A Parametric Copula Approach for Modelling Shortest-Path Trees in Telecommunication Networks

David Neuhäuser[1], Christian Hirsch[1], Catherine Gloaguen[2], and Volker Schmidt[1]

[1] Ulm University, Institute of Stochastics,
Helmholtzstr. 18, 89069 Ulm, Germany
{david.neuhaeuser,christian.hirsch,volker.schmidt}@uni-ulm.de
http://www.uni-ulm.de/mawi/mawi-stochastik
[2] Orange Labs,
38-40 rue du General Leclerc, 92794 Issy-Moulineaux Cedex9, France
catherine.gloaguen@orange.com

Abstract. We extend the *Stochastic Subscriber Line Model* by the introduction of shortest-path trees which are obtained by splitting up the segment system of the typical serving zone at its crossings and endings. Due to reasons in the complex field of cost and capacity estimation in telecommunication networks, it is desirable to gain knowledge about distributional properties of the branches of these trees. The present paper shows how to obtain parametric approximation formulas for the univariate density functions of the lengths of the two main branches in shortest-path trees. Besides, we derive a joint bivariate distribution for the lengths of these branches by means of copula functions, i.e., we give a parametric composition formula of the marginals. These approximative parametric representation formulas can be used in order to prevent time consuming computer experiments.

Keywords: Parametric copula, parametric marginal distribution, stochastic geometry, network planning, Palm calculus, shortest–path tree, telecommunication network

1 Introduction

In [4], [8] and [13], the *Stochastic Subscriber Line Model* (SSLM) has been developed and extended, especially in order to model access networks in urban areas. So far, the research focus has been put on so-called typical shortest-path lengths where engineers are mainly interested in minimising the total length of the telecommunication network in a city, see [4]. Besides this, also other cost functionals have to be considered for reasonable optimising. Physical links, e.g., optical fibres emanating from several nodes of lower order in the network are merged into thicker fibres at nodes of higher order. For cost estimation of the network, knowledge of capacities in the network is an important factor and so far, only preliminary results are available, see [12]. The idea in the present paper is to extract the so-called shortest-path tree from the typical segment system of a typical serving zone. Having information about this tree, research departments of telecommunication companies such as *Orange Labs* can draw conclusions about capacity problems and cost estimation of communication networks. As the geometry of

the shortest-path tree can become extremely complex, it is not clear how to describe its structure in a simple way which can also be useful for engineers. A general roadmap for attacking this problem is to build models of increasing complexity which are able to describe the tree in greater and greater details. In the present paper, we therefore provide the first step towards this goal by deriving an analytical approximation formula for the joint distribution of the two main branches of the shortest-path tree.
The paper is organised as follows. Section 2 briefly discusses the tools of stochastic geometry and their properties which are used in the SSLM. Especially Palm calculus, allowing to deal with typical cells of random tessellations as well as Cox processes are described. Besides this, we explain how to extract the shortest-path tree from the typical segment system in order to investigate capacity problems in telecommunication networks. Then, in Section 3, a simulation algorithm for the main branches of shortest-path trees is described. In particular, we provide approximation formulas for the density functions of their lengths. Besides, as the lengths of these branches cannot be assumed to be independent, a copula approach is used in order to model the correlation structure between them. Finally, Section 4 concludes the paper and gives an outlook to possible future research topics.

2 A stochastic model representing access networks

2.1 The SSLM for urban areas

The SSLM has been introduced as a spatial network model using tools of stochastic geometry in each of its three parts, see [4]. These three parts are the geometrical support, the network component part and the topological part. For the convenience of the reader, we briefly discuss the model existing so far before we proceed to expand the SSLM by shortest-path trees.

Geometrical support Cables and fibres in telecommunication networks are installed along the road system of a city or town in order to reach as many customers as possible. The whole network of main roads, side streets, dead ends, etc. is modelled in the SSLM by stationary random geometric graphs. In the present paper, these graphs are restricted to stationary random tessellations in the Euclidean space $\mathbb{R}^2$. We call a subdivision of $\mathbb{R}^2$ into a sequence $\Xi_1, \Xi_2, \ldots$ of random convex and compact polygons a planar random tessellation T if the following three conditions are fulfilled.

1. $\bigcup_{i=1}^{\infty} \Xi_i = \mathbb{R}^2$,
2. $\operatorname{int} \Xi_i \cap \operatorname{int} \Xi_j = \emptyset$ for $i \neq j$,
3. $\#\{i : \Xi_i \cap B \neq \emptyset\} < \infty$ for each bounded $B \subset \mathbb{R}^2$.

Important examples of random tessellations considered in this paper are Poisson-Voronoi tessellations, Poisson line tessellations and Poisson-Delaunay tessellations. For further information on these random graphs, the reader can consult [10] and [11].

Placement of components Network nodes such as antennas, wire centre stations, service area interfaces as well as the subscribers themselves are also assumed to be located along the road system. We restrict our considerations to two-hierarchy-level networks, i.e., we investigate path properties between higher level components (HLC) and lower level components (LLC). Therefore, we model these nodes as point processes located on the edge system of the underlying tessellation T. To be more precise, we use two stationary Cox point processes (for further information, see [11]), say $X_H = \{X_{H,n}\}$ and $X_L = \{X_{L,n}\}$. Their random intensity measures are concentrated on the edge set $T^{(1)} = \bigcup_{n=1}^{\infty} \partial \Xi_n$ of the underlying tessellation T and are proportional to the 1-dimensional Hausdorff measure ν_1 on T, i.e., $\mathbb{E}X_H(B) = \lambda_\ell \, \mathbb{E}\nu_1(B \cap T^{(1)})$ and $\mathbb{E}X_L(B) = \lambda'_\ell \, \mathbb{E}\nu_1(B \cap T^{(1)})$ for each Borel set $B \subset \mathbb{R}^2$ and linear intensities $\lambda_\ell, \lambda'_\ell > 0$. The planar intensities are thus given by $\lambda_H = \lambda_\ell \gamma$ for HLC and $\lambda_L = \lambda'_\ell \gamma$ for LLC, where $\gamma = \mathbb{E}\nu_1([0,1]^2 \cap T^{(1)})$.

Topology model and Palm calculus To define connection rules in the network, an LLC is assumed to be linked with its nearest HLC in the Euclidean sense. In other words, we consider the Voronoi tessellation $T_H = \{\Xi_{H,n}\}_{n \geq 1}$ which is generated by the Cox point process $X_H = \{X_{H,n}\}$ of high level components, i.e., the cell $\Xi_{H,n}$ around its nucleus $X_{H,n}$ is given by

$$\Xi_{H,n} = \{x \in \mathbb{R}^2 : \|x - X_{H,n}\| \leq \|x - X_{H,m}\| \text{ for all } m \neq n\},$$

where $\|\cdot\|$ denotes the Euclidean norm. Each LLC $X_{L,i}$ which is located within a so-called *serving zone* $\Xi_{H,n}$ is connected to the corresponding HLC $X_{H,n}$ along the edges of the underlying tessellation T, i.e., via the cable system along the roads of the city. The connection is arranged in a way such that the distance between LLC and corresponding HLC measured along $T^{(1)}$ is the smallest, i.e., we consider shortest paths, see [8]. In the present paper, we consider the Palm version X_H^* of X_H whose distribution can be interpreted as conditional distribution of X_H given that there is a HLC located at the origin $o = (0,0)^\top \in \mathbb{R}^2$. To be more precise, the distribution of X_H^* is given 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 $g : \mathbb{L} \to [0,\infty)$ is an arbitrary measurable function and $\mathbb{L}$ denotes the family of all locally finite sets of $\mathbb{R}^2$. Note that in particular, we have by definition that $\mathbb{P}(o \in X_H^*) = 1$. In addition, the *typical Voronoi cell* Ξ_H^* of X_H is defined as the Voronoi cell associated with the cell centre 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, 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 Cox point process of the HLC $X_{H,i}$ marked with the corresponding segment systems $S_{H,i}$, see Figure 1 for an illustration. For further details on marked point processes and Palm mark distribution, the reader is referred to [3] and [11].

2.2 Extracting shortest-path trees

In this part of the present paper, we want to extend the SSLM by a further, new component, the so-called *shortest-path tree G*. Following any leaf, say *v*, in this tree towards its root *o* corresponds to tracing the shortest path from *v* to *o* along $T^{(1)}$ in the typical serving zone Ξ_H^*. An illustration of this situation can be found in Figure 2.

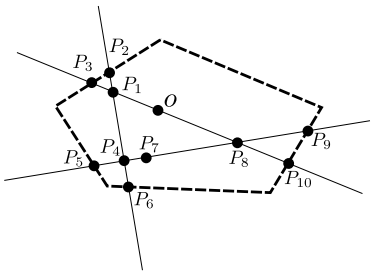

Fig. 1. Typical serving zone Ξ_H^* (dashed) and corresponding segment system S_H^* (solid)

