Thinning-Stable Point Processes as a Model for Spatial Burstiness

Sergei Zuyev
Chalmers University of Technology and University of Gothenburg,
Department of Mathematical Sciences,
412 96 Gothenburg, Sweden.
Email: sergei.zuyev@chalmers.se

***Abstract*—In modern telecommunications, spatial burstiness of data traffic poses challenges to traditional Poisson-based models. This paper describes application of thinning-stable point processes, which provide a more appropriate framework for modeling bursty spatial data. We discuss their properties, representation, inference methods, and applications, demonstrating the advantages over classical approaches.**

***Index Terms*—wireless systems, modelling, inference, point process, spatial burstiness, discrete stability, cluster process, random measure**

I. Introduction

The classical Poisson point process has been a foundational model in queueing theory for telecommunications since the early 20th century [6], [8]. It assumes that events (such as call arrivals) occur independently, with a constant or slowly varying mean rate, see, e.g., [7]. However, real-world data, particularly in mobile communications, often exhibit burstiness—periods of high activity interspersed with inactivity. The activity rate can change rapidly by orders of magnitude; for instance, the size of a typical SMS message or email is several orders of magnitude smaller than that of a streamed video.

Traditional Poisson models struggle to capture this burstiness, leading to inefficient network resource allocation and inaccurate performance predictions. While heavy-tailed queueing models and fractional Brownian motion have been employed to account for temporal burstiness, they are poorly suited for modeling spatial burstiness, i.e., high spatial inhomogeneity.

Since the performance characteristics of mobile and, in particular, wireless systems are strongly influenced by their spatial layout and topology, point processes and stochastic geometry models have gained popularity over the last few decades, see, e.g., [1], [10], [2], [4]. However, spatial burstiness has received little attention despite its critical importance for network performance analysis, dimensioning, and strategic planning. The following example illustrates this issue.

Figure 1 presents data on calling activity collected during the 2008 Fête de la Musique festival in Paris.

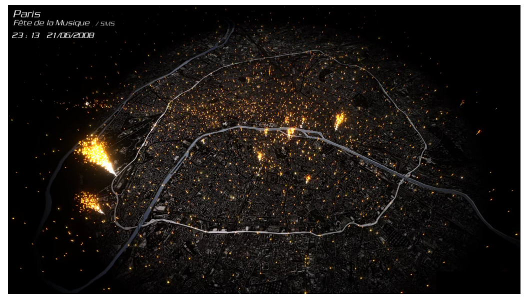


Fig. 1. Wireless communication activity during *Fête de la Musique* festival in Paris in 2008, courtesy of Orange.fr. The height of the torches represents the number of calls in progress at their locations.

The height of the torches represents the number of people making calls in different areas of the city. The data reveal bursty activity, particularly in locations where large crowds have gathered, such as stadiums and public squares. When these torches are projected onto the city's ground plane, it becomes evident that accurately modeling this distribution requires a spatial model capable of capturing highly irregular and dynamic spatial patterns.

This paper develops inference methods for thinning-stable point processes (also known as discrete-stable processes, or $\mathrm{D}\alpha\mathrm{S}$ processes), a generalisation of the Poisson model that better accommodates bursty spatial patterns.

The structure of the paper is as follows: Section 2 introduces the concepts of stability and discrete stability in probability theory and defines thinning-stable ($\mathrm{T}\alpha\mathrm{S}$) processes. Section 3 discusses statistical inference methods. Section 4 presents simulations and applications. Section 5 provides concluding remarks.

