This is a post-print of a paper published in Elsevier Computer Communications:
M. Wellens, J. Riihijärvi, and P. Mähönen, "Spatial Statistics and Models of Spectrum Use",
Elsevier Computer Communications, XX, YY-ZZ (2009).

http://dx.doi.org/10.1016/j.comcom.2009.08.004

Spatial Statistics and Models of Spectrum Use

Matthias Wellens[a,*], Janne Riihijärvi[a], Petri Mähönen[a]

[a] *Department of Wireless Networks, RWTH Aachen University, Kackertstrasse 9, D-52072 Aachen, Germany*

Abstract

In order to opportunistically exploit unused radio spectrum nodes of dynamic spectrum access (DSA) networks monitor the spectrum around them. Such cognitive radios can greatly benefit from a spatial characterization of spectrum use. However, there is need to find an efficient way to describe spatial use, something which has not been studied in details so far. In this paper, we introduce spatial statistics techniques as promising methods to describe spectrum use and enable optimization of DSA networks. We discuss two approaches to spatial modelling of spectrum, namely a deterministic approach based on a system model of the complete radio environment and an empirical approach that exploits passive measurements of the spectrum use. We elaborate on the impact of different network properties on the models and provide realistic parameter sets for generation of simulation scenarios. Additionally, we investigate cooperative sensing as a use case for spatial statistics based runtime optimization of the network configuration.

Key words: Spatial statistics, spectrum model, dynamic spectrum access, node location distribution, spectrum measurements

1. Introduction

The success of wireless communication services and the advent of broadband systems have resulted in an increased need for radio spectrum. At the same time, regulatory agencies have given out licenses for most of the radio spectrum with commercially interesting propagation characteristics and allocating further spectrum bands to new services or technologies is a difficult and complicated process. However, several extensive measurement campaigns have recently shown that the utilization of radio spectrum is significantly lower than indicated by the regulatory situation [1, 2, 3, 4, 5]. These findings and the availability of more flexible radio hardware have motivated the vision of dynamic spectrum access (DSA) [6, 7]. Secondary users[1] search for unused spectrum bands and access those in an opportunistic fashion without causing harmful interference to the primary users:

The task of identifying vacant and used spectrum bands, usually known as spectrum sensing, is crucial for efficient and reliable system operation. Single spectrum sensors may miss primary transmissions if the primary signal is subject to deep fading or shadowing effects. Cooperative sensing has been proposed and evaluated in detail as solution to this problem [9, 10, 11]. In particular, it has been shown that correlated shadowing further complicates cooperative sensing because the amount of new information that other secondary users can provide is limited. Ghasemi and Sousa have shown in [12] that cooperating with farthest away secondary users optimizes the cooperation gain. Most of these results have also been confirmed through practical measurements [13, 14, 15]. The theoretical derivations have been based on models for

*Corresponding author
Email addresses: mwe@mobnets.rwth-aachen.de (Matthias Wellens), jar@mobnets.rwth-aachen.de (Janne Riihijärvi), pma@mobnets.rwth-aachen.de (Petri Mähönen)

[1] Parties that possess official licenses to use a certain spectrum band are usually called primary users [8]. Users that opportunistically access vacant spectrum are called secondary users, respectively. Primary users are also referred to as incumbent users.

correlated shadowing initially developed in the context of cellular network planning and optimization [16, 17, 18, 19]. Although these models turned out to be very useful in the DSA context, their applicability for further extensions is limited due to their focus on cellular frequencies and a single spatial dimension. The sole dependence on distance should be replaced by a more complete system description taking into account at least two independent spatial dimensions.

In this paper, we address the need to have more flexible spatial spectrum models. We apply spatial statistics techniques [20] to spectrum modelling. In particular, we use the theories of random fields and point processes to model the power spectral density (PSD) over space. Both techniques have been applied to selected problems in wireless communications before. Random fields have been used to model different phenomena in the wireless sensor networks community [21] and we have introduced point processes and their statistics as appropriate models for node location distributions in our earlier work [22, 23, 24]. In previous work we have also explored spatial statistics for spectrum modelling from a more conceptual point of view [25]. Additionally, Willkomm *et al.* have used some methods of spatial statistics to evaluate their measurement results on base station load in a cellular network [26].

The goal of this paper is to introduce a versatile methodology for spatial spectrum modelling and to provide realistic model parameters that enable direct use of the spectrum model by other researchers. Additionally, we elaborate on selected impact factors in order to develop an understanding which aspects of wireless communications change spatial spectrum use in which ways. Without such an understanding no reasonable modelling can be achieved. We do not aim at a specific model application or a direct performance improvement of DSA systems although the proposed techniques have the potential to achieve this. Moreover, we do not want to focus on a single and very detailed scenario but keep the discussion intentionally on a rather generic level in order not to limit the applicability of the introduced techniques. Nevertheless, we discuss possible applications in the system modelling as well as in the online system adaptation and optimization areas.

We differentiate between two modelling approaches. First, empirical models reproduce the spatial statistics of spectrum use as previously measured in real-life experiments. Such models are rather simple to use and can be sufficiently described with few parameters. However, none of the characteristics of the primary user system can be changed. We investigated the empirical approach in part in [27]. Second, deterministic models reconstruct the complete primary user network and compute the spectrum use based on a system model. The complexity of this approach is significantly higher but the impact of single factors such as the node location distribution or the propagation environment can be investigated separately. We studied the deterministic model in a slightly different context in [28]. In this paper, we will discuss both approaches. We will present measurement-based model parameters that enable direct use of the spectrum model, and discuss the impact of single factors using the deterministic model. We will show that the propagation model and the node location distribution have significant impact on the spatial statistics. In contrast, the effect of the primary user transmission power is nearly negligible.

There is a multitude of applications for spatial spectrum models. The most obvious is the generation of simulation scenarios for evaluation of DSA algorithms and protocols. Further theoretical work on cooperative sensing can also benefit from the flexible description of the correlation properties of spectrum use. Additionally, cooperative localization of the primary user [29, 30, 31] and subsequent estimation of the no-talk area [32] are further examples. Finally, cognitive radios can use the described metrics to classify the radio environment they are in and increase the awareness of their surroundings, which Mitola introduced as one major system characteristic in his initial definition of the term *cognitive radio* [33]. In particular, cognitive radios can build up spectrum models during runtime and use them in order to estimate the spectrum use at locations from which no sampling data are available. Detailed information on the wireless network topology can also be extracted and used for protocol parameter optimization. Similar approaches have been proposed as applications of radio environment maps [34, 35].

The remainder of this paper is structured as follows. We introduce spatial statistics in more detail in Section 2 and describe how we apply these techniques to the modelling of spectrum in Section 3. In Section 4 we present results for the deterministic model and continue in Section 5 with the analysis of the measurement-based empirical model. We discuss further aspects in Section 6 and, finally, conclude the paper in Section 7.

2. Spatial statistics

Spatial statistics have been applied in a vast variety of problem areas ranging from social and medical sciences to engineering and technology. Here, we focus on two specific fields of spatial statistics, namely random fields and point processes. We focus on those theoretical foundations of both areas that we apply in this paper and refer the reader to the excellent existing literature for a more comprehensive review [20, 36].

2.1. Random fields

Random fields belong to the field of geostatistics which is seen as a subfield of spatial statistics. These methods have been originally developed in order to describe different properties of soil, e.g., the concentration of ore, based on a limited number of samples. Deterministic models cannot fully describe these phenomena since sampling the soil is expensive and variation of soil is too complex. Instead, we describe the examined phenomena as realization of a *random field* $Z(\mathbf{u})$ based on the model

$$Z(\mathbf{u}) = \mu + \epsilon(\mathbf{u}), \tag{1}$$

where μ is the average over the whole random field and ϵ is a random residual. The position $\mathbf{u} \in D$ is located in the domain of interest D which depends on the chosen model and the sampling process and is a part of $\mathbb{R}^2$ or $\mathbb{R}^3$. We use the capital letter Z in order to show the relationship between the one-dimensional random process and the multi-dimensional random field.

The initial goal of geostatisticians has been to reliably estimate soil properties also at locations that have not been sampled. They achieve this goal based on the correlation characteristics of the random residual $\epsilon(\mathbf{u})$ over the domain of interest D. They have developed the theoretical foundations and multiple metrics to describe the correlation characteristics. Here, we introduce the *semivariogram*

$$\begin{aligned} \gamma(\mathbf{u},\mathbf{v}) &= \frac{1}{2}\operatorname{Var}(Z(\mathbf{u}) - Z(\mathbf{v})) \\ &= \frac{1}{2}\mathbb{E}\big[(Z(\mathbf{u}) - Z(\mathbf{v}))^2\big] \\ &= \frac{1}{2}\mathbb{E}\big[(\epsilon(\mathbf{u}) - \epsilon(\mathbf{v}))^2\big], \text{ with } \mathbf{u},\mathbf{v} \in D, \end{aligned} \tag{2}$$

where $\mathbb{E}$ signifies expectation.[2] In the literature γ is also referred to only as variogram although this term is usually defined without the additional factor 1/2. The semivariogram γ will only depend on $\mathbf{h} \equiv \mathbf{v} - \mathbf{u}$ if Z is (intrinsically) stationary. If Z is also isotropic γ can be completely described by its dependence on the distance $h = \|\mathbf{h}\|$. In the case of second order stationarity we can rewrite equation (2) to the equivalent formula defining the spatial autocovariance

$$\begin{aligned} C(\mathbf{h}) &= \mathbb{E}\big[\big(Z(\mathbf{u}) - \mu\big)\big(Z(\mathbf{u}+\mathbf{h}) - \mu\big)\big] \\ &= C(\mathbf{0}) - \gamma(\mathbf{h}), \end{aligned} \tag{3}$$

where $C(\mathbf{h})$ denotes the autocovariance at lag $\mathbf{h}$ and $C(\mathbf{0})$, the autocovariance at lag zero, is the variance of the process. We can also define the spatial analog to the autocorrelation function $\rho(\mathbf{h}) = C(\mathbf{h})/C(\mathbf{0})$, which is usually referred to as *correlogram* but in the literature it is also known simply as the correlation function.

When coming back to the geostatistical background we rewrite the semivariogram as

$$\gamma(\mathbf{h}) = \tau^2 + \sigma^2\big(1 - \rho_0(\mathbf{h}/\phi)\big), \text{ with } \tau^2 \in \mathbb{R}_0^+ \text{ and } \sigma^2, \phi \in \mathbb{R}^+, \tag{4}$$

where the newly introduced parameters have more descriptive meanings. The sampling error τ^2 is also referred to as *nugget variance* because larger deviations occurred in the sampling of soil when rare nuggets

[2] Often, the mean μ as introduced in equation (1) is not only a single scalar but replaced by a location-dependent variable in order to describe trends present in the data. In this case the simplification in the last step in equation (2) cannot be applied anymore.