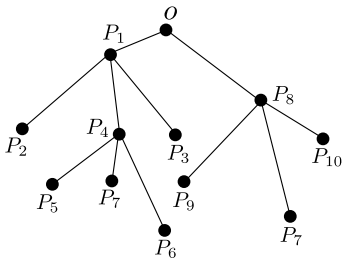

Fig. 2. Extracted shortest-path tree *G* with origin *o* as root

For cost estimation and capacity calculations in telecommunication networks, it is useful to have knowledge of this tree with the lengths of its subparts, the number of links etc. Focus is put on two main branches building the backbone and skeletal structure of *G*. The first main branch is defined as the longest branch of the whole tree. It is called the longest shortest path and denoted by *LSP* in the following. Its length, C_{LSP}, is a well-defined random variable as the supremum over all path lengths is actually a maximum due to the fact that we have a random but finite number of endpoints in the tree. Observe that with probability 1, the origin *o* has two emanating edges (see Figure 1). Thus, we can subdivide the shortest-path tree *G* into two subtrees, a half-tree G_1^h and a half-tree G_2^h as shown in Figure 3. The graph G_1^h is defined as the half-tree which contains *LSP*. The second main branch, denoted by *LSP*′, is now defined as the longest branch in the second half-tree G_2^h, see also Figure 4. Its length will be denoted by $C_{LSP'}$. Note that $C_{LSP} \geq C_{LSP'}$ holds, but *LSP*′ does not necessarily have to be the second longest branch of *G*.

3 Modelling shortest-path trees via copulas

The goal in this section is to find suitable two-dimensional (joint) distribution functions for the lengths C_{LSP} and $C_{LSP'}$ of the two main branches *LSP* and *LSP*′. In general, it cannot be assumed that these lengths are independent random variables. Indeed, both lengths certainly depend on the size of the typical serving zone in a sense that the lengths of both *LSP* and *LSP*′ are positively correlated with the size of Ξ_H^*. As a consequence, the bivariate joint distribution function cannot be written as product of the two

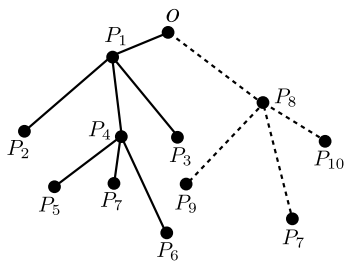

Fig. 3. Half-trees G_1^h (solid) and G_2^h (dashed) of G emanating from the root

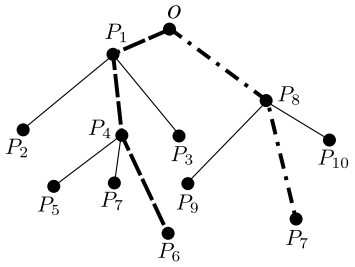

Fig. 4. Main branches *LSP* (dashed) and *LSP'* (dot-dashed) of the corresponding half-trees

univariate marginal distribution functions, i.e.,

$$F_{(C_{LSP},C_{LSP'})}(x,y) \neq F_{C_{LSP}}(x) \cdot F_{C_{LSP'}}(y).$$

A first attempt could be to consider a non-parametric approach for the fitting of a multivariate density function. Although a good fit can be achieved for a specific choice of parameters λ_ℓ and γ, it is not suitable for our goals, as we want to obtain a parsimonious parametric model which applies for any values of λ_ℓ and γ. A second attempt could be to fit a well-known parametric multivariate distribution function to the data, such as e.g. a multivariate normal or t-distribution. However, as we will see in Section 3.1, the univariate distribution of $C_{LSP'}$ can be closely approximated by a mixed gamma-distribution and we are not aware of any commonly used multivariate distribution with mixed-gamma marginals. Both types of problems mentioned above can be avoided by using a family of parametric copulas.

A copula function, roughly speaking, combines marginal distributions to a joint distribution by adding some correlation structure in a way which has to be precised. In general, a function $K : [0,1]^2 \to [0,1]$ is called a 2-dimensional copula if there exists a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ supporting a random vector $\mathbf{U} = (U_1, U_2)^\top$ such that

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

and $U_i \sim U[0,1]$ for $i \in \{1,2\}$. Note that Sklar's theorem (see [7]) guarantees the existence of a (not necessarily parametric) copula function $K_{\mathbf{C}} : [0,1]^2 \to [0,1]$ such that the bivariate joint distribution function of the random vector $\mathbf{C} = (C_{LSP}, C_{LSP'})^\top$ can be written as

$$F_{\mathbf{C}}(\mathbf{c}) = K_{\mathbf{C}}(F_{C_{LSP}}(c_1), F_{C_{LSP'}}(c_2)),$$

where $\mathbf{c} = (c_1,c_2)^\top$, $c_1,c_2 > 0$. Note that the density of $\mathbf{C}$ is given by

$$f_{\mathbf{C}}(\mathbf{c}) = f_{C_{LSP}}(c_1)\, f_{C_{LSP'}}(c_2) \cdot k_{\mathbf{C}}(F_{C_{LSP}}(c_1), F_{C_{LSP'}}(c_2)), \tag{1}$$

where $k_{\mathbf{C}}(u_1,u_2) = \frac{\partial^2}{\partial u_1 \partial u_2} K_{\mathbf{C}}(u_1,u_2)$ denotes the density function of the copula $K_{\mathbf{C}}$ and $f_{C_{LSP}}$ and $f_{C_{LSP'}}$ are the density functions of $F_{C_{LSP}}$ and $F_{C_{LSP'}}$, respectively.

A well-known tool in order to fit parametric models to data is the maximum-likelihood method. Suppose that we have parametric models $F_{C_{LSP}}(\cdot \mid \eta_1)$, $F_{C_{LSP'}}(\cdot \mid \eta_2)$ and $K_{\mathbf{C}}(\cdot \mid \eta)$ with parameter vectors η_1, η_2 and η for the marginals as well as for the copula, respectively, and assume that we have an i.i.d. sample $\mathbf{C}_i = (C_{LSP,i}, C_{LSP',i})^\top$, $i = 1, \ldots, n$ where n denotes the sample size. Considering (1), one obtains for the log-likelihood function $\log L$ the following representation

$$\begin{aligned} &\log L(\eta_1 \eta_2, \eta) \\ &= \sum_{i=1}^{n} \big(\log f_{C_{LSP}}(C_{LSP,i} \mid \eta_1) + \log f_{C_{LSP'}}(C_{LSP',i} \mid \eta_2) \\ &\quad + \log \left[k_{\mathbf{C}}(F_{C_{LSP}}(C_{LSP,i} \mid \eta_1), F_{C_{LSP'}}(C_{LSP',i} \mid \eta_2) \mid \eta) \right] \big) . \end{aligned} \tag{2}$$

A quite annoying handicap of the maximum-likelihood method is the fact that maximising the log-likelihood function is a challenging numerical problem if we have several parameters in the model. One way out of this unpleasant situation is the usage of the so-called *parametric pseudo-maximum-likelihood-method* in order to fit a suitable model, see [9]. More precisely, we follow a similar way as we do in maximum-likelihood estimation but this time, we have an optimisation process in two steps. First, the marginal distributions are estimated and represented by parametric families using the common maximum-likelihood method, each on its own. Second, we have to determine the best copula in a way which still has to be defined. Note that by estimating parameters of the marginal distributions and the copula separately, the pseudo-maximum likelihood approach avoids a higher-dimensional optimisation.
In the following, we precisely describe the two optimisation steps for the marginals and copula cases and provide the numerical results.

3.1 Fitting parametric marginal distributions

Parametric density function for C_{LSP} The distribution of C_{LSP} depends both on the linear intensity λ_ℓ of the HLC and the length intensity γ of T. For Poisson-Voronoi tessellations (PVT), Poisson line tessellations (PLT) and Poisson-Delaunay tessellations (PDT) it was shown in [13] that we have a certain scaling invariance in our model and therefore it suffices to investigate the dependence of the distribution of C_{LSP} on the ratio $\kappa = \frac{\gamma}{\lambda_\ell}$ called scaling parameter. In contrast to the situation observed for the limit cases where $\kappa \to 0$ or $\kappa \to \infty$ (see [5]), for general values of κ it is hardly possible to derive an explicit formula for the distribution of C_{LSP}. In this section, we therefore aim at finding a suitable parametric representation formula for the density function $f_{C_{LSP}}$ of C_{LSP} which approximates $f_{C_{LSP}}$ sufficiently well. For three different types of random tessellations (PVT, PDT, PLT) and for various values of the scaling parameter κ, we run a sufficiently large number n of realisations of the typical segment system S_H^* via Monte Carlo simulation and extract the shortest-path tree G. According to the obtained histograms, we fit a suitable family of parametric distributions which can be characterised by just a few parameters. This can be achieved by manual choice of an eligible class of distributions and via subsequent maximum-likelihood estimation. It turns out that a suitable class of parametric distributions for C_{LSP} is the family of scaled

Gamma distributions. To be more precise, it approximately holds that $C_{LSP} \sim \Gamma(k,\lambda)$ with density function

$$f_{\Gamma(k,\lambda)}(x;k,\lambda) = \frac{1}{\lambda^k \Gamma(k)} x^{k-1} \exp(-\frac{x}{\lambda}) \mathbb{1}_{[0,\infty)}(x), \tag{3}$$

for some shape parameter $k > 0$ and scale parameter $\lambda > 0$. This class of distributions is applicable to all three types (PVT, PDT, PLT) of underlying tessellations representing the infrastructure of the city, see Figure 5. Note that the parameters k and λ depend on the type of the underlying tessellation and on κ.

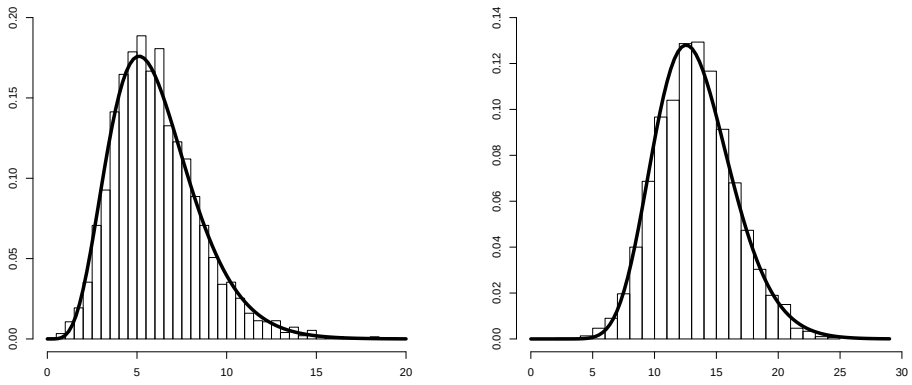

Fig. 5. Parametric densities of the scaled Gamma distribution for C_{LSP} where the underlying tessellation is a *PLT* with $\kappa = 20$ (left) and *PDT* with $\kappa = 120$ (right)

Parametric density function for $C_{LSP'}$ Next, we want to proceed in an analogous way as above and derive a parametric approximation formula for the density of $C_{LSP'}$. Looking at the histogram in Figure 6, the mindful reader may observe that, in contrast to the distribution of C_{LSP} (see Figure 5), we now have a bimodal type of distribution. In order to cope with this situation, the usage of a mixture of two scaled Gamma distributions is reasonable (note that this type of distribution indeed fulfills the necessary bimodality). To be more precise, we get that approximately $C_{LSP'} \sim \Gamma_{mix}^{\alpha,k,\lambda,\ell,\theta}$ with density function

$$\begin{aligned} & f_{\Gamma_{mix}^{\alpha,k,\lambda,\ell,\theta}}(x;\alpha,k,\lambda,\ell,\theta) \\ = & \left[\alpha \cdot \left(\frac{1}{\lambda^k \Gamma(k)} x^{k-1} \exp(-\frac{x}{\lambda})\right)\right. \\ & \left. + (1-\alpha) \cdot \left(\frac{1}{\theta^\ell \Gamma(\ell)} x^{\ell-1} \exp(-\frac{x}{\theta})\right)\right] \cdot \mathbb{1}_{[0,\infty)}(x), \end{aligned}$$

for some mixing parameter $\alpha \in [0,1]$, shape parameters $k, \ell > 0$ and scale parameters $\lambda, \theta > 0$. It turns out that this class of distributions is applicable to all three types of

underlying tessellations which have been considered so far, i.e., PVT, PDT, PLT. Note that the parameters α, k, λ, ℓ and θ depend on the type of the underlying tessellation and on κ. Formally, we can also consider the distribution of C_{LSP} as a mixture of gamma distributions (indeed, with mixing parameter $\alpha \in \{0,1\}$).

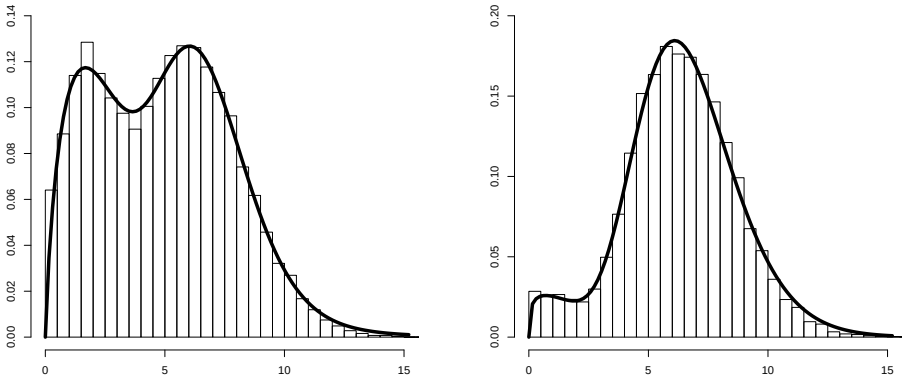

Fig. 6. Parametric densities of the mixture of two scaled Gamma distributions for $C_{LSP'}$ where the underlying tessellation is a PDT with $\kappa = 25$ (left) and PVT with $\kappa = 50$ (right)

3.2 Choosing a suitable copula type

Now having information about the (marginal) distributions of the lengths of the two main branches each considered on its own, the remaining task is to find a suitable approximative model for correlation structure between them. Similar as in the univariate case, some asymptotic results on the joint distribution of $\mathbf{C}$ can be obtained for the limit cases where $\kappa \to 0$ or $\kappa \to \infty$ (see [6]). However, for general values of κ it is hardly possible to derive an explicit formula for the joint distribution of $\mathbf{C}$. In order to achieve a good approximation formula for this distribution, we choose a copula approach, i.e., we consider several common parametric copulas and investigate which of them represents the correlation structure best, in a way which still has to be defined. Choosing a suitable type of copula is essential as the several types differ notably from each other, e.g., some have tail dependance while others have not, some allow negative correlation and others not, etc.

Preprocessing of data Before we can start to find an appropriate copula $K_{\mathbf{C}}$ adding information about the correlation of C_{LSP} and $C_{LSP'}$ when combining its marginal distributions in a joint distribution, we have to manage some difficulties resulting from special properties of our data. More precisely, our data is completely asymmetric in terms of the fact that for each shortest-path tree G, we have $C_{LSP} \geq C_{LSP'}$. This means that the scatterplot of a sample $\mathbf{C}_i$, $1 \leq i \leq n$, where n denotes the sample size, is completely located beneath the first angle bisector, see the left-hand side of Figure 7.

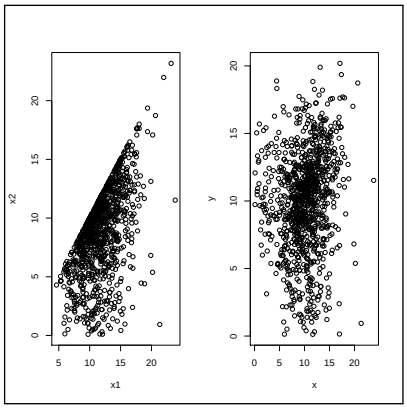

Fig. 7. Scatterplots of original and symmetrized data

Scatterplots of data which was sampled from the vast majority of commonly used bivariate copula types are however more or less some kind of symmetric, i.e., the corresponding point pairs are located beneath as well as above the first angle bisector. This is due to the fact that for many copulas $K : [0,1]^2 \to [0,1]$ it holds that $K(u_1,u_2) = K(u_2,u_1)$ for any $u_1,u_2 \in [0,1]$. In order to manage this problem, each point pair $\mathbf{C}_i = (C_{LSP,i}, C_{LSP',i})^\top$ will be associated with an independent $U[0,1]$ distributed mark U_i. If $U_i < 0.5$, we put $\tilde{\mathbf{C}}_i = (C_{LSP,i}, C_{LSP',i})^\top$, and if $U_i \geq 0.5$, then we put $\tilde{\mathbf{C}}_i = (C_{LSP',i}, C_{LSP,i})^\top$. The effect of the transformation $\mathbf{C}_i \mapsto \tilde{\mathbf{C}}_i$ can be seen on the right-hand side of Figure 7. In order to obtain the bivariate distribution of $\mathbf{C}_i$, we proceed now in the following way. First, we determine the distribution of $\tilde{\mathbf{C}}_i$ by using copulas. Then, we put

$$\mathbf{C}_i^{(1)} = \max\{\tilde{\mathbf{C}}_i^{(1)}, \tilde{\mathbf{C}}_i^{(2)}\}$$

and

$$\mathbf{C}_i^{(2)} = \min\{\tilde{\mathbf{C}}_i^{(1)}, \tilde{\mathbf{C}}_i^{(2)}\}.$$

Here, $\mathbf{C}_i^{(j)}$ and $\tilde{\mathbf{C}}_i^{(j)}$ denote the j-th component of $\mathbf{C}_i$ and $\tilde{\mathbf{C}}_i$, respectively, where $j \in \{1,2\}$.

Another possibility to bypass this problem would be to avoid the explicit ordering in $\mathbf{C}$, so that both branch lengths are identically distributed. Note that this would lead to a representation of the marginals as a mixture of C_{LSP} and $C_{LSP'}$. However, for certain values of κ, the branch lengths C_{LSP} and $C_{LSP'}$ can be dramatically different which makes a separate consideration of C_{LSP} and $C_{LSP'}$ necessary for capacity analysis of real-world telecommunication networks.

Note that the endpoints of LSP and LSP' are either located on the boundary of the typical cell Ξ_H^* or they form a so-called distance-peak, i.e., there is a point in the interior of the typical segment system S_H^* for which the shortest path to the origin is not unique (see [8] for further explanations). The point P_7 in Figure 1 illustrates that distance-peaks can occur with positive probability. Note however that in this example the point P_6 (and not the distance-peak P_7) is the endpoint of LSP, see also Figure 4. Another problem occurs

if LSP and LSP' share the same distance-peak as endpoint (i.e., $C_{LSP,i} = C_{LSP',i}$). Then the distribution of $\mathbf{C}$ cannot be modelled by commonly used parametric copulas (since these are absolutely continuous with respect to 2-dimensional Lebesgue measure). In order to prevent this unsatisfying situation, we (temporarily) remove distance-peaks and treat them separately.

Gumbel copula Recall that we already determined parameters for the marginal distributions by using a pseudo-maximum-likelihood approach, see Section 3.1. Therefore, only the parameters of the copula model remain to be estimated. Copula types which are considered in the present paper are the Archimedean copulas of Clayton, Gumbel, and Frank type as well as the Gaussian and the t-copula. In order to find a suitable copula type and to prevent overfitting due to too many parameters, a reasonable decision tool in this context is Akaike's information criterion (AIC) which is defined as

$$AIC = 2\left(p - \log L(\hat{\eta})\right),$$

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

$$\log L(\eta) = \sum_{i=1}^{n} \log\left[k_{\mathbf{C}}\left(\hat{F}_{C_{LSP}}(C_{LSP,i}), \hat{F}_{C_{LSP'}}(C_{LSP',i}) \mid \eta\right)\right]. \tag{4}$$

Note that (4) is obtained from (2) by replacing the marginal distribution functions by their estimators (which are in fact nothing else than the empirical distribution functions $\hat{F}_{C_{LSP}}$ and $\hat{F}_{C_{LSP'}}$) and omitting summands which do not depend on the parameter vector η of the copula. The type of copula which has the smallest AIC value is now chosen to work with. This seems quite reasonable as maximising $\log L(\eta)$ and minimising the number of parameters p is the goal we aim at. It turned out that for all three underlying tessellation models (PVT,PLT,PDT) we can use the same family of copulas since the Gumbel copula minimises AIC among the five considered copula models for PVT, PDT and PLT. To be more precise, we use the copula

$$K_{\mathbf{C}}(u,v) = \exp\left(-\left((-\log u)^{\eta} + (-\log v)^{\eta}\right)^{1/\eta}\right),$$

where we obtain $\eta \approx 1.21$ for each κ and for each type of the considered tessellations. This means that neither the type of the copula nor the value of its parameter η depend on the type of the underlying tessellation and the scaling parameter κ which is a quite surprising result.

Distance peaks In this section, we describe how to handle the distance peaks mentioned before, i.e., we suggest a model for those point pairs $\mathbf{C}_i = (C_{LSP,i}, C_{LSP',i})^{\top}$, where $C_{LSP,i} = C_{LSP',i}$. Due to the fact that those points are located along the first angle bisector $f(x) = x$, it suffices to consider the (conditional) univariate distribution of C_{LSP} given that LSP and LSP' have the same length. It turns out that we can model the distance peaks $\mathbf{C}_i = (C_{LSP,i}, C_{LSP',i})^{\top}$ by again using a scaled gamma distribution.

Note that the parameters are different from those obtained in Section 3.1, but nevertheless also depend on the underlying tessellation T as well as the scaling parameter κ. In particular, we sample from a density function with representation formula

$$f(x;k,\lambda) = \frac{1}{\lambda^k \Gamma(k)} x^{k-1} \exp(-\frac{x}{\lambda}) \mathbb{1}_{[0,\infty)}(x).$$

Putting a realisation x of a random variable $X \sim \Gamma(k,\lambda)$ into both components of $\mathbf{C}_i$, i.e., $\mathbf{C}_i = (X,X)^\top$, yields the desired quantity. The probability ρ of distance-peaks can be easily estimated by the ratio

$$\widehat{\rho} = \frac{\#\text{ distance-peaks in empirical data}}{n},$$

where n denotes the sample size.

Combining simulated data Finally, we join data sampled from the Gumbel copula and from the scaled Gamma distribution of the distance peaks. To be more precise, for some large integer $N \geq 1$, we generate $\lfloor (1-\widehat{\rho})N \rfloor$ points from the copula model and $\lfloor \widehat{\rho} N \rfloor$ points from the distance-peak model and consider the union of these data. In Figures 8, 9 and 10, the reader can compare symmetrised empirical data which has already been extracted from the shortest-path tree G, to data which has been directly sampled by the copula approach presented above. For each type of tessellation and a wide range of κ, we obtain quite good results. For example, see Figure 8 for $\kappa = 375$ in the PVT case, Figure 9 for $\kappa = 20$ in the PDT case and Figure 10 for $\kappa = 120$ in the PLT case.

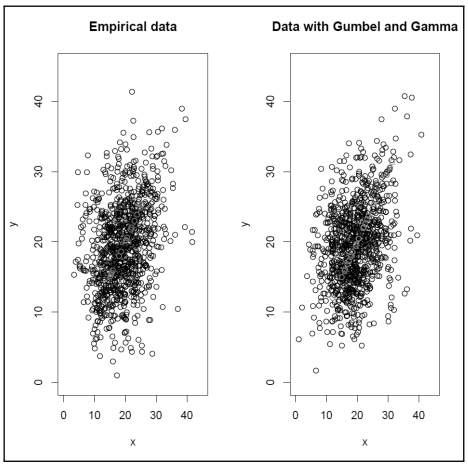

Fig. 8. Scatterplots of empirical data (left) and directly simulated data (right) where $\kappa = 375$ and T is a PVT

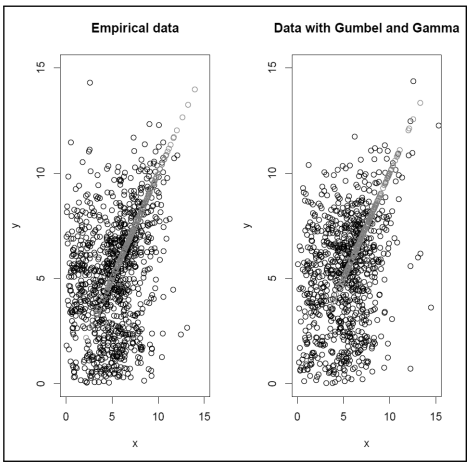

Fig. 9. Scatterplots of empirical data (left) and directly simulated data (right) where $\kappa = 20$ and T is a PDT

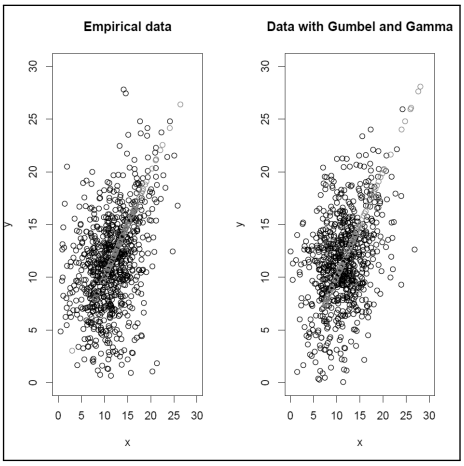

Fig. 10. Scatterplots of empirical data (left) and directly simulated data (right) where $\kappa = 120$ and T is a PLT

Note that with increasing values of κ, the number of distance peaks (grey points) compared to non-distance-peaks (black points) decreases for each of the three types of underlying tessellations.

4 Conclusions and Outlook

We extended the SSLM by a further component, the shortest-path tree G which can be extracted out of the typical segment system S_H^*. In order to simulate structural characteristics of G directly instead of simulating Ξ_H^* respectively S_H^* and extracting the tree, we developed a method how to obtain the skeletal backbone – the main branches – of G. We used a pseudo-maximum-likelihood approach to achieve this. Parametric approximation formulas for the (marginal) density functions of the longest shortest-path-lengths C_{LSP} and $C_{LSP'}$ were derived as well as a parametric copula type representing the (joint) distribution of $\mathbf{C} = (C_{LSP}, C_{LSP'})$. In our future research, we will investigate further structural characteristics of the shortest-path tree G using the copula approach developed in the present paper. Moreover, other possible topics for future research could be the investigation of shortest-path trees with other types of underlying tessellations, e.g., iterated tessellations, or even completely different types of stationary random geometric graphs as β-skeletons (see [1]). Besides, further types of copulas used for the joint bivariate distribution for $\mathbf{C}$ can be investigated in order to examine if they yield even better results for directly simulated characteristics of shortest-path trees compared to those ones obtained in the present paper.

Acknowledgement

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.

References

1. P. Bose, L. Devroye, W. Evans, and D. Kirkpatrick. *On the spanning ratio of Gabriel graphs and β-skeletons.* In: Proceedings of the 5th Latin American Symposium on Theoretical Informatics (LATIN'02), Lecture Notes in Computer Science 2286:479-493, Springer, Berlin, 2002.
2. P. Calka. The distributions of the smallest disks containing the Poisson-Voronoi typical cell and the Crofton cell in the plane. *Advances in Applied Probability*, 34:702-717, 2002.
3. D.J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes*. Vol. I/II, Springer, New York, 2005/08.
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, C. Gloaguen and V. Schmidt. First-passage percolation on random geometric graphs and its applications to shortest-path trees. *Advances in Applied Probability* (submitted).
6. C. Hirsch, D. Neuhäuser, C. Gloaguen and V. Schmidt. Asymptotic properties of Euclidean shortest-path trees and applications in telecommunication networks. Working paper (under preparation).
7. J.-F. Mai and M. Scherer. *Simulating Copulas: Stochastic Models, Sampling Algorithms, and Applications.* Imperial College Press, London, 2012.
8. 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.
9. D. Ruppert. *Statistics and Data Analysis for Financial Engineering.* Springer Texts in Statistics, New York, 2010.
10. R. Schneider and W. Weil. *Stochastic and Integral Geometry*. Springer, Berlin, 2008.
11. D. Stoyan, W.S. Kendall, and J. Mecke. *Stochastic Geometry and its Applications.* J. Wiley & Sons, Chichester, 1995.
12. F. Voss, C. Gloaguen and V. Schmidt, Capacity distributions in spatial stochastic models for telecommunication networks. *Image Analysis and Stereology*, 28:155-163, 2009.
13. 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.