II. Mathematical Background

A. Poisson Limit Theorem

A Poisson point process is the basic model for point events happening in time or space. It plays the same role as the Gaussian processes for continuous data because of the following result mimicking the Central Limit Theorem. Let (Ψ_i) be a sequence of i.i.d. point processes in $\mathbb{R}^d$ with a locally finite intensity measure $\mathbf{E}\,\Psi_i(B) = \mu(B)$, i.e. taking finite values for all bounded Borel sets B. Writing $p \circ \Psi$ for the result of *independent thinning* of the process Ψ points with retention probability p, and $+$ for superposition of the processes, the following *Poisson limit theorem* holds:

$$\frac{1}{n} \circ (\Psi_1 + \cdots + \Psi_n) \Rightarrow \Pi_\mu, \tag{1}$$

ISBN 978-3-903176-73-7

where Π_μ is a Poisson process with intensity measure μ, and $\Rightarrow$ denotes the weak convergence.

This result explains why Poisson processes naturally emerge as limits of superpositions of many 'thin' sources. The thinning operation is thus a discrete analogue of the scaling for random variables making the processes 'thin'. However, the assumption of locally finite intensity measure μ is inadequate for bursty data, where variations can span multiple orders of magnitude. For these processes the expectation $\mathbf{E}\,\Psi_i(B)$ is infinite even for a bounded B so the intensity measure does not exist. For such processes, the limit in the thinning-superposition scheme is no longer Poisson. Namely, if for some $0 < \alpha < 1$ one has a non-trivial limiting process Φ such that

$$\frac{1}{n^\alpha} \circ (\Psi_1 + \cdots + \Psi_n) \Rightarrow \Phi, \tag{2}$$

then the process Φ is necessarily *thinning α-stable* or TαS, i.e. it satisfies the following distributional identity: for any $p \in (0,1)$,

$$p^{1/\alpha} \circ \Phi' + (1-p)^{1/\alpha} \circ \Phi'' \stackrel{\mathcal{D}}{=} \Phi, \tag{3}$$

where Φ', Φ'' are independent realisations of Φ and $\stackrel{\mathcal{D}}{=}$ denotes equality in distribution. Thinning-stable point processes were introduced and studied in details in [5] where we call them Discrete α-Stable (DαS). Here we prefer TαS-notation as it reflects better the used scaling operation $\circ$: a more general *continuous branching* operation can also lead to a non-trivial limit in (2), see [9].

Observe a direct parallel to the classical notion of stability. Recall that a random variable ξ is strictly α-stable (StαS) if for any $t \in (0,1)$,

$$t^{1/\alpha}\xi' + (1-t)^{1/\alpha}\xi'' \stackrel{\mathcal{D}}{=} \xi, \tag{4}$$

where ξ', ξ'' are independent copies of ξ. Stability plays a fundamental role in limit theorems, as only stable distributions can appear as weak limits:

$$n^{-1/\alpha}(\zeta_1 + \cdots + \zeta_n) \Rightarrow \xi. \tag{5}$$

Recall that a *Cox process* directed by a random measure ζ is a double-stochastic Poisson process which may be thought of as been generated by the following procedure: first, a realisation of $\zeta(\omega)$ is obtained, second, a realisation of a Poisson process with $\zeta(\omega)$ as the intensity measure is produced. By [5, Th. 15], TαS processes are exactly Cox processes directed by a random measure ζ which is StαS, i.e. for every set B, the random variable $\zeta(B)$ is StαS. Note that Poisson process is thinning-stable with exponent $\alpha = 1$. Thus, for $0 < \alpha < 1$, the random measure ζ plays the same role as a (non-random) intensity measure of a Poisson process when $\alpha = 1$. An explicit description of TαS process distribution is available in the form of the probability generating functional, see [5, Eqns. (34) and (40)].

The most practically important class of TαS processes comprise the *regular*[1] processes which can be represented as a Poisson cluster processes. This representation allows direct modeling of burstiness through the underlying Poisson structure. For this, state a few definitions.

A random variable ν taking values in $\mathbb{N}$ has Sibuya distribution $\mathsf{Sib}(\alpha)$ with parameter $\alpha \in (0,1]$ if

$$\mathbf{P}\{\nu = n\} = \prod_{k=1}^{n-1}\Big(1 - \frac{\alpha}{k}\Big)\frac{\alpha}{n},\ n = 1,2,\ldots$$

It corresponds to the number of Bernoulli trials until the first success if the probability of success in the k-th trial is α/k.

A finite point process Υ on a measurable phase space X is called *Sibuya point process* with parameter $\alpha \in (0,1]$ and a probability measure μ on X (notation: $\Upsilon \sim \mathsf{Sib}(\alpha,\mu)$) if the total number $\nu = \Upsilon(X)$ of its points is Sibuya $\mathsf{Sib}(\alpha)$-distributed and, given ν, these points are independently scattered according to the measure μ.

The main representation theorem for a regular TαS process Φ states that it is characterised by two parameters: $\alpha \in (0,1]$ and a *spectral measure* σ which is a measure on the space $\mathbb{M}_1$ of the probability distributions on X. It has the following *cluster representation*: first, a Poisson process $\widetilde{\Pi}_\sigma$ with intensity measure σ on $\mathbb{M}_1$ is generated (the *germs* process or cluster *centres*). And then superposition of independent Sibuya processes (the *daughter* processes or *clusters*) is taken:

$$\Phi \stackrel{\mathcal{D}}{=} \sum_{\mu_i \in \widetilde{\Pi}_\sigma} \Upsilon_i, \tag{6}$$

where $\Upsilon_i \sim \mathsf{Sib}(\alpha,\mu_i)$, see [5, Th. 24].

The simplest, yet flexible and rich family of regular *stationary ergodic* TαS processes on $X = \mathbb{R}^d$ is obtained when the spectral measure σ is concentrated on the set $\mathbb{M}_{\mu_0} = \{\mu_x = \mu_0(\cdot - x) :\ x \in \mathbb{R}^d\}$ of shifts of a fixed measure $\mu_0 \in \mathbb{M}_1$ and invariant with respect to these shifts. The image of such σ with respect to mapping $\mu_x \mapsto x$ from $\mathbb{M}_{\mu_0}$ to $\mathbb{R}^d$ is thus a multiple of the Lebesgue measure λdx for some $\lambda > 0$. Such processes represent a superposition of independent realisations Υ_i of $\mathsf{Sib}(\alpha,\mu_0)$-processes shifted to the points of a homogeneous Poisson process Π_λ with density λ in $\mathbb{R}^d$:

$$\Phi \stackrel{\mathcal{D}}{=} \sum_{x_i \in \Pi_\lambda} (\Upsilon_i + x_i), \tag{7}$$

where

$$\Upsilon_i + x_i = \sum_{y_j \in \Upsilon_i} \delta_{y_j + x_i} \sim \mathsf{Sib}(\alpha, \mu_0(\cdot - x_i)).$$

This presentation serves a basis for their easy simulations. Figure 2 shows a realisation of such a process when μ_0 is a 2D symmetrical Gaussian distribution. Different clusters are coloured differently for better visibility, but in real data points may or may not have its cluster identity tag.

[1] TαS processes are infinitely divisible, so *regular* here means the corresponding KLM (or Lévy) measure is supported by the set of finite measures, see [5, Sec. 5] for details.

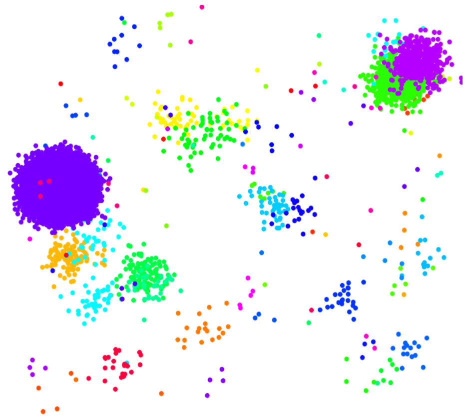


Fig. 2. $\lambda = 0.4$, $\alpha = 0.6$ and $\mu_0 \sim \mathcal{MVN}(0, 0.5^2\mathrm{I})$

III. Statistical Inference for TαS Processes

For inference, we assume that we observe a stationary ergodic TαS process Φ given by (7) inside a bounded set W (the observation window).

Given an observed realisation of Φ, statistical estimation aims to recover:

- λ: the intensity of the cluster centers,
- α: the stability parameter,
- μ_0: the Sibuya parameter measure, i.e. the distribution of a cluster's points.

Since Φ is a cluster process, we rely on well developed methods to separate clusters for estimation of μ_0 in Sec. III-A. In Sec. III-B, we propose novel estimation methods for estimation of α and λ which are based on specific stability property of TαS processes with respect to thinning.

A. Estimation of the Sibuya parameter measure

The crucial step in statistical inference is estimation of the measure μ_0 according to which all the clusters are distributed.

One could have three possible situations:

- μ_0 is already known or assumed;
- μ_0 is known to belong to a parametric class (for instance, a Gaussian distribution with unknown covariance matrix or a uniform distribution in a ball of unknown radius);
- μ_0 is totally unknown.

The bursty nature of TαS processes provides a very intuitive way of estimating μ_0: if one identifies a big cluster and isolates it from the rest of the data, it would give an accurate estimation for the measure μ_0 because it contains a large number of independent points drawn from a shifted version of μ_0. The center of mass could be taken as an estimate of this shift.

Once a cluster C of a large size $|C|$ is identified, an obvious estimator of μ_0 is the empirical distribution

$$\widehat{\mu}_0 = \frac{1}{|C|} \sum_{y_i \in C} \delta_{y_i - b}, \tag{8}$$

where $b = \sum_{y_i \in C} y_i$ is the center of mass. This empirical measure provides a non-parametric approximation of the spatial distribution of points in a cluster and serves a basis for subsequent smoothing, for example, with a kernel density estimation or for estimation of the parameters via the likelihood maximisation. It is effective when well-separated clusters can be identified. Then also a few large clusters could be pooled together after the shift compensation by their centers of mass b_k providing yet a larger dataset to estimate μ_0. But detecting cluster may be hard and unreliable when clusters overlap significantly and the cluster affiliations are not known.

In these cases, a mixture model provides a better framework for estimation and it is theoretically supported by the cluster structure of a TαS process. Assuming that $\mu_0(dy)$ has a density, the data (y_i) are interpreted as a realisation from a mixture of cluster-generating processes with the likelihood

$$\prod_k \sum_i p_k f(y_i | \theta_k), \tag{9}$$

where p_k, $\sum_k p_k = 1$, are the probabilities that an observation point belong to the k-th cluster (the mixture weights) and $f(y \mid \theta_k)$ is the probability density of the k-th Sibuya cluster process parametrised by θ_k.

An Expectation-Maximisation (EM) algorithm can then be applied to iteratively estimate the parameters:

- E-step: Compute the expected cluster assignments based on current parameter estimates.
- M-step: Maximize the likelihood function with respect to θ_k.

This method improves estimation accuracy when the true clusters are not directly observable, see, e.g., [3].

B. Estimation of the Stability Parameter and the Cluster Density

Once an estimate of μ_0 is available, we aim to estimate the other two parameters: the density λ of the clusters' centres and the stability parameter α. Because of availability of explicit formulas, we tried two methods. One is based on the probability generating functional (p.g.fl.) which, for a function $u(x)$ such that $0 \le 1-u < 1$ has a bounded support, is given by

$$\log G_\Phi[u] = -\int_{\mathbb{M}_1} \langle 1-u, \mu\rangle^\alpha \sigma(d\mu), \tag{10}$$

see [5, Cor. 16]. Here, $\langle f, \mu\rangle$ stands for $\int_X f(x)\,\mu(dx)$, for short. For the model we consider, this translates into the log-p.g.f. $\log \mathbf{E}\, z^{\Phi(B)}$ of the number of points in a set B which equals:

$$-\lambda(1-z)^\alpha \int_{\mathbb{R}^d} \mu_0(B-x)^\alpha dx. \tag{11}$$

Fitting empirical p.g.f. estimates λ and α. A boundary correction is also possible to implement when the integration domain $\mathbb{R}^d$ is replaced by the observation window W.

When there is no multiple points in the data, the process may be assumed simple (without multiple points). This is the case when μ_0 has a density. Since void probabilities: $\mathbf{P}\{\Phi(B) = 0\}$ for bounded B, determine the distribution of a simple point process Φ, an alternative method can be used: the parameters are estimated via fitting the void probabilities for a system of balls of different radii and positions. From (11), the void probability $\mathbf{P}\{\Phi(B) = 0\}$ equals

$$\exp\Big\{ -\lambda \int_{\mathbb{R}^d} \mu_0(B - x)^\alpha dx \Big\}, \tag{12}$$

and the method is equivalent to fitting the contact distribution tail function $G(r) = \mathbf{P}\{\Phi(B_r(0)) = 0\}$, where $B_r(0)$ is the ball of radius r at the origin. Let $\{x_i\}_{i=1}^n \subset W$ be a sequence of *test points* and r_i be the distances from x_i to the closest point of Φ. It is straightforward to show that

$$\widehat{G}(r) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{\{r_i > r\}} \tag{13}$$

is an unbiased estimator for $G(r)$. Then α and λ are estimated by the best fit to this curve: if $r_{(i)}, 1 = 1, \dots, n$, denotes the increasingly ordered sequence of $r_i' s$,

$$\begin{aligned}(\widehat{\alpha}, \widehat{\lambda}) &= \operatorname*{arg\,min}_{\alpha \in (0,1),\, \lambda \geq 0} \sum_{i=1}^{n} (\widehat{G}(r_i) - G(r_i))^2 \\ &= \operatorname*{arg\,min}_{\alpha \in (0,1),\, \lambda \geq 0} \sum_{i=1}^{n} ((n-i)/n - G(r_{(i)}))^2.\end{aligned}$$

It can also be shown that the variance of the estimator is smaller when the test points are taken on a grid rather than randomly scattered. We also observed poor performance of both estimation methods in the presence of large clusters, so for highly bursty data. It is explained by the fact that when a test point x_i falls inside a large cluster, the distance to the closest point may be by orders of magnitude smaller than for the test points in relatively void areas. So a novel estimation method is proposed that makes use of a particular structure of $\mathrm{T}\alpha\mathrm{S}$ processes.

The method is based on the observation that a thinned $\mathrm{T}\alpha\mathrm{S}$ process is again $\mathrm{T}\alpha\mathrm{S}$: it is easy to see from (10) and that $G_{p\circ\Phi}[u] = G_\Phi[1 - p(1-u)]$ for independent thinning. The parameters μ_0 and α stay the same for $p \circ \Phi$, but the clusters' density becomes λp^α. Therefore,

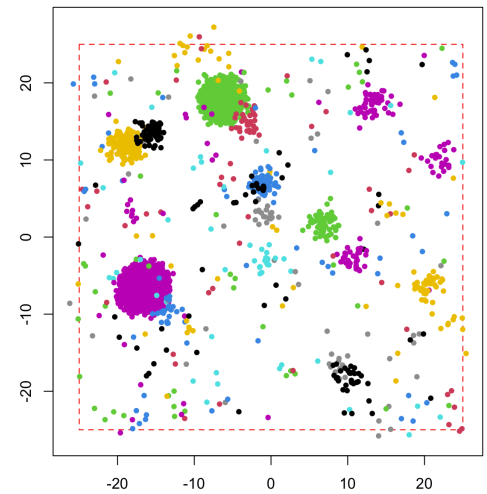
$$\begin{aligned}&\mathbf{P}\{(p \circ \Phi)(B_r(0)) = 0\} \\ &= \sum_{k=1}^{\Phi(W)} p(1-p)^{k-1} \mathbf{P}\{r_k > r\}.\end{aligned}$$

Here $p(1-p)^{k-1}$ is the probability that the closest point of Φ which survived the p-thinning is the k-th. This results in that

$$\frac{1}{n} \sum_{i=1}^{n} \sum_{k=0} p(1-p)^{k-1}\, \mathbb{I}_{\{r_{i,k} > r\}} \tag{14}$$

is also an unbiased estimator for $G(r)$. In practice we regress the values of λ and α from the curves obtained for a range of thinning probability p.

The right plot on Figure 3 exemplifies the relative error of the estimation of $G(r)$ for various values of the thinning probability p for the model with Gaussian clusters illustrated on the left plot. Value $p = 1$ corresponds to the formula (13). For small values of p, the error starts to grow because of



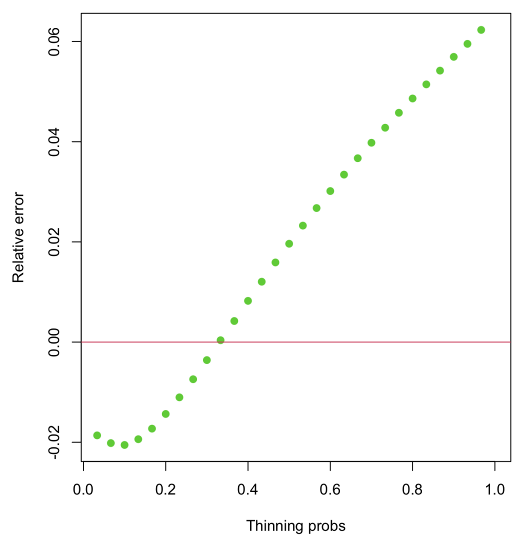


Fig. 3. Realisation of a Gaussian cluster model with $\mu_0 \sim \mathrm{MVN}(0, \mathrm{I})$, $\lambda = 0.1, \alpha = 0.7$ and the relative error of estimation of $G(1)$ for various values of p in (14).

the gradual loss of information: a boundary correction starts to prevail because of small number of points survived such a heavy thinning. We therefore recommend a moderate range of p-values in (14) which proves a sufficient number of survived points for the analysis.

C. Estimation of the stability parameter from cluster sizes

If the clusters are well separated or if they possess an identifying tag, the stability parameter α may be estimated from their sizes as these are independent $\mathsf{Sib}(\alpha)$-distributed random variables. The corresponding p.g.f. is given by $g(t) = 1-(1-t)^\alpha$, and, given the sizes $n_1, \dots, n_K$ of K observed clusters we may use an empirical p.g.f. $g_K(t) = K^{-1} \sum_{k=1}^{K} t^{n_k}$ to estimate

$$\alpha = \frac{\log(1-g(t))}{\log(1-t)}.$$

Fix $0 \le t_1 < \dots t_q < 1$. Then

$$\alpha_K = \sum_{j=1}^{q} \log(1-g_n(t_j))b_j, \text{ where}$$
$$b_j = \frac{l_j - \bar{l}}{\sum_{j=1}^{q}(l_j-\bar{l})^2},$$
$$l_j = \log(1-t_j) \text{ and } \bar{l} = q^{-1}\sum_{j=1}^{q} l_j$$

is a consistent estimator of α. Indeed,

$$\alpha_K = \sum_{j=1}^{q} \frac{\log(1-g_n(t_j))}{l_j} l_j b_j$$

converges a.s. as $K \to \infty$ to

$$\sum_{j=1}^{q} \frac{\log(1-g(t_j))}{l_j} l_j b_j = \alpha \sum_{j=1}^{q} l_j b_j = \alpha.$$

Other parameter estimation methods for heavy-tailed distributions can be employed too.

IV. Applications and Simulations

We implemented simulation studies to analyse estimation biases under various conditions:

- Large vs. moderate datasets;
- Overlapping vs. separated clusters;
- Heavy burstiness (small α, large clusters) vs. moderate burstiness.

For the estimation of the distribution μ_0, we observed that treating some proportion of small clusters as a Poisson noise may benefit computation time by significantly limiting the possible number of clusters and hence, possible associations.

To check our estimation methods described in Section III-B, we simulated one-dimensional TαS process with μ_0 being Uniform in [-1,1] distribution in the window $W = [-500, 500]$ for various values of parameters ranging from $\alpha = 0.6$ (heavy clusters) to $\alpha = 0.8$ (light clusters) and from $\lambda = 0.02$ (separated clusters) to $\lambda = 0.4$ (overlapping clusters). We obtained the accuracy of the estimates are shown in Table I with intermediate values in between these extremes. Our simulations show that parameters α and λ are best estimated by void probabilities with thinning method which also produces best estimates in all the situations apart from moderate separated clusters, however, at the expense of being more computationally expensive. The simplest void probabilities method is well suited for moderate datasets with separated clusters. It estimates well the burstiness parameter α, but in the latter case λ is best estimated by counts' p.g.f. fitting. We also tried fitting $\log G(r)$, because it depends linearly on the parameter λ as seen in (12). Then the least squares approximation error has explicit solution for the root of the partical derivative with respect to λ thus leading to numeric optimisation over just one parameter α.

		Cluster density λ	
		0.02	0.4
Stability parameter α	0.6	(0.6045, 0.0208)	(0.6019, 0.3978)
	0.8	(0.8168, 0.0215)	(0.8093, 0.4007)

TABLE I
ESTIMATES $(\widehat{\alpha}, \widehat{\lambda})$ FOR THE PARAMETERS α LISTED AS THE ROW LABELS AND λ AS COLUMN LABELS FROM 50 REALISATION OF TαS WITH UNIFORM μ_0 IN 1D.

As common in modern Statistics, all methods should be tried and consistency in estimated values gives more trust to the analysis and the model.

For the Fête de la Musique data presented on Figure 1, estimated value of α ranges between 0.17-0.28 for various time snapshots. Since the clusters are associated with geographical positions, we employed the method described in Section III-C. The obtained estimates indicates a heavy burstiness of the data and speaks in favour of the use of TαS as an appropriate model. We also observed a poor fit of the Sibuya distribution to small values of the cluster sizes and tried more general branching-stable process model [9] to improve it.

V. Conclusion

Thinning-stable processes provide a powerful framework for bursty spatial data. Their estimation requires careful selection of inference methods based on dataset characteristics, specific novel inference methods are developed in this paper. Further generalisations, such as branching-stable processes, see [9], can enhance modeling flexibility.

Acknowledgements

The author is thankful to his Master students: Brunella Spinelli and Stefano Crespi for algorithmic implementation in R of the described above methods and for extensive simulation studies.

References

[1] F. Baccelli, M. Klein, M. Lebourges, and S. Zuyev. Stochastic geometry and architecture of communication networks. *J. Telecommunication Systems*, 7:209–227, 1997.

[2] François Baccelli and Bartłomiej Błaszczyszyn. *Stochastic Geometry for Wireless Networks*. Cambridge University Press, 2018.

[3] C.M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

[4] David Coupier. *Stochastic Geometry Modern Research Frontiers: Modern Research Frontiers*. Springer, 01 2019.

[5] Yu. Davydov, I. Molchanov, and S. Zuyev. Stability for random measures, point processes and discrete semigroups. *Bernoulli*, 17(3):1015–1043, 2011.

[6] A.K. Erlang. Sandsynlighedsregning og telefontrafik. *Tidsskrift for Matematik B*, 20, 1909.

[7] L. Kleinrock. *Queueing systems. Volume I: Theory*. Wiley, 1975.

[8] C. Palm. Intensitätsschwankungen im fernsprechverkehr. *Ericsson Technics*, 1943.

[9] G. Zanella and S. Zuyev. Branching-stable point processes. *Electronic J. Probab.*, 20(119):1–26, 2015.

[10] S. Zuyev. Stochastic geometry and telecommunications networks. In Kendal. W. S. and I. Molchanov, editors, *Stochastic Geometry: Highlights, Interactions and New Perspectives*, pages 520–554. Oxford University Press, 2009.