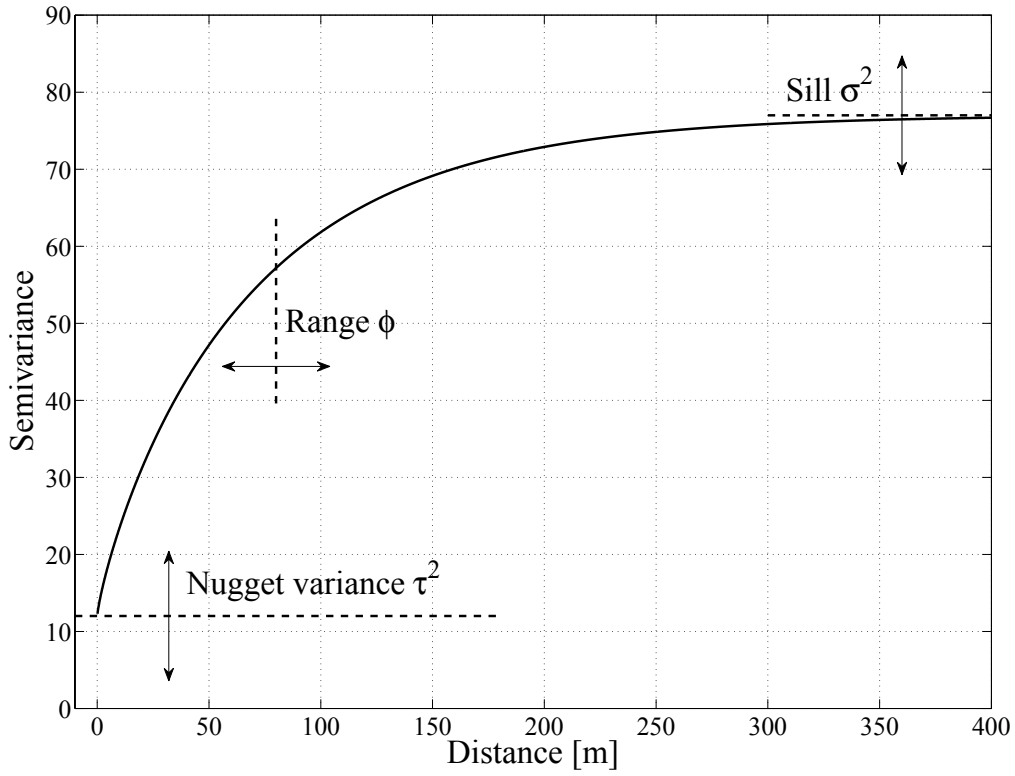

Figure 1: Meaning of the semivariogram model parameters and their impact on the shape of the semivariogram based on the example of the Mátern model.

have been found at the sampling locations. In most scenarios the nugget variance should be a small value or even zero. The parameter σ^2 is known as *sill* because it describes the maximum of the semivariogram $\lim_{\tau^2=0,\|\mathbf{h}\|\to\infty}\gamma(\mathbf{h})=\sigma^2$. It is equal to the variance of the process $C(\mathbf{0})=\sigma^2$. Finally, the third parameter ϕ is known as the *range* of the semivariogram and describes at which distance two samples are not correlated anymore.[3] The parameter is introduced as an additional scaling factor to the definition of the correlogram as $\rho(\mathbf{h})=\rho_0(\mathbf{h}/\phi)$.

Figure 1 shows an example for the semivariogram and how the three described parameters change its shape. The nugget variance τ^2 is the y-axis intercept and will move the whole curve up if $\tau^2>0$. Together with the nugget variance the sill σ^2 determines the asymptote of the semivariogram $\tau^2+\sigma^2$ that will be reached if $\rho_0(\mathbf{h}/\phi)=0$. Finally, as discussed above, the range ϕ scales the whole semivariogram on the x-axes and describes the distance at which the phenomena under study will not be correlated anymore.

In addition to the three usually scalar parameters τ^2, σ^2, and ϕ we need to specify the form of ρ_0 in order to sufficiently describe a semivariogram model. The *exponential model*

$$\gamma_{\exp}(h)=\tau^2+\sigma^2\big(1-\exp(-h/\phi)\big), \tag{5}$$

[3]For asymptotically silled semivariograms, two points will only be spatially uncorrelated in the limit as $\|\mathbf{h}\|\to\infty$. In this case, let us denote ϕ as the *effective range*, which is not uniquely defined [37] but is often set to 95% of the range. Additionally, there exist semivariogram models that are not asymptotically silled but for the empirical model we do not consider those because the modelled processes are non-stationary.

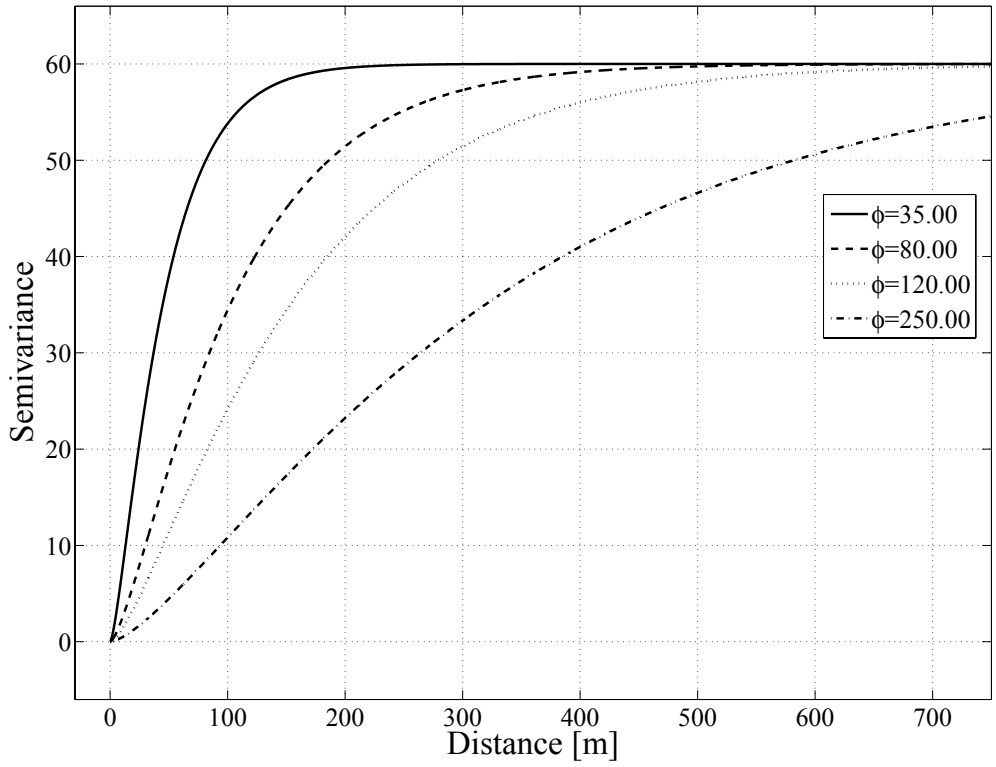

Figure 2: Comparison of different values for the parameter ϕ of the Mátern model on the shape of the semivariogram.

and the *Gaussian model*

$$\gamma_{\mathrm{G}}(h) = \tau^2 + \sigma^2\big[1 - \exp\big(-(h/\phi)^2\big)\big], \tag{6}$$

are two popular examples that only rely on the described three parameters. The very general *Matérn model*

$$\gamma_{\mathrm{M}}(h) = \tau^2 + \sigma^2\left(1 - \frac{1}{2^{\kappa-1}\Gamma(\kappa)}\left(\frac{h}{\phi}\right)^{\kappa} K_{\kappa}\left(\frac{h}{\phi}\right)\right), \tag{7}$$

is a very flexible model, where $\kappa \in \mathbb{R}^+$ is introduced as a fourth model parameter, Γ is the gamma-function, and K_κ is the modified Bessel function of the second kind. The Matérn model can reproduce the exponential and the Gaussian model as the special cases $\kappa = 1/2$ and $\kappa \to \infty$, respectively. Additionally, we consider the *Cauchy model*

$$\gamma_{\mathrm{Cau}}(h) = \tau^2 + \sigma^2\left[1 - \left(1 + (h/\phi)^2\right)^{-\kappa}\right], \text{ with } \kappa \geq 0, \tag{8}$$

as one example for models with power-law correlation characteristics.

Figures 2 and 3 show further examples how the shape of the semivariogram changes depending on the chosen parameter values. We selected the range ϕ and the fourth parameter κ of the Mátern model since the meaning of the nugget variance τ^2 and the sill σ^2 is rather straightforward. Figure 2 shows that a higher range results in a lower slope of the semivariogram. Higher values for κ result in the addition of an inflexion point to the shape of the semivariogram, as shown in Figure 3. Accepting the increased complexity of a four-parameter model gives the additional freedom to change also the basic type of the shape of the semivariogram compared to the models with only three parameters.

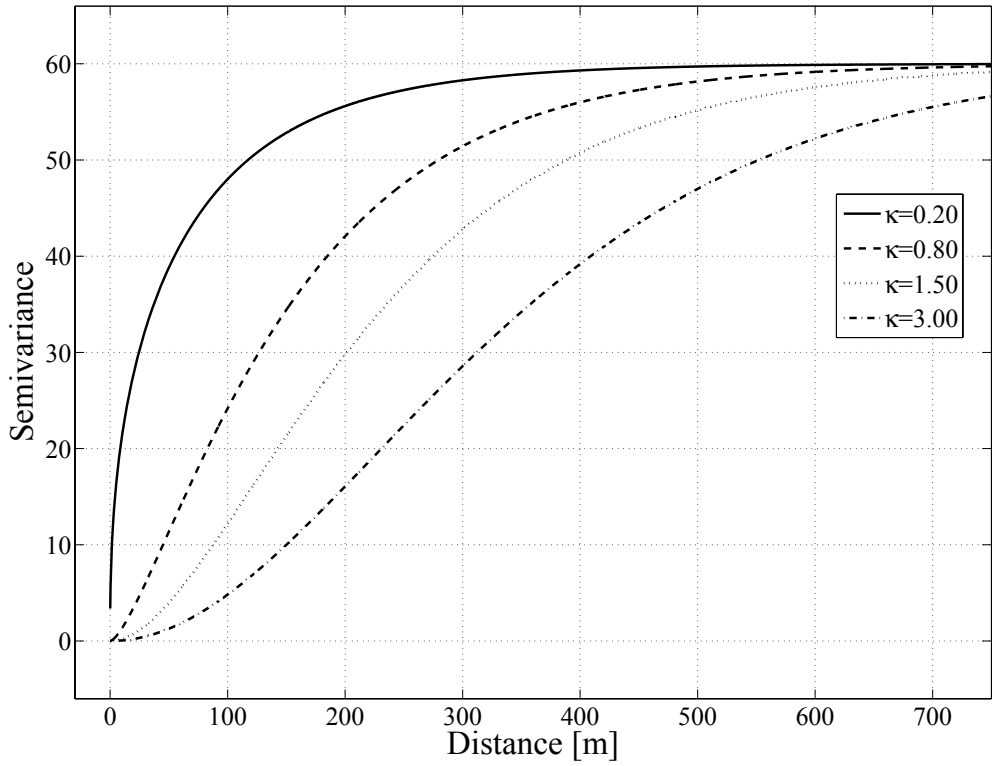

Figure 3: Comparison of different values for the parameter κ of the Mátern model on the shape of the semivariogram.

2.2. Point processes

In order to study the influence of transmitter locations on the spatial statistics of spectrum, we shall introduce now some simple stochastic models for those locations. For the purposes of this paper it suffices to think of such models, technically called *point processes*, as simply random processes yielding as realizations collections of points $\{x_i\} \in \mathbb{R}^2$ (for more precise definitions we refer the reader to [38]). Both the coordinates of the points as well as their overall number can in general be random, and the coordinates are in general *not* independent as random variables. Typically in theoretical calculations and simulations the *Poisson point process* is considered, for which the number of points is a Poisson random variable, and the coordinates of the points are uniformly and independently distributed over the area under consideration. Recent work indicates that the Poisson approximation does not seem to be very accurate in any length scale for real networks, so other types of point processes should be considered as well [22, 24].

In the following we shall consider three types of point processes as models for transmitter locations. We do not claim that any of these models is particularly realistic, but instead use them as a tool to understand the impact of different features of the location distribution on the spatial statistics of spectrum. The first model we shall consider is the Poisson point process discussed above. Due to its complete randomness and high prevalence in the literature it makes for an important reference case. We shall also consider the grid process, in which points are placed on a regular grid with specified inter-node distances in horizontal and vertical directions. Finally, we shall use the *Thomas process* as an example of a clustered node distribution. Realizations of the Thomas process are constructed by first choosing number and locations of cluster centers from a Poisson point process with mean λ. Then, each cluster is formed by Poisson distributed number of nodes, with mean ν, the coordinates of which are sampled from a two-dimensional symmetric normal distribution with variance σ^2. The average total number of nodes is thus $\lambda\nu$.

3. Modelling spectrum with spatial statistics

Until now we have introduced the underlying theory of random fields and point processes. Next, we describe how appropriate model parameters for a spatial spectrum model can be determined. We discuss two major approaches. The empirical approach relies on extensive measurements and reproduces the correlation characteristics of the measured data. The deterministic method is based on a system model for the primary user network and the radio environment. If all factors are modelled appropriately we can compute realistic power spectral density (PSD) samples at arbitrary locations. We can interpret these samples as a random field in both approaches. If we estimate the semivariogram and afterwards fit one of the introduced model types we can extract parameter sets that ensure accurate reproduction of the primary user system.

We start with the empirical case and describe in detail how the model parameters can be determined from measurement results in Section 3.1. Afterwards, we introduce the system model required for the deterministic approach in Section 3.2.

3.1. Sampling the spectrum

As mentioned above the empirical approach relies on extensive measurements of the random field. Depending on the wireless system of interest the PSD has to be sampled at a certain frequency and at appropriate sampling locations. These locations depend on the typical coverage area of a single transmitter because samples measured at locations much further away from each other will most probably not be correlated anymore. Moreover, the area where we expect non-negligible correlation has to be sampled with high enough resolution in order to provide sufficient data for model fitting later on. Further discussion on the selection of an appropriate sampling grid is provided in [20, 39].

At each sampling location a sufficient number of PSD samples should be measured. The corresponding value of the random field is computed by simple averaging in order to lower the impact of measurement artefacts such as receiver noise. The number of samples taken at a single location depends mainly on the desired confidence interval for the average PSD value, obtained as the standard error of mean. The number of locations chosen influences then the accuracy of the semivariogram estimate. Typically with already 30–40 measurement locations good estimates are possible, assuming a regular sampling pattern such as a

grid. Depending on the measurement methodology and the wireless system preceding data processing steps may be required in order to ensure an accurate measurement of the signal strength of the examined service.

The next step is the fitting of the semivariogram model parameters to the measured PSD field. Multiple techniques have been applied by the spatial statistics community, namely least squares fitting [20, 40], maximum likelihood parameter estimation [41], and more advanced approaches based on Bayesian inference [37]. Here, we focus on the least squares approach because of its simplicity and popularity. Specifically, we apply the weighted least squares estimation (WLSE) as introduced by Cressie [20, 40].

In contrast to the maximum likelihood approach that estimates the model parameters from the random field, for the WLSE we need an intermediate estimate for the semivariogram. We use the *empirical semivariogram* as introduced by Matheron [42], which is the most popular one in the literature. It is also referred to as *method of moments*. Other estimators are presented in [20] and [40]. The algorithm computes the empirical semivariogram at quantized distances based on a binning approach. Let $N(h)$ denote the set of pairs (i, j) such that $h - \Delta \leq \|\mathbf{x}_j - \mathbf{x}_i\| < h + \Delta$, where $\mathbf{x}_i, \mathbf{x}_j \in D$ and $\Delta \in \mathbb{R}^+$. Self-matches should not be taken into account, either by explicitly not considering those pairs or by placing the first bin such way that its lower border does not include zero distance. The empirical semivariogram $\hat{\gamma}(h)$ per bin is computed after averaging over all sample-pairs in each bin:

$$\hat{\gamma}(h) \equiv \frac{1}{2|N(h)|} \sum_{N(h)} (Z_j - Z_i)^2. \tag{9}$$

For a detailed discussion on the accuracy of the empirical semivariogram estimate as a function of the number of samples per bin, see [20].

The binning of the distances opens possibilities for further enhancements of the estimation process. In the described technique each estimated semivariogram value $\hat{\gamma}(h)$ is placed at the middle of the bin. However, the distances grouped in a bin might not be uniformly distributed over the bin. Using the average $\bar{h}$ of all grouped distances instead of the middle of the bin enables more accurate estimation of the semivariogram $\hat{\gamma}(\bar{h})$. Since the WLSE does not rely on equidistant bins the fitting process can be used unchanged. It may also be more appropriate for the available sampling data to apply bins of different widths. We chose a constant bin width such that the number of samples per bin is sufficient in order to ensure a low enough estimation variance. Further discussion on adapted binning is given in [39].

In the final step the WLSE determines the parameter set that fits the empirical semivariogram best. During the fitting process the WLSE weights each bin based on the number of sample-pairs that belong to that bin. A bin with less sample-pair is weighted less. Finally, we can also use the goodness-of-fit metric, provided by the WLSE, to compare different model types and decide if the added complexity of a fourth parameter in the case of, e.g., the Matérn model provides more realistic reproduction of the measurement data or a simpler model based on three parameters is sufficient. However, since the goodness-of-fit metric is not normalized comparing fits across multiple data sets is not possible.

3.2. System model

In the deterministic case we use a system model to simulate a realistic environment and compute the PSD that could be received at any location in the simulated environment. Therefore, we replace the measurement process used for the empirical model by the computational approach based on the system model but both techniques result in a field of PSD values. The latter steps of estimating the empirical semivariogram and fitting it to one of the model types are same for both approaches as shown in Figure 4. The sampling grid will most probably be different in both approaches because the effort required to cover more points is significantly higher in the empirical case. We will focus here purely on the system model, which describes the part unique to the deterministic model.

In order to compute a realistic estimate of the received PSD over space we have to consider the following factors:

- Transmitter location distribution.
- Transmitter density.

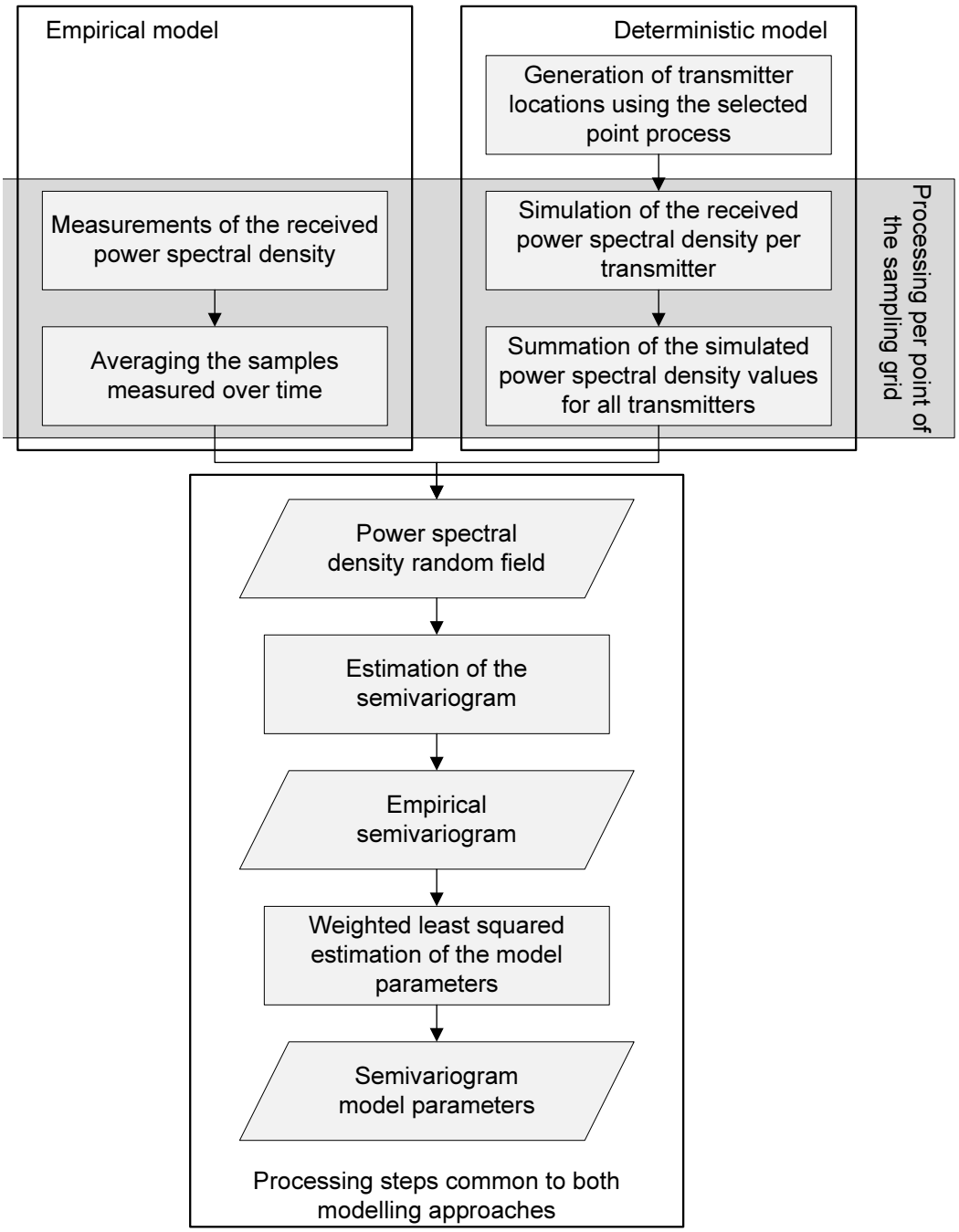

Figure 4: Flowchart of the different processing steps required to extract model parameters using either the empirical or the deterministic approach.

- Transmit power and operating frequency.
- Receiver and transmitter antenna characteristics.
- Propagation environment and terrain.
- Transmitter activity model.

The transmitter location distribution has direct impact on the coverage of a wireless system. An even distribution over space will result in better coverage compared to a clustered distribution. Additionally, for planned networks also the required network capacity will change the transmitter location distribution. Operators deploy more base stations in urban environments compared to rural areas with less users. However, for some wireless systems, such as wireless local area networks (WLANs) no centralized network planning is applied and the nodes are distributed in an unplanned way. Varying the transmitter location distribution but keeping the other parts of the system model constant enables us to investigate the impact of such differences. The transmitter density is closely related to the distribution because less transmitters obviously result in worse network coverage.

The transmit power, the working frequency, and also the antenna characteristics including its installation height describe the hardware configuration of each transmitter. Together with the propagation environment and the terrain these parameters determine the communication range of the simulated system. Using these factors we can examine different system configurations, and differentiate rural and urban scenarios. The propagation modelling is probably the most complex building block of the system model and the tradeoff between computational complexity and simulation accuracy has to be decided upon. Additionally, very detailed models such as raytracing techniques often require information on buildings or accurate terrain data that may not always be available.

Finally, the transmitter activity model describes the traffic carried by the network as well as the system behaviour over time. There are simple cases such as broadcasting applications that are simplex systems and do not apply power control due to the lack of feedback. Duplex systems will require more accurate modelling especially if time duplexing is applied. Additionally, the chosen multiple access scheme and the traffic patterns have significant impact on the transmitter activity. The PSD received at a selected location is in most cases the superposition of signals emitted by multiple transmitters. Therefore, a very accurate model may also require to consider correlations in the activity patterns of multiple transmitters over time. Due to the well defined framing structure of several wireless systems and the dependence on the activity of the users, such correlations are to be expected.

However, if we do not assume any primary user cooperation complex primary user behaviour may render DSA impossible because the loss of single frames may not be avoidable. Since DSA is based on the requirement not to cause harmful interference a focus on primary systems with simple activity patterns is reasonable.

4. Deterministic spectrum model

The aim of this paper is to introduce a spatial spectrum model and explore its sensitivity to selected characteristics of wireless communication systems. The latter enables researchers to quickly and reliably decide if the proposed model is appropriate for their purpose and to which level of detail their system model has to reproduce the reality for their research problem at hand. In order to enable a variety of model applications we limit the number of scenario specific configurations to the required minimum.

The deterministic approach and its underlying system model enable us to investigate the impact of different network properties on the spatial statistics of spectrum use. We can keep all settings same but increase, e.g., the transmit power and re-evaluate the resulting random field. Additionally, this methodology builds a basis for comparison of empirical models and greatly helps to explain differences in the measured spatial statistics. In order to provide this understanding we do not require the system model to reproduce all aspects of real systems in very realistic manner. Instead, we can abstract some of these properties under appropriate conditions. Here, we base our system model on the following assumptions.

We focus on a typical broadcasting application and do not apply complex models to describe the activity of the primary user over time but assume continuous transmission without any power control. Additionally, we do not consider antenna characteristics but limit our analysis to idealized omnidirectional and isotropic antennas for both the transmitter and the receiver. We also do not evaluate the impact of hills or buildings and use a flat terrain. Adding one of the omitted details to our models would be straightforward. For further system parameters we mostly mimic the Global System for Mobile (GSM) communications and, e.g., use the frequency $f = 915\,\mathrm{MHz}$.

We examine four parameters in detail, namely the transmitter location distribution, the propagation model, the transmit power, and the node density. We compare different settings based on the semivariogram, which we compute using few selected software tools. We generate the node location samples using the `spatstat` [43] package in the `R` environment [44]. For propagation simulation we use the `WinProp` software suite [45]. Finally, we compute the empirical variogram using an appropriate `MATLAB` script. Since the deterministic model does not use realistic models for all components of the system model we do not apply the WLSE parameter fitting to the computed semivariograms but limit it to the empirically measured data sets.

4.1. Transmitter location distribution

The first impact factor we investigate is the transmitter location distribution, which is a fundamental property of a communication network. We want to compare how the spatial statistics of the PSD change based on different transmitter location distributions. For this purpose we use different point processes to generate transmitter location samples that we use for simulation of the PSD random field. We have shown in our earlier work that the methodology of point processes is flexible enough to accurately reproduce node location distributions of various wireless systems [22, 23, 24].

In particular we consider the Poisson process, a regular equidistant grid, and the Thomas process as an example for clustered point processes. Unless stated otherwise the simulations were carried out in a square area of $5\,\mathrm{km} \times 5\,\mathrm{km}$, with transmit power of the nodes set to $80\,\mathrm{W}$, transmitter antenna height of $30\,\mathrm{m}$, receiver antenna height of $2\,\mathrm{m}$, and operating frequency of $915\,\mathrm{MHz}$. The point process parameters used were as follows. For a grid we used the regular arrangement of 6×6 nodes, for the Poisson process mean intensity yielding on average 40 nodes in the simulation area, and for Thomas process the parameter values of $(\lambda, \nu, \sigma) = (3, 13, 500\,\mathrm{m})$.

We assume the transmitters to be static in correspondence to most of the common DSA-related scenarios. However, we expect that the shapes of the arising semivariograms are not very sensitive to mobility provided that the overall density of the nodes satisfies a suitable stability property. In this case, using the stationary distribution of the node locations as the point process and proceeding as in the static case should yield good approximation for the semivariance of the mean PSD field at any given time [46].

Figure 5 shows the semivariogram for the described parameter sets. The grid results in a wavelike shape because of its periodic structure. PSD samples computed at locations at similar relative positions inside the grid will always be correlated although they may be served by different transmitters. Locations at other relative positions are less correlated. When neglecting border effects, the same relative position in the grid occurs every integer-multiple of the grid size giving the wavelike shape of the semivariogram. As expected, the distance between two maxima of the semivariogram corresponds to the grid step size of $833\,\mathrm{m}$.

Due to its regular structure the maximum semivariance computed for the grid structure is lower compared to all other point processes. The grid step size is obviously also the minimum distance between two transmitters. This minimum distance is larger than the minimum distance in the other two scenarios and results in the highest semivariance on shorter distances up to about $600\,\mathrm{m}$ because, starting from each transmitter, the PSD steadily decreases until half of the grid step size.

The Poisson process yields nearly as high variance at close distances as the grid case because very closely located transmitters are improbable but may occur. Due to the random nature of the point process, at each location a different superposition of emitted signals is simulated, no specific distances guarantee highly correlated PSD samples, and the semivariance steadily increases over distance. However, the random distribution of nodes across the simulated area limits the probability of highly varying transmitter densities

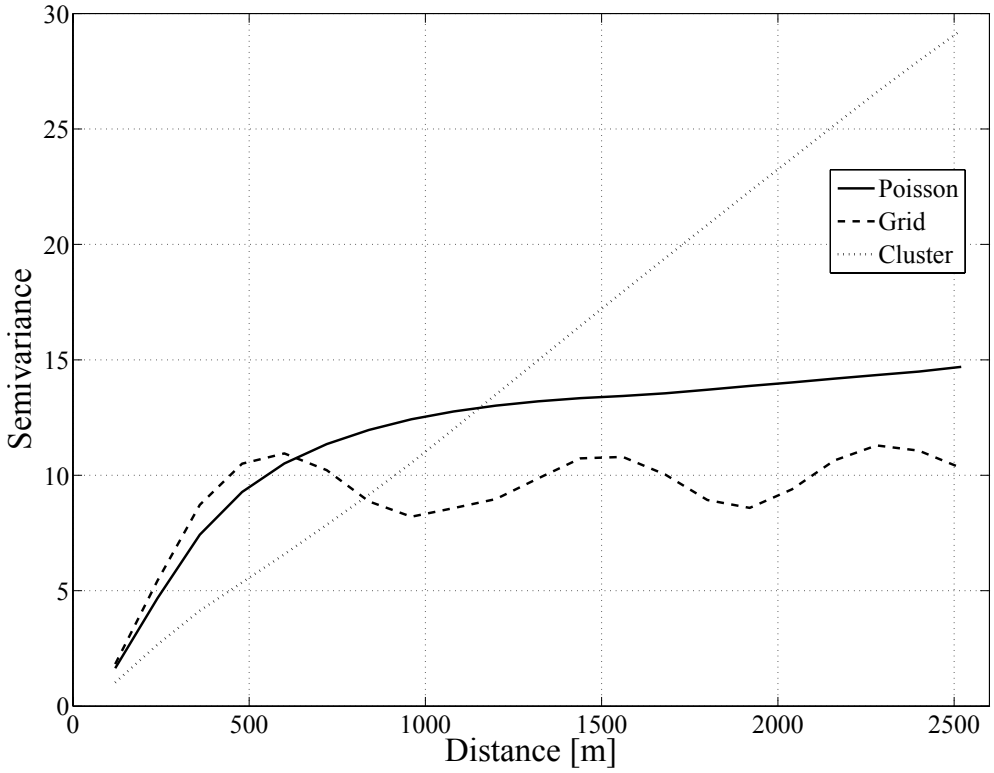

Figure 5: Impact of the underlying point process on the semivariogram based on system parameters configured to resemble a GSM network.

between subareas and, thus, also limits the variance of the PSD, which corresponds to the sill of the semivariogram.

Finally, the clustered process describes networks in which each transmitter has probably other transmitters close by. The PSD inside these clusters is high but changes also less if compared with the other two scenarios. These minor changes result in the lowest semivariance at close distances. In contrast, outside all clusters the PSD will drop further compared to the other cases and the semivariance reaches the highest simulated values. The coverage of such a network will be very good as long as the user is located inside a cluster but will steadily degrade the further the user is located from the nearest cluster.

4.2. Propagation model

For the propagation environment various models can be considered ranging from the simple free space propagation model to accurate raytracing approaches that take into account information on buildings and the terrain. Since we do not focus on a specific scenario in this paper, we do not consider the terrain or the type of housing present in a certain area. Instead, we start from fundamental free space propagation and compare it to the well known model that Hata developed [47] leveraging the previous work of Okumura *et al.* [48]. In the following we will refer to this model as Hata propagation model or, simply, Hata model.

The pathloss L between transmitter and receiver is given by the Hata model as

$$\begin{aligned}\frac{L}{\mathrm{dB}} = 69.55 + 26.16\log(\frac{f}{\mathrm{MHz}}) - 13.82\log(\frac{h_t}{\mathrm{m}}) - \frac{a(h_r)}{\mathrm{dB}} \\ + \big(44.9 - 6.55\log(\frac{h_t}{\mathrm{m}})\big)\log(\frac{d}{\mathrm{km}}),\end{aligned} \tag{10}$$

with

$$\frac{a(h_r)}{\mathrm{dB}} = \big(1.1\log(\frac{f}{\mathrm{MHz}}) - 0.7\big)\frac{h_r}{\mathrm{m}} - \big(1.56\log(\frac{f}{\mathrm{MHz}}) - 0.8\big), \tag{11}$$

where h_r is the receiver antenna height, h_t is the effective transmitter antenna height, and d is the distance between transmitter and receiver. We use the model configuration for small and medium cities.

The higher pathloss described by the Hata propagation model results in a higher variance of the PSD random field. Since the semivariogram can, in addition to the correlation-based interpretation, be also seen as variance measure all simulated scenarios based on the Hata model give significantly higher semivariances. This effect is present for all three transmitter location distributions as shown in Figure 6. However, the shape of each curve did not change and all major results discussed for the free space case are also valid for the Hata model scenario.

Since free space propagation modelling is unrealistic for most wireless systems and the Hata model was designed with our use case of cellular systems in mind we will limit our further evaluation to the more appropriate Hata propagation model. Additionally, we will focus on the Poisson point process due to its random nature and its popularity for generation of transmitter locations in the wireless communication research community.

4.3. Transmit power

In the next step we evaluate the impact of the transmit power on the semivariogram. Figure 7 shows that changing the transmit power has only negligible impact on the semivariogram. The random nature of the underlying Poisson point process causes the small differences between the shown curves.

If we assume the single transmitter case the simulated PSD random field will have the shape of a circular, isotropic cone because the Hata model does not include any random component. At some distance this cone stops because it hits the thermal noise floor. The slope of the cone is not constant over distance due to the logarithmic behaviour of the Hata model but the PSD steadily decreases with distance. A change in transmit power will shift the height of the cone up or down and change the distance at which the thermal noise floor is hit. However, this distance may never be reached because other transmitters may cover those further away areas. Since the empirical variogram, as defined in eq. (9), considers only the difference between

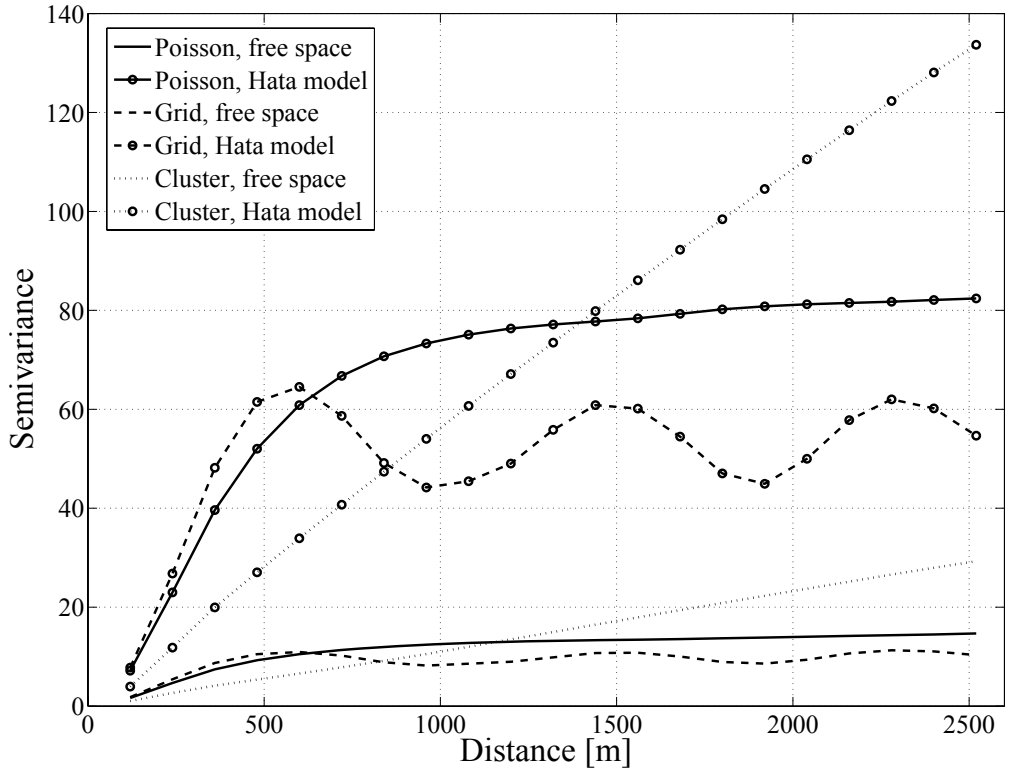

Figure 6: Impact of the propagation model on the semivariogram based on system parameters configured to resemble a GSM network.

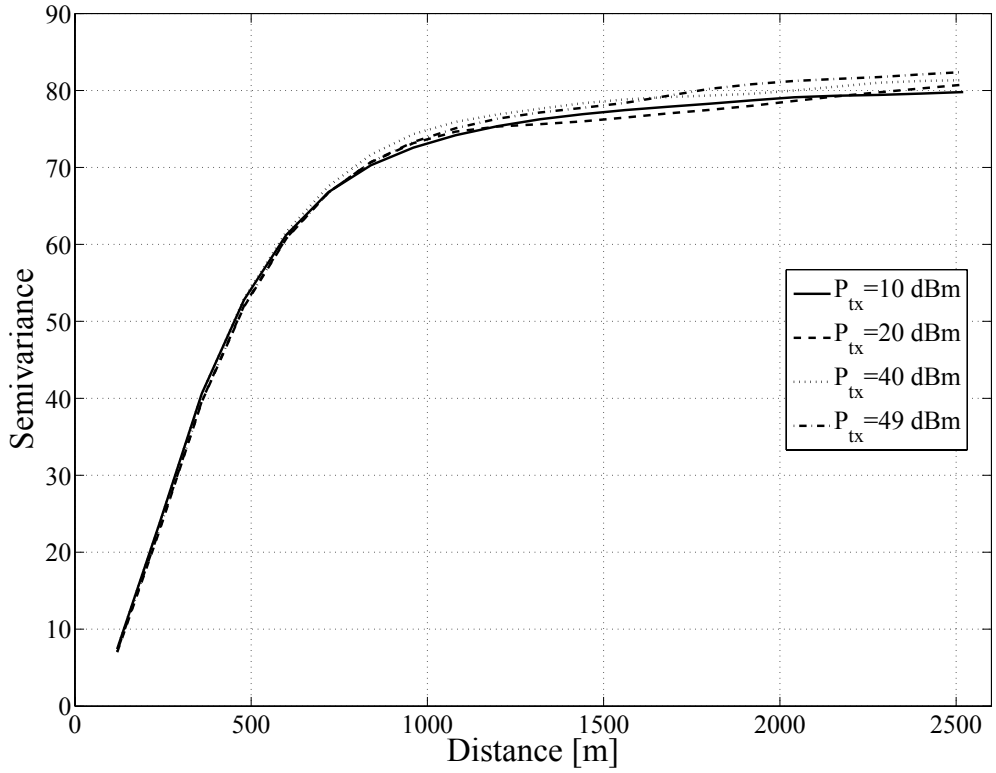

Figure 7: Impact of the simulated transmit power on the semivariogram based on further system parameters configured to resemble a GSM network.

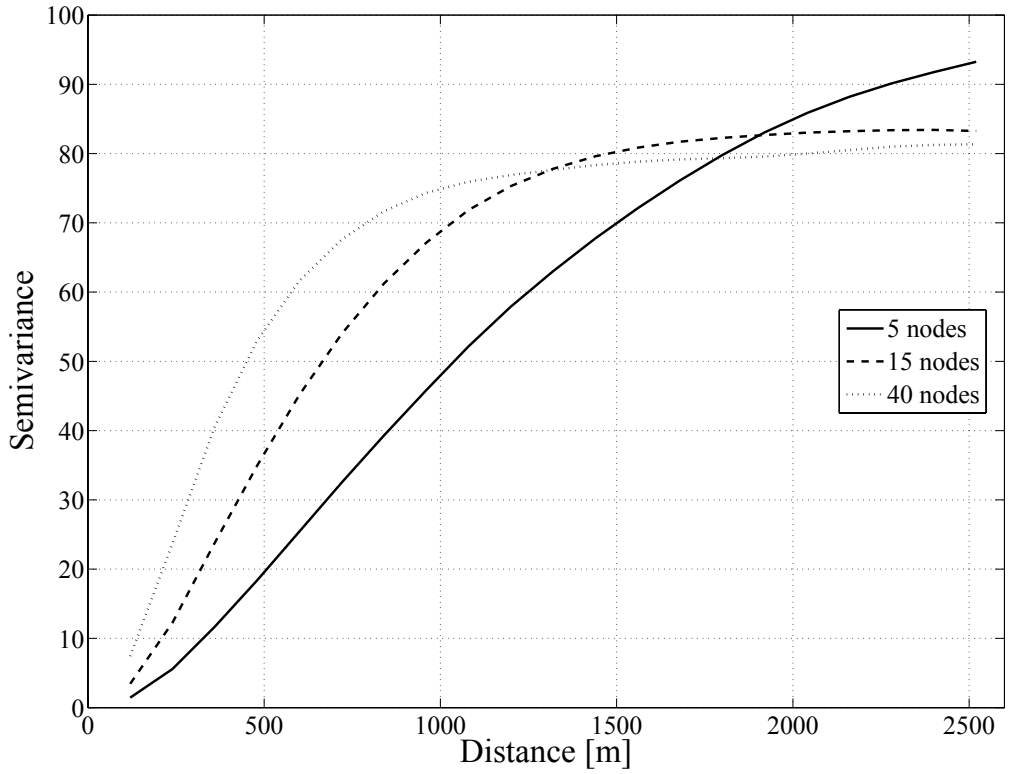

Figure 8: Impact of the node density on the semivariogram based on system parameters configured to resemble a GSM network.

the random field value pairs located at certain distance, a relative shift in the random field will not change the result if the thermal noise floor is never hit.

Only if the PSD will at some distance hit the thermal noise floor a variation in transmit power will change the relative shape of the PSD random field. All configurations shown in Figure 7 result in situations where the PSD simulated for the complete area is above the noise floor. Therefore, the transmit power does not have any impact on the semivariogram.

4.4. Node density

Finally, we investigate the node density as impact factor on the semivariogram. We do not change the size of the scenario and also continue to use the Poisson point process and the Hata propagation model as underlying scenario.

Figure 8 shows that the node density has considerable impact on the semivariogram. It will not change the fundamental shape of the graph but will result in different model parameters if model fitting is applied in a second step. In particular, a decreased node density has two consequences: First, the sill σ^2 of the semivariogram increases. Second, the slope of the semivariogram, which is partially described by the range parameter ϕ, decreases. Both effects are caused by the increased random field variance. If fewer transmitters are deployed in the simulated area the minimum and average distance to the nearest neighbouring transmitter will inherently increase and the minimum and average PSD will decrease, respectively.

The area predominantly covered by a single of the described cone shapes increases and the typical distance to a location with similarly high PSD estimate also increases. Both effects result in the lower semivariances at shorter distances. Since the sill of the semivariogram σ^2 represents the variance of the random field its

increase follows from the worse coverage and higher difference between maximum and minimum simulated PSD.

5. Empirical spectrum model

After discussing the deterministic case based on power spectral density simulations we continue with the empirical modelling approach. Our results are based on the evaluation of extensive spectrum measurements that we performed in spring of 2008 [13]. We start with a short description of the measurement setup and continue with commenting on data preprocessing steps. Afterwards, we discuss the fit of the introduced semivariogram models to the measurement data and the extraction of model parameters for realistic spatial spectrum models. We list fitted parameter sets for several popular wireless systems.

5.1. Measurement setup

We are interested in the spatial statistics of the PSD. Therefore, we performed extensive PSD measurements at two locations, namely the downtown area of Aachen, Germany, and the CeBIT industry fair in Hanover, Germany. The CeBIT is the world's largest computer and electronics fair with around 495,000 visitors in March 2008 and is an extreme measurement scenario because, compared to usual environments, significantly more wireless communication services are used at such events. The spectrum was highly utilized for the most popular wireless services such as broadcasting applications, cellular networks, WLANs, etc. However, frequency bands that have not been allocated to these commonly used services have been underutilized also at the CeBIT industry fair [13]. Thus, this scenario shows the need for additional radio spectrum for wireless communication but it also shows that considerable amount of spectrum is vacant and could be made available for secondary use.

We collected PSD samples at positions in regular grids of different step size. At the CeBIT we investigated the WLAN use inside two connected exhibition halls based on a step size of $\approx 12.5\,\mathrm{m}$. In Aachen we used two different step sizes. In the case of shorter range systems such as WLAN and the Digital Enhanced Cordless Telecommunications (DECT) system we set the step size to $\approx 15\,\mathrm{m}$. In order to cover a large enough area we extended the step size to $\approx 250\,\mathrm{m}$ for evaluation of cellular networks and possibly also the terrestrial version of Digital Video Broadcasting (DVB-T).

The measurement system consists of two equal setups. Each setup can run autonomously for at least ten hours and is based on a portable spectrum analyzer, wideband omnidirectional antennas, and a laptop for instrument control and data saving purposes. Both setups use GPS for localization[4] and time synchronization. Further details are given in [13]. We placed one setup statically at a central position and measured continuously and moved the other setup from grid point to grid point and gathered appropriate amount of data at each location. Using the measurement trace of the static setup we validated that the spectrum usage did not change significantly throughout the day and we can combine the PSD samples taken at different points in time to a PSD spatial field.

5.2. Preprocessing of the measurement data

We applied two major preprocessing steps in order to ensure accurate reproduction of the measurement data during the model fitting process. First, many of the investigated systems do not transmit continuous signals but employ certain duplexing and multiple access schemes. In some cases, the spectrum analyzer sweep measures a frequency during the transmission of a frame but it also sometimes misses an active transmission. Additionally, most wireless systems will not transmit any signal if no information is to be exchanged. If no signal is transmitted the measurement will only describe the thermal and any other background noise processes. Since we are solely interested in the spatial statistics of the primary user communication we have to sort out the additional noise samples. As we examined multiple different systems we applied generic energy detection based on the detection threshold $\delta = -105\,\mathrm{dBm}/100\,\mathrm{kHz}$. Previous

[4] At the CeBIT we examined two connected exhibition halls with indoor GPS coverage installed to enable live demonstrations of navigation devices and services.

tests showed that this threshold value sufficiently limits the probability of false alarms triggered by strong noise samples but at the same time is sensitive enough for most primary user systems.[5]

The second preprocessing step combines different data sets based on the examined service. Let us consider the Universal Mobile Telecommunications System downlink (UMTS DL) as an example. Each channel is 5 MHz wide and based on the configured resolution bandwidth of 100 kHz without any overlap, we get 50 PSD samples per channel. Additionally, multiple channels are used by partially different network operators at one location. In the first step we average the PSD samples that were collected sequentially at the same frequency. In the next step, we average those samples that belong to the same channel. Afterwards, we compute the empirical semivariogram for each UMTS channel separately and, finally, before we perform the parameter fitting, we average the empirical semivariograms over all actively used channels.

In addition to these preprocessing steps we apply the adapted binning process to the estimation of the empirical semivariogram as described in Section 3.1.

5.3. Semivariogram fitting

After we have discussed the measurement methodology we continue with presenting results for the empirical semivariogram and parameter fits. We will focus on the Matérn model for the correlogram because it combines high flexibility with high accuracy.

The first results are based on the measurements taken in the large grid used in Aachen. Recall that we selected the grid step size of $\approx 250\,\mathrm{m}$ in order to properly sample cellular networks such as the GSM or UMTS. Additionally, we performed measurements on lower frequencies in order to possibly also determine model parameters for DVB-T. Figure 9 shows the empirical semivariogram for UMTS, DVB-T, and a pager service. Since all three services emit continuous signals we did not apply any detection threshold but adopted the other discussed preprocessing steps. The pager system is a narrowband service and works at $f = 466\,\mathrm{MHz}$. Although the propagation characteristics of lower frequencies are better and signals are less attenuated, the fitted semivariogram falsely results in very short ranges for both the pager service as well as the DVB-T system. The fitting process failed because the geographical area ($\approx 1750\,\mathrm{m} \times 1000\,\mathrm{m}$) that our sampling grid covered is too small. Either a wider sampling grid or additional grid points are required to gather sufficient amount of data to capture the spatial statistics of these lower frequency services.

In contrast, the sampling grid is appropriate for the UMTS network. The fitted graph matches well to the empirical semivariogram and the nugget variance has been fitted to a low value as expected. Additionally, the graph reaches its asymptote around 800 m which is a quite typical value for UMTS cell sizes in urban areas. The sill is significantly higher compared to the DVB-T and the pager service. In addition to the higher working frequency of UMTS the cellular base stations emit lower transmit powers compared to broadcasting towers. Together, these two factors result in a less uniform PSD over space in the UMTS case and thus the sill, which corresponds to the variance of the random field, is higher for the cellular system.

In the next step we evaluate the impact of the data preprocessing. Figure 10 compares the shorter range DECT and WLAN systems. We start with the comparison of the WLAN results. At the CeBIT the preprocessing lowers the sill of the model fit. If we sort out the noise samples we will take out the lowest values and will lower the variance. However, the preprocessing does not decrease the variance for the Aachen results. Here, two effects compensate each other. First, again, the noise sample rejection lowers the difference between the extreme PSD samples. Second, the duty cycle[6] in Aachen was significantly lower compared to the busy CeBIT radio environment. Since most of the samples were below the detection threshold the distribution of the PSD values is more uniform after the threshold application. The distribution change increases the variance and compensates the described variance decrease due to the separation of the lowest samples.

[5]The measurement setups are based on portable spectrum analyzers since more sensitive instruments cannot be reasonably run autonomously on batteries. Therefore, very sensitive detection thresholds such as proposed by the IEEE 802.22 standardization committee could not be applied [49].

[6]The duty cycle (DC) is defined as the ratio of the number of detections to the number of all measured samples. Continuous signals will result in $DC = 1$ if the signal is clearly above the detection threshold. Other systems that apply, e.g., time division multiple access (TDMA), will usually result in lower DCs. A vacant channel is indicated by $DC = 0$.

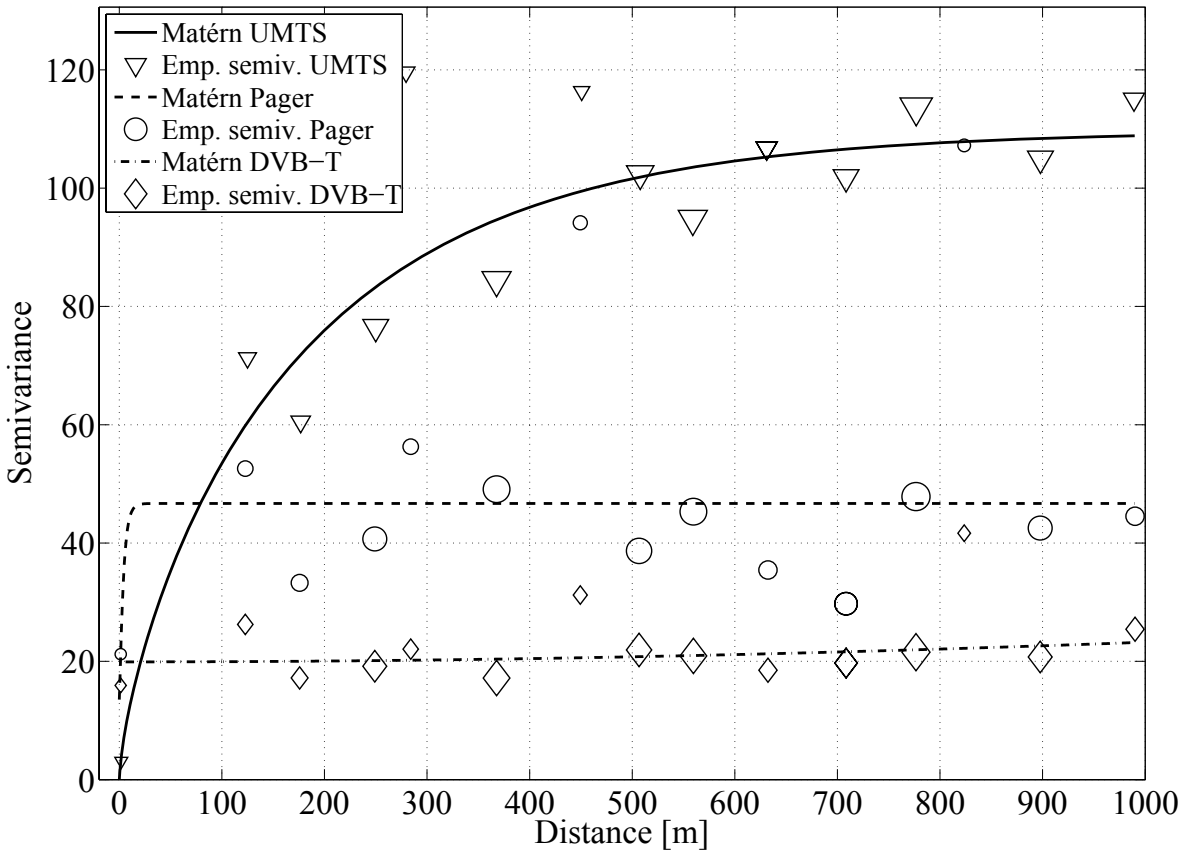

Figure 9: Fitted semivariograms based on the Matérn-model and estimated empirical semivariograms for the UMTS DL (5 MHz per channel, average over five channels at frequencies $f \in [2110, 2170]$ MHz), a pager-service (200 kHz bandwidth at $f = 466$ MHz), and DVB-T (6 MHz bandwidth per channel, average over three channels at $f_1 = 490$ MHz, $f_2 = 498$ MHz, and $f_3 = 514$ MHz).

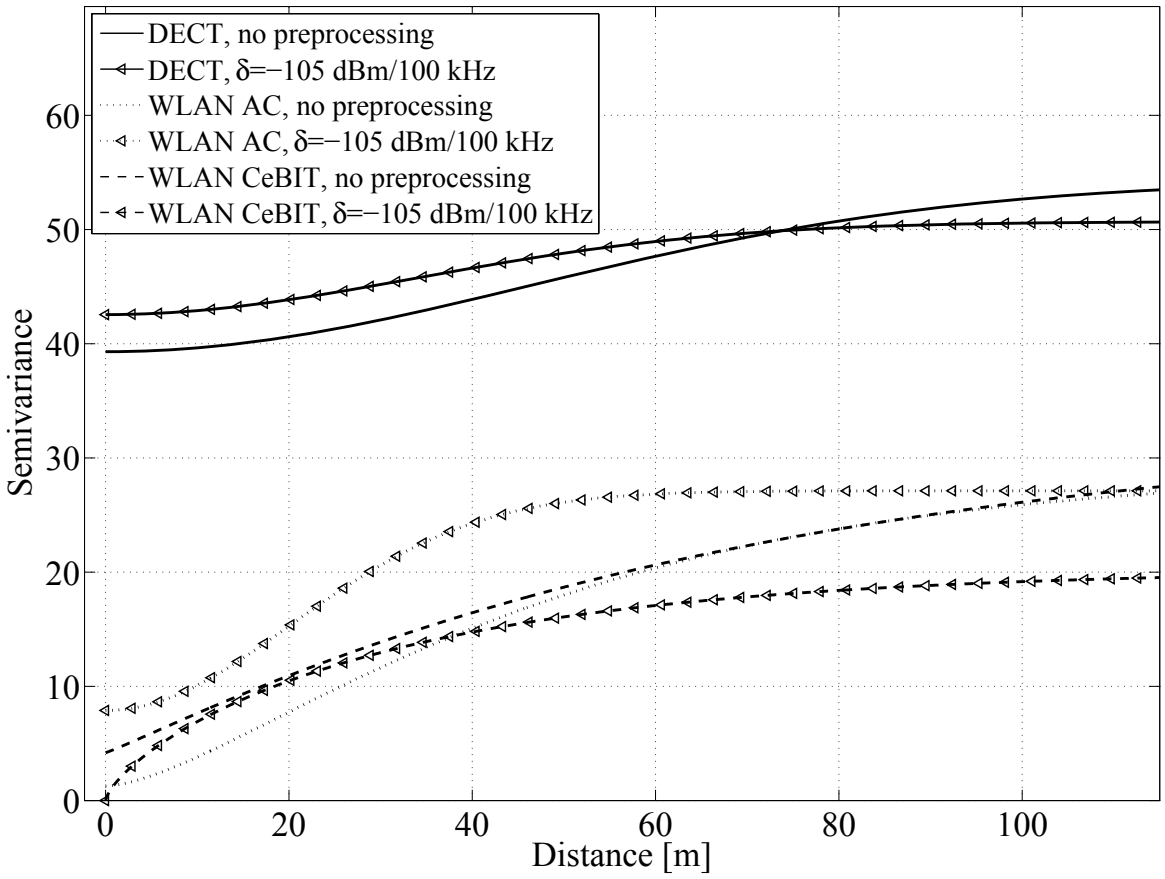

Figure 10: Fitted semivariograms based on the Matérn-model for the unchanged and the preprocessed input data sets for DECT networks ($\approx 1.7\,\text{MHz}$ bandwidth per channel, average over ten channels at frequencies $f \in [1880, 1900]\,\text{MHz}$) and the WLAN system (20 MHz bandwidth per channel, average over the most popular channels 1, 6, and 11 in the 2.4 GHz ISM-band).

The two WLAN data sets are also good examples how the deterministic model improves the understanding of differences in empirically measured results. We have shown in Sections 4.2 and 4.4 that both, higher path loss, modelled by the Hata model, and a lower node density, result in higher values for the sill. Both effects describe also differences between the Aachen and the CeBIT environments. The propagation environment in Aachen is more diverse due to the buildings and changing density of WLANs. At the CeBIT the examined exhibition halls were very well covered and we did not consider locations outside the exhibition halls. Additionally, also the density of WLANs was significantly higher in the CeBIT exhibition hall compared to the Aachen downtown area.

We continue with the discussion of the DECT results. Similarly to the CeBIT WLAN results the preprocessing lowers the sill because the noise samples have been sorted out. Sill and range are both higher for the DECT system compared to the previously discussed WLAN results. As for the mid-range systems, the different transmit power and working frequency are the major factors causing this difference. However, the range of 60 m is also realistic for the DECT case.

The very high value fitted to the nugget variance was a surprise. However, one can show that it is another measurement artefact due to the suboptimal sampling grid design. Although the grid step size of 15 m applied in the Aachen downtown campaign is appropriate for the DECT communication range more sampling points are required at very small distances. The WLSE considers the amount of sampling pairs in each distance bin but it still overweights the few outliers at close distances and fits a very high nugget variance.

5.4. Parameter sets for various wireless services

Finally, we list fitted model parameters for multiple wireless technologies in Table 1. We also include goodness of fit results and lower values are better due to the least squares based parameter estimation technique. We provide results for three different correlogram models. The exponential model gives sufficiently good fits and is interesting because three parameters are enough to fully describe it. The Matérn model results in on average the best goodness of fit and is often used to describe semivariograms. The third model is the Cauchy model because it is a good compromise between other models. On the one hand, it does not require theoretically less tractable terms such as the Γ- or the modified Bessel-function and on the other hand it still results in very good fits.

The interested researcher may choose himself which model fits his requirements best. In analytical studies the exponential model might be the best option but in simulation studies the additional complexity of a fourth model parameter might not cause any problems. We do not list model parameters for the DVB-T and the pager service because the fitting method has turned out not to provide reliable results. Instead, we focus on those systems for which the sampling grid was appropriate.

When having a closer look on the values of the model configurations few parameters are surprising. Sometimes very low numbers for the sill or the range are given. These cases usually correspond to situations with high nugget variances or high values for the fourth model parameter κ. In geostatistics the parameter κ is often limited to low values in the order of $\kappa \in [0.5, 3.5]$. Since few parameter sets outside these limits resulted in more accurate fits we did not apply similar limitations. Thus, although some parameter sets do not allow direct interpretation of the numbers they still provide an appropriate reproduction of the measurement results.

5.5. Nugget variance

In those cases where the best fit results in a nugget $\tau^2 > 0$ we also include parameter fitting results for the case of predefined $\tau^2 = 0$ into Table 1. Both parameter sets can be used to generate artificial scenarios for simulations or other purposes but the models without nugget variance may be preferable for theoretical studies due to their improved analytical tractability. In order to enable better selection of parameters for the use case at hand, we discuss the possible causes for nugget variance in the context of wireless communications in this section.

We start from a vector-valued random field that has multiple entries per location but all of them provide exactly the same value. This may happen if we investigate a time-invariant phenomena but collect multiple

Table 1: Fitted model parameters for selected wireless technologies.

Correlogram model	Wireless service	Measurement location	δ $[\frac{\text{dBm}}{100\,\text{kHz}}]$	$\overline{\mu_{ch}}$ $[\frac{\text{dBm}}{100\,\text{kHz}}]$	std(μ_{ch}) $[\frac{\text{dBm}}{100\,\text{kHz}}]$	Nugget τ^2	Sill σ^2	Range ϕ	κ	Quality of fit Q
Exponential	UMTS	Aachen	-	-87.83	8.0	2.0	105.27	161.14	-	19.71
Exponential	UMTS	Aachen	-	-87.83	8.0	0.0	104.82	129.81	-	25.05
Matérn	UMTS	Aachen	-	-87.83	8.0	0.0	109.73	229.58	0.34	18.47
Cauchy	UMTS	Aachen	-	-87.83	8.0	3.1	144.28	33.78	0.21	17.30
Cauchy	UMTS	Aachen	-	-87.83	8.0	0.0	144.28	5.01	0.13	22.73
Exponential	GSM900	Aachen	-105	-82.68	6.1	17.0	98.42	191.40	-	10.10
Exponential	GSM900	Aachen	-105	-82.68	6.1	0.0	106.50	9.35	-	31.70
Matérn	GSM900	Aachen	-105	-82.68	6.1	2.5	122.15	490.91	0.19	8.31
Matérn	GSM900	Aachen	-105	-82.68	6.1	0.0	124.49	508.78	0.18	8.33
Cauchy	GSM900	Aachen	-105	-82.68	6.1	15.1	124.49	50.63	0.29	9.53
Cauchy	GSM900	Aachen	-105	-82.68	6.1	0.0	122.99	1.98	0.18	25.40
Exponential	DECT	Aachen	-105	-91.55	1.5	37.7	13.52	34.08	-	6.87
Exponential	DECT	Aachen	-105	-91.55	1.5	0.0	49.45	9.00	-	8.08
Matérn	DECT	Aachen	-105	-91.55	1.5	42.5	8.13	3.42	50.00	6.70
Matérn	DECT	Aachen	-105	-91.55	1.5	0.0	51.60	56.57	0.07	6.92
Cauchy	DECT	Aachen	-105	-91.55	1.5	42.4	8.23	214.54	21.54	6.72
Cauchy	DECT	Aachen	-105	-91.55	1.5	0.0	52.95	1.99	0.38	7.04
Exponential	WLAN	Aachen	-105	-96.56	0.5	0.0	28.44	23.93	-	10.51
Matérn	WLAN	Aachen	-105	-96.56	0.5	7.9	19.24	2.05	50.00	8.45
Matérn	WLAN	Aachen	-105	-96.56	0.5	0.0	27.40	12.51	1.14	8.65
Cauchy	WLAN	Aachen	-105	-96.56	0.5	7.9	19.24	202.72	49.89	8.45
Cauchy	WLAN	Aachen	-105	-96.56	0.5	0.0	28.10	20.56	1.24	8.65
Exponential	WLAN	CeBIT	-105	-84.59	2.1	2.4	17.46	32.38	-	16.23
Exponential	WLAN	CeBIT	-105	-84.59	2.1	0.0	18.85	24.77	-	17.54
Matérn	WLAN	CeBIT	-105	-84.59	2.1	0.0	20.26	39.90	0.35	16.09
Cauchy	WLAN	CeBIT	-105	-84.59	2.1	4.7	20.63	15.02	0.32	16.11
Cauchy	WLAN	CeBIT	-105	-84.59	2.1	0.0	20.63	10.33	0.48	19.12

samples at each location consecutively. Next, we revisit the definition of the semivariogram as given in eq. (2). It is based on the expectation of the difference between samples taken at two locations, which will be zero when using the same location twice; $\gamma(\mathbf{u} \equiv \mathbf{v}, \mathbf{v}) = 0$.

However, this result does not necessarily imply that the underlying process does not exhibit a nugget variance. We have to differentiate location-independent and location-dependent random processes when describing the reasons for nugget variances. Both groups contribute to and build up the random residual $\epsilon(\mathbf{h})$ whose spatial statistics we are interested in.

First, location-independent random processes affect a whole area in the same way, with multiple measurement points located in this area. If we stick to the use case of PSD measurements over space examples for location-independent random processes will be the receiver and other background noise or the power control applied by, e.g., a UMTS base station.[7] Changes in transmit power affect all measurement locations in the coverage area. However, since power control is a time-dependent process it would be averaged out if enough samples over time were collected. Similarly, background noise processes can also be averaged out over time. Therefore, these processes may only cause a nugget variance if an insufficient number of samples are collected. For example, if the measurements at different locations are not performed in parallel and differ only due to power control effects, in the limit towards infinite number of samples these differences over time will be averaged out resulting in a nugget variance of zero. Alternatively, location-independent processes could also be modelled by considering a location-dependent mean in the underlying model for the random field as given in eq. (1) because whole areas are affected on a larger scale compared to location-dependent random processes that cause differences also on smaller scales.

In contrast, location-dependent effects cannot be averaged out. Typical examples for location-dependent random processes in wireless communications are different types of fading effects. Also deterministic system characteristics such as directive transmitter antennas will result in system inherent differences in the PSD measured at nearby locations. Therefore, we state the definition of the semivariogram more explicitly than initially given in eq. (4) as

$$\gamma(\mathbf{h}) = \begin{cases} 0, & \text{if } \|\mathbf{h}\| = 0, \\ \tau^2 + \sigma^2\big(1 - \rho_0(\mathbf{h}/\phi)\big), & \text{if } \|\mathbf{h}\| > 0, \text{with } \tau^2 \in \mathbb{R}_0^+ \text{ and } \sigma^2, \phi \in \mathbb{R}^+. \end{cases} \tag{12}$$

This definition allows for a discontinuity at $\|\mathbf{h}\| = 0$ since $\lim_{\|\mathbf{h}\| \to 0} \gamma(\mathbf{h}) \neq 0$ although $\gamma(\mathbf{h} = 0) = 0$.

For some data sets our model fitting results in low nugget variances which seem realistic. The significantly higher nugget variances in the case of the DECT system still result in very good fits and represent the measurement data well but seem to be too high and motivate further investigation possibly through additional measurements.

Our results discussed for the deterministic system model do not show any nugget variance because all the effects that may realistically cause a nugget variance in the use case of PSD measurements over space are not modelled explicitly, yet. For example, no multipath fading is considered and also hardware characteristics such as very directive transmitter antennas that may cause nugget variances are not taken into account.

5.6. Generation of spectrum scenarios

Simulation of random fields can be achieved, e.g., with the `RandomFields` package [50] for the `R` environment [44]. It supports a variety of correlogram models and is well documented. The listed parameters can be used to generate simulation scenarios based on Gaussian random fields. Additionally, the average PSD of a random field can easily be added to a previously generated zero-mean random field.

In order to enable the use of such more realistic PSD values we also list the average PSD for each service in Table 1. We computed these average values as follows: Several wireless systems employ wider channels

[7] UMTS power control is strictly speaking not a random process because the power control algorithms are deterministic over a known environment. However, we investigate a specific scenario chosen from a large ensemble of environments and interpret the environment selection process as being random. Therefore, also the overall process looks like a random process and we can treat the power control statistically as well.

than our resolution bandwidth of 100 kHz. In the first step, we compute μ_{ch} as average over all samples of the random field belonging to the same channel. This step involves averaging over frequency as well as over space. Afterwards, we also average over those channel-specific values resulting in the listed value $\overline{\mu_{ch}}$, again combining data in the frequency domain. The two-step averaging enables us to also list the standard deviation over different channels as denoted by std(μ_{ch}).

6. Discussion

We have shown how the spatial statistics of spectrum use can be assessed from a deterministic and an empirical viewpoint. In this section we add two further aspects. First, we bring both approaches together and discuss which insight we can get when taking into account both. Second, we elaborate on the use case of cooperative sensing as an example how we envision spatial statistics to be a useful tool for efficient operation of DSA networks.

Since the deterministic model does not fully consider all characteristics of a real network a detailed comparison between the simulated semivariograms and those computed from the measurement results is not reasonable. Nevertheless, certain aspects can be investigated and used to validate the underlying assumptions. Here, we focus specifically on the transmitter location distribution because the different considered point processes resulted in clearly differentiable shapes for the semivariogram.

The empirical semivariograms computed for the WLAN measurement data are most similar to the basic shape of the semivariogram simulated for the Poisson point process. In contrast, clustered point processes have been shown to describe WLAN systems very well in [22, 23]. However, both results are not contradictory when the details are taken into account. In [22, 23] the access point (AP) distribution over large areas of the United States of America (USA) has been considered and our measurements cover only parts of the Aachen downtown area or two CeBIT exhibition halls. Therefore, through our measurements we investigated, at maximum, a single cluster but more probably only the center of a single cluster.

In our simulations we used the Thomas point process to generate clustered network topologies. Inside each cluster the node distribution follows a Gaussian probability distribution function (PDF) over the distance with the maximum located at the center of the cluster. If we assume that the whole downtown Aachen area is described by a single cluster the standard deviation of the corresponding Gaussian PDF will be rather high and the curve will have a rather flat shape. Since our WLAN measurements were based on the smaller grip step size we covered only the core part of Aachen downtown, which is, seen from the theoretical modelling viewpoint, the area in which the Gaussian PDF has a rather constant value and thus behaves very similarly to a constant PDF as in the case of the Poisson point process. Therefore, the measured semivariogram describes in more detail the spatial statistics of the internal topology of a single cluster and confirms that there is no specific structure that should be modelled in greater detail than achieved by the Thomas process.

One may expect that a hard core process could provide further enhancements because it limits the probability that several APs are located very close by. However, our sensitivity analysis showed that hard core processes result in very similar results for the semivariogram as determined for the Poisson case. Adding one more parameter does not seem to be required. We have seen that the area covered by the small grid in Aachen downtown is small enough not to exhibit node density differences and to enable realistic modelling through a Poisson point process. In contrast, the area of a complete city should be modelled by a cluster point process because more industrial and residential areas obviously have different characteristics.

The CeBIT data also results in a semivariogram shape that is similar to the Poisson case. Again, this is an expected result when considering where APs have been deployed in the exhibition hall. Several exhibitors have put up their own APs in order to provide wireless network access throughout their booth. Additionally, further permanent APs have been installed on the walls of the exhibition halls. The latter APs have prevented the AP density from decreasing towards the borders of the exhibition areas as one may have expected. Thus, the uniform probability distribution as indicated by the Poisson point process can also be well explained in the CeBIT scenario. In the examined exhibition hall scenario a hardcore process may be a more intuitive model since no multi-story buildings have to be considered and multiple APs will rarely

be projected to the same position on a two-dimensional plane. However, as we have mentioned above, the results for hardcore processes are very similar to those based on the Poisson process.

In the next step, we comment on a selected application of spatial statistics of spectrum use. In particular, we discuss how this kind of information can be used to enhance cooperative sensing in DSA networks. It has been shown through theoretical analysis as well as practical measurements that correlated spectrum measurements lower the achievable cooperation gain [9, 10, 11, 12, 13, 14, 15]. In [13] we have also shown that such correlations can be reliably accessed through measurements during system runtime. Additionally, Selén *et al.* have described in [51] efficient algorithms for selection of other users to cooperate with under consideration of the correlation characteristics. However, the detailed correlation properties depend on the radio environment, the frequency band in use, etc. and runtime adaptations of the cooperative sensing system are certainly a promising approach.

In this context, we face the trade-off between the amount of information that is available at a node for optimization of the cooperative sensing process and the involved signalling overhead for exchange of such measurement results. A detailed correlation analysis including the tracking of changes in the correlation characteristics over time may even require time-synchronized measurements to be performed by the cooperative nodes. Additionally, the processing of a significant amount of measurement data requires a node with high computational power which often requires the introduction of a central basestation-like node to the system architecture. An approach based on less complex computation could also be implemented in a distributed fashion using appropriate load-balancing between the cooperating nodes.

The semivariogram-based modelling requires only a subset of the measurement results, namely the average PSD over frequency as measured by each spectrum sensor. These data can be collected in an online manner and the task can be carried out by diverse types of nodes ranging from efficient specialized sensors in a sensor network via end user devices to highly flexible and powerful infrastructure nodes. The additional load for the participating stations is manageable since only averaged PSD values have to be computed. The data of all cooperative nodes have to be fused to a single node for estimation of the semivariogram. Afterwards, the parameters and the correlogram type can be sent back to all nodes. Although the transfer of these parameters causes very little overhead, the semivariogram provides information on several key properties of the radio environment, e.g., its diversity through the sill, σ^2, or the correlation through the shape of the semivariogram and the range, ϕ. Now, the participating nodes can optimize the cooperative sensing process using distributed algorithms because all have the same view of the radio environment augmented by detailed local information. For example, the sets of nodes that cooperate can be selected more efficiently. Either the reliability of the sensing system is increased or the overhead is reduced by distributing the sensing task across space, frequency, and/or time. First, we can lower the spatial resolution and save the resources spent for sensing at some nodes. Second, not all nodes have to sense the whole frequency band of interest and, third, we can exploit the correlation by lowering the sensing rate at selected nodes. Additionally, system-wide adaptations can be implemented in a more synchronized fashion. An example is the adjustment of the sensing rate due to changes in the primary user activity as proposed in [26]. The reduced activity could either be detected by a decrease in the average PSD over the whole random field or a lower sill caused by a smaller variation in the primary user system.

A node joining the network would contact the node currently in charge of computing the semivariogram and retrieve the valid model parameters. Afterwards, the node can estimate the minimum distance to possible nodes worth cooperating with in order to initiate efficient cooperative sensing. If multiple nodes are close by it may also decide not to sense and rely on the results gathered by the other nodes. How the measurement data collection and semivariogram parameter distribution is organized in this example scenario is independent of the applicability of the semivariogram-based modelling. The data may also be saved in a distributed data structure such as a radio environment map [34, 35].

Another application for the semivariogram models is the prediction of spectrum occupancy in regions at which no direct measurements are available. In the geostatistics literature procedures of this kind are called different variants of *kriging*, most common example being the use of best linear unbiased estimator for the random field based on measurements and the semivariogram model known as *ordinary kriging*. Naturally the prediction error will depend heavily on the correlation structure of the mean PSD field as well as number and density of the measurement points compared to the range of the variogram. For a thorough discussion

on these issues we refer the reader to [20].

Semivariogram-based spectrum modelling can also be applied in the non-cooperative sensing case. A single mobile node could enhance the efficiency of its sensing process when taking advantage of the spectrum model. Due to the movement of the node the spectral environment around the node will constantly change. The semivariogram provides means to estimate the probability that a spectrum channel, which has previously been sensed occupied, is vacant at the new location. If the semivariogram indicates a high range for the spectral band under test the sampling frequency of the sensing process should be reduced in order to avoid sensing steps with low probability of finding a vacant spectrum band.

7. Conclusion

In this paper, we have introduced the methodology of spatial statistics to describe radio environments and in particular the spectrum use in wireless networks. We have shown that the power spectral density (PSD) measured over space can be interpreted as a random field and its properties can be analyzed in detail using powerful techniques such as the semivariogram modelling originally developed in the geostatistics community.

We have discussed two approaches for scenario evaluation. First, the deterministic approach is simulation-based and enables the investigation of individual impact factors. We have shown that the transmitter location distribution modelled through a point process determines the shape of the semivariogram. The propagation model has high impact on the variance of the random field, the transmit power has negligible impact in the case of systems with good network coverage, and the node density describes the slope of the semivariogram and has also impact on the variance of the random field. Second, the empirical approach is based on real-life measurements of the PSD over space. It provides a realistic description of the current spectrum use and enables the evