exp\left(-\left(\frac{x}{\theta}\right)^\ell\right)$  for some  $\ell, \theta > 0$  and  $x \geq 0$  (short:  $Wei(\ell, \theta)$ ).

Table 4.1 shows the resulting approximative distributions for  $I_2$ .

**Table 4.1** Approximative distributions for  $I_2$

$T$	distribution in $R_{\text{peak}}$	distribution in $R \setminus R_{\text{peak}}$
PVT	$\rho \cdot \log N(\nu, \sigma) + (1 - \rho) \cdot Naka(\mu, \Omega)$	$Wei(\ell, \theta)$
PDT	$\rho \cdot \log N(\nu, \sigma) + (1 - \rho) \cdot Naka(\mu, \Omega)$	$Wei(\ell, \theta)$
PLT	$\rho \cdot \log N(\nu, \sigma) + (1 - \rho) \cdot Naka(\mu, \Omega)$	$Wei(\ell, \theta)$

Note that, analogously to the  $I_1$ -case, the parameters  $\rho, \nu, \sigma, \mu, \Omega, \ell$  and  $\theta$  depend on both  $\kappa$  and the type of  $T$ .

#### 4.2 Choosing a Suitable Parametric Copula

Among the considered copula types mentioned in Section 3.3, the Gaussian copula is the one which minimises (3.3). It is given by

$$C_\Sigma^{Gauss}(u, v) = \Phi_\Sigma\left(\Phi^{-1}(u), \Phi^{-1}(v)\right),$$

where  $\Phi_\Sigma$  denotes the joint cumulative distribution function of a bivariate normal distribution with mean vector  $o$  and covariance matrix  $\Sigma$  and where  $\Phi^{-1}$  denotes the inverse cumulative distribution function of the one-dimensional standard normal distribution. The covariance matrix  $\Sigma$  with correlation coefficient  $\psi \in [-1, 1]$  has representation

$$\Sigma = \begin{pmatrix} 1 & \psi \\ \psi & 1 \end{pmatrix}.$$

This optimisation result is true for all three types of the underlying tessellation  $T$ . Furthermore, the copula type does not depend on the scaling parameter  $\kappa$ . On the other hand, the parameter  $\psi$  of the Gaussian copula depends on the type of  $T$  and  $\kappa$ , i.e.,  $\psi = \psi(T, \kappa)$ . However, it turns out that the dependence of  $\psi(T, \kappa)$  in  $\kappa$  is negligible, so that we can assume  $\psi = \psi(T)$ . Note that in our considered data, the total lengths of the two half-trees differ at least by  $10^{-6}$  length units in 99.5% of all cases. This justifies the usage of a copula which is absolutely continuous with respect to the two-dimensional Lebesgue measure.

### 4.3 Joint Distribution for Infinitely Sparse and Infinitely Dense Road Systems

In addition to the parametric approximation results obtained in Sections 4.1 and 4.2, we also provide a description of the joint distribution of  $I$  for infinitely sparse and infinitely dense networks. Limit scenarios where the underlying road system becomes infinitely sparse or infinitely dense are of special interest since the extracted data of the typical segment system is available for a large but only finite range of  $\kappa$ . For very small  $\kappa$ , fitting methods are unstable and for increasing values of  $\kappa$ , Monte Carlo simulations become very time consuming.

#### 4.3.1 Infinitely Sparse Networks

First, we consider the case of infinitely sparse networks, i.e.,  $\lambda_\ell \rightarrow \infty$  and  $\gamma = 1$  (and therefore  $\kappa \rightarrow 0$ ). In this scenario, the typical segment system  $S_H^*$  consists of the union of two segments, say  $L_1$  and  $L_2$ , with independent and exponentially distributed lengths joined at the origin. In particular, we have that  $\nu_1(L_i) \sim \text{Exp}(2\lambda_\ell)$ ,  $i \in \{1, 2\}$ . We thus obtain  $\mathbb{P}(I_1 \leq x) = \mathbb{P}(\max\{\nu_1(L_1), \nu_1(L_2)\} \leq x) = \mathbb{P}(\nu_1(L_1) \leq x, \nu_1(L_2) \leq x) = (1 - \exp(-2\lambda_\ell x))^2$  and therefore

$$f_{I_1}(x) = 4\lambda_\ell (\exp(-2\lambda_\ell x) - \exp(-4\lambda_\ell x)).$$

Similarly, we obtain  $\mathbb{P}(I_2 \leq x) = 1 - \mathbb{P}(I_2 > x) = 1 - \mathbb{P}(\min\{\nu_1(L_1), \nu_1(L_2)\} > x) = 1 - \mathbb{P}(\nu_1(L_1) > x, \nu_1(L_2) > x) = 1 - (\exp(-2\lambda_\ell x))^2 = 1 - \exp(-4\lambda_\ell x)$  and therefore

$$f_{I_2}(x) = 4\lambda_\ell \exp(-4\lambda_\ell x).$$

To summarise, the joint distribution of  $I = (I_1, I_2)$  is given by the distribution of the random vector  $(\max\{V_1, V_2\}, \min\{V_1, V_2\})$ , where  $V_1, V_2$  are independent exponentially distributed random variables with parameter  $2\lambda_\ell$ .

#### 4.3.2 Infinitely Dense Networks

Second, we provide a description of the joint distribution of  $I$  for infinitely dense networks, i.e.,  $\lambda_\ell \rightarrow 0$  and  $\gamma = 1$  (and therefore  $\kappa \rightarrow \infty$ ). In order to achieve this goal, we first note that as  $\lambda_\ell \rightarrow 0$ , the rescaled typical serving zone  $\sqrt{\lambda_\ell} \Xi_{H, \lambda_\ell}^*$  at intensity  $\lambda_\ell$  converges in distribution to a typical Poisson-Voronoi cell  $\Xi_{PV}^*$ . This can be heuristically explained by observing that, as the road system becomes infinitely dense, the rescaled Cox process of HLC converges to a homogeneous Poisson point process. Moreover, the boundary between the half-trees  $G_1^h$  and  $G_2^h$  asymptotically approaches the union of two rays corresponding to the asymptotic directions of these boundaries. Hence, the division of the typical serving zone induced by the half-trees can be recovered as a decomposition  $\Xi_{PV}^* = A_1 \cup A_2$ , where the two sets  $A_1, A_2$  are defined by two rays whose enclosed angle only depends on the type of the underlying tessellation. In particular, as  $\lambda_\ell \rightarrow 0$ , the suitably rescaled total lengths of the half-trees converge to the areas of the two parts  $A_1$  and  $A_2$ . Combining these asymptotic properties, the joint distribution of the rescaled total lengths  $(\lambda_\ell I_1, \lambda_\ell I_2)$  converges in distribution to the distribution of the random vector  $(\nu_2(A_1), \nu_2(A_2))$ , where  $\nu_2$  denotes the two-dimensional Lebesgue measure. In a forthcoming paper [5], we will provide a rigorous proof for this heuristically derived asymptotic result.

## 5 Validation of the Model

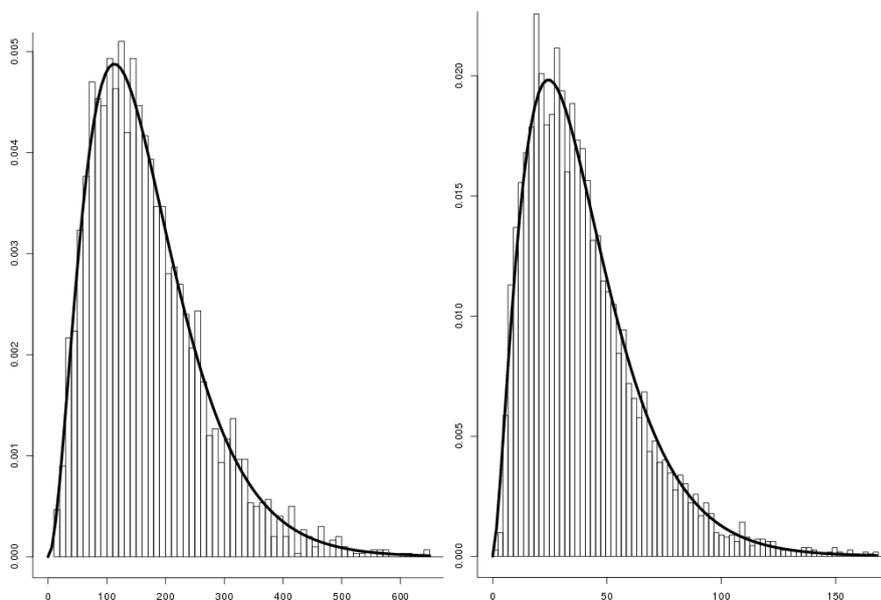
In the following, we give two possibilities to validate the model introduced in the present paper. The first one is rather heuristic while the second one relies on statistical tools. In particular, we used the following validation methods.

### 5.1 Visual Validation

A first possibility to validate our model for the functionals  $I_i$ ,  $i \in \{1, 2\}$  of  $G$  is a visual comparison between empirical data extracted from  $S_H^*$  on the one hand and directly simulated data on the other hand. We compare both the univariate datasets of the marginals and the bivariate datasets of  $I = (I_1, I_2)$ .

#### 5.1.1 Marginal Densities

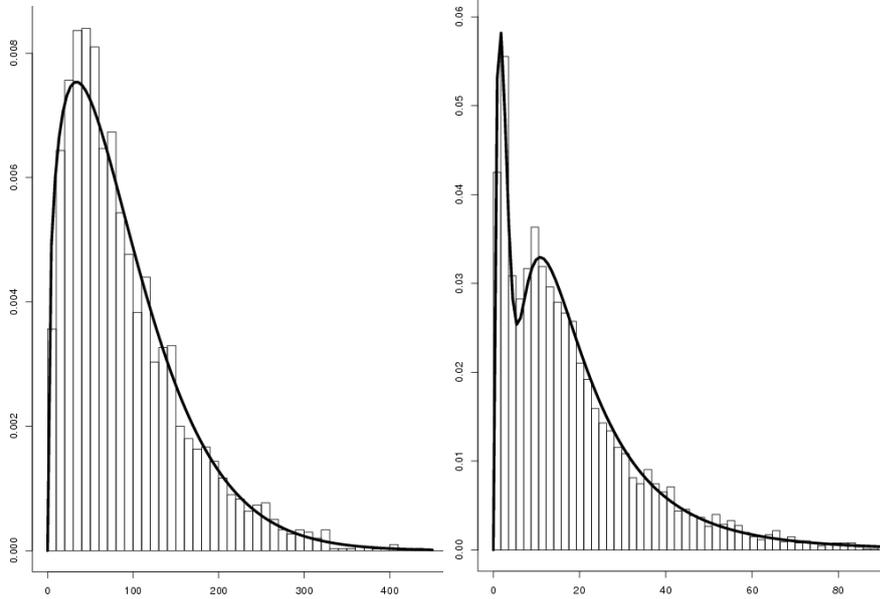
In Figure 5.1, the histograms of  $I_1$  are plotted together with the fitted density functions of the corresponding gamma distributions for two different scenarios, namely for an underlying PVT with scaling parameter  $\kappa = 250$  (left) and an underlying PDT with scaling parameter  $\kappa = 50$  (right). Both densities fit to the empirical data quite well.



**Fig. 5.1** Histograms of  $I_1$  with underlying PVT for  $\kappa = 250$  (left) and with underlying PDT for  $\kappa = 50$  (right) together with the fitted densities according to Section 4.1.1

In Figure 5.2, we present the histograms of  $I_2$  together with the fitted density functions of the corresponding Weibull distributions for the same scenarios as in

Figure 5.1. Both densities fit to the empirical data quite well, too. Using our suggested modelling approach, especially the difficulties due to the high peaks in the histograms for lower values of  $\kappa$  mentioned in Section 3 can thus obviously be handled well.

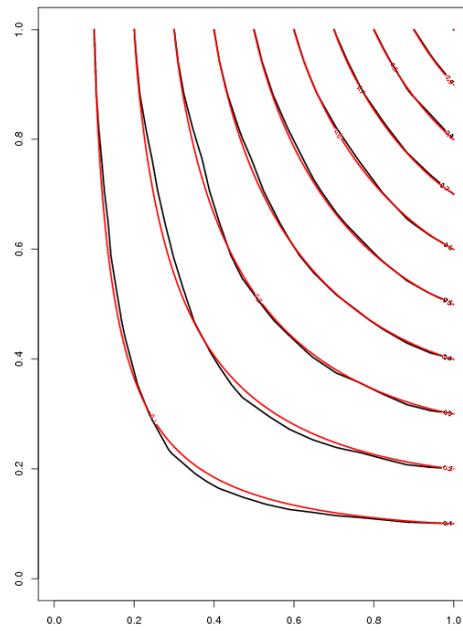


**Fig. 5.2** Histograms of  $I_2$  with underlying PVT for  $\kappa = 250$  (left) and with underlying PDT for  $\kappa = 50$  (right) together with the fitted densities according to Section 4.1.2

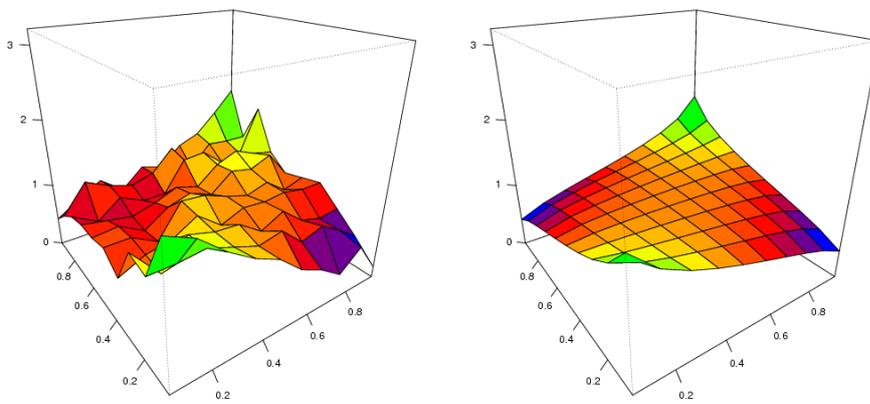
### 5.1.2 Joint Distribution

Figure 5.3 shows the contour plot of the empirical copula, coloured black, and the contour plot of the fitted Gaussian copula, coloured red. Only slight and negligible differences between the contour plots can be observed, which is a first hint that our fitted copula should be suitable.

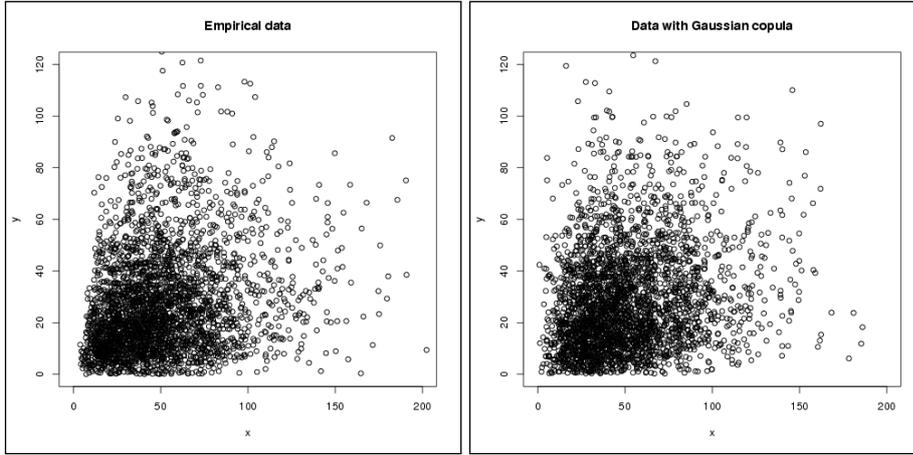
Furthermore, we take a look at the shape of the empirical density function of  $I$  and the density function of the corresponding fitted bivariate Gaussian copula. Figure 5.4 shows the corresponding plots with underlying PVT and  $\kappa = 75$ . We can observe that empirical and simulated density plots seem to be nearly equal. In order to see if the empirical data extracted from the typical segment system  $S_{\Xi}^*$  and the data directly simulated from the Gaussian copula coincide, we take a look at the corresponding scatterplots, see Figure 5.5. The final results are very similar, so that we can conclude that the model provides realistic scenarios.



**Fig. 5.3** Contour plots of the empirical copula (black) and the fitted Gaussian copula (red) for  $I$  in PVT case with  $\kappa = 75$



**Fig. 5.4** Empirical density function (left) and fitted Gaussian copula density (right) for  $I$  in PVT case with  $\kappa = 75$



**Fig. 5.5** Scatterplots of the empirical data (left) and directly simulated data (right) for  $I$  in PVT case with  $\kappa = 75$

## 5.2 Multivariate Wald-Wolfowitz Test

Besides a visual comparison, it is also desirable to validate the model with statistical tools, where we make use of a multivariate generalisation of the well-known univariate Wald-Wolfowitz two-sample test. This generalisation has been extensively discussed in [3]. For the convenience of the reader, we briefly recall the general settings of the test. Let  $X_1, \dots, X_n, Y_1, \dots, Y_m \in \mathbb{R}^2$  be independent bivariate vectors and let furthermore  $X_1, \dots, X_n \sim F_X$  and  $Y_1, \dots, Y_m \sim F_Y$  for some arbitrary distribution functions  $F_X$  and  $F_Y$ . We want to test the null-hypothesis  $H_0 : F_X = F_Y$  against the alternative  $H_1 : F_X \neq F_Y$ . In our case,  $\{X_j\}_{1 \leq j \leq n}$  corresponds to empirical data extracted from the typical segment system  $S_H^*$  whereas  $\{Y_k\}_{1 \leq k \leq m}$  corresponds to data simulated directly from the copula and the marginal distributions. In order to define a decision rule whether  $H_0$  will be rejected or not, we have to take a look at the *minimal spanning tree* (MST) based on (the Euclidean distances of) the joint sample  $X_1, \dots, X_n, Y_1, \dots, Y_m$ . Let  $S$  denote the number of disjoint subtrees resulting from erasing all edges of MST with nodes from different samples. Additionally, let  $D$  denote the number of edge pairs that share a common node in MST,  $\text{Var}(S|D)$  denote the conditional variance of  $S$  given  $D$  and let  $N = n + m$ . Then, it is shown in [3] that under  $H_0$ ,  $\mathbb{E}S = \frac{2nm}{N} + 1$ ,  $\text{Var}(S|D) = \frac{2nm}{N(N-1)} \left( \frac{2nm-N}{N} + \frac{D-N+2}{(N-2)(N-3)} (N(N-1) - 4nm + 2) \right)$  and furthermore  $\frac{S - \mathbb{E}S}{\sqrt{\text{Var}(S|D)}}$  is asymptotically  $N(0, 1)$ -distributed. The null-hypothesis  $H_0$  is therefore rejected for specific realisations  $x_1, \dots, x_n$  of  $X_1, \dots, X_n$  and  $y_1, \dots, y_m$  of  $Y_1, \dots, Y_m$  if

$$\left| \frac{S - \mathbb{E}S}{\sqrt{\text{Var}(S|D)}} \right| > z_{1-\alpha/2},$$

where  $z_{1-\alpha/2}$  denotes the  $(1 - \alpha/2)$ -quantile of the  $N(0, 1)$ -distribution. In Table 5.1, the results of the multivariate generalisation of the Wald-Wolfowitz test for

various scenarios are listed. We chose  $n = m = 1000$  and the confidence level  $\delta = 1 - \alpha = 95\%$ . For all considered examples of structural scenarios,  $H_0$  is not rejected which supports our visual validation.

**Table 5.1** Results of the Wald-Wolfowitz test

Scenario	$\left  \frac{S-ES}{\sqrt{\text{Var}(S D)}} \right $	$z_{0.975}$	$H_0$
PVT $\kappa = 40$	0.074	1.96	not rejected
PVT $\kappa = 250$	0.054	1.96	not rejected
PDT $\kappa = 20$	0.044	1.96	not rejected
PDT $\kappa = 120$	0.050	1.96	not rejected
PLT $\kappa = 75$	0.062	1.96	not rejected
PLT $\kappa = 750$	0.010	1.96	not rejected

## 6 Conclusions and Outlook

The present paper provides a substantial extension of the copula methodology which was previously proposed in [9]. We derived parametric approximation formulas for the univariate density functions of the total lengths  $I_i$  of both half-trees of the shortest-path tree. Besides, a one-parametric bivariate Gaussian copula has been used to model the correlation structure between the components of the random vector  $I = (I_1, I_2)$ . Visual validation and the multivariate Wald-Wolfowitz test showed that this approach to model the bivariate distribution of the functional  $I$  in shortest-path trees and their corresponding half-trees is suitable to directly simulate random vectors representing the data extracted from the typical segment system  $S_H^*$  in telecommunication networks.

In a forthcoming paper [5], we will provide a rigorous proof for the asymptotic distribution of  $I$  as the scaling parameter  $\kappa$  tends to  $\infty$ . Additional research topics in future could be the investigation of further, e.g. capacity-related functionals of the shortest-path tree  $G$ . Note that studying the asymmetry of the half-trees of  $G$  is an important topic from the telecommunication point of view as knowing the corresponding capacities determines the choice of size for the two biggest cables from higher to lower nodes. For instance, it answers the question whether engineers should use two cables each with capacity of 50% (since the half-trees have more or less the same capacities) or rather a partition of 70% and 30%.

## References

1. D.J. Daley and D. Vere-Jones. An Introduction to the Theory of Point Processes. Vol. I/II, Springer, New York, 2005/08.
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*, 79:939-957, 2009.
3. J. H. Friedman and L. C. Rafsky. Multivariate generalizations of the Wald-Wolfowitz and Smirnov two-sample tests. *The Annals of Statistics*, 7:967-717, 1979.
4. C. Gloaguen, F. Voss and V. Schmidt. Parametric distributions of connection lengths for the efficient analysis of fixed access network. *Annals of Telecommunications*, 66:103-118, 2011.

5. C. Hirsch, D. Neuhäuser and V. Schmidt. Asymptotic properties of Euclidean shortest-path trees and applications to telecommunication networks. Working paper (under preparation).
6. J. Illian, A. Penttinen, H. Stoyan and D. Stoyan. *Statistical Analysis and Modelling of Spatial Point Patterns*. J. Wiley & Sons, Chichester, 2008.
7. J.-F. Mai and M. Scherer. *Simulating Copulas: Stochastic Models, Sampling Algorithms, and Applications*. Imperial College Press, London, 2012.
8. R. B. Nelsen. *An Introduction to Copulas*. Springer, New York, 2006.
9. D. Neuhäuser, C. Hirsch, C. Gloaguen, and V. Schmidt. A parametric copula approach for modelling shortest-path trees in telecommunication networks. In: A. Dudin and K. Turck (eds.) *Analytical and Stochastic Modeling Techniques and Applications*. Lecture Notes in Computer Science 7984, Springer, Berlin 2013, 324-336.
10. D. Neuhäuser, C. Hirsch, C. Gloaguen and V. Schmidt. On the distribution of typical shortest-path lengths in connected random geometric graphs. *Queueing Systems*, 71:199-220, 2012.
11. D. Ruppert. *Statistics and Data Analysis for Financial Engineering*. Springer Texts in Statistics, New York, 2010.
12. R. Schneider and W. Weil. *Stochastic and Integral Geometry*. Springer, Berlin, 2008.
13. D. Stoyan, W.S. Kendall, and J. Mecke. *Stochastic Geometry and its Applications*. J. Wiley & Sons, Chichester, 1995.
14. F. Voss, C. Gloaguen, F. Fleischer and V. Schmidt. Densities of shortest path lengths in spatial stochastic networks. *Stochastic Models*, 27:141-167, 2011.
15. F. Voss, C. Gloaguen and V. Schmidt. Scaling limits for shortest path lengths along the edges of stationary tessellations. *Advances in Applied Probability*, 42:936-952, 2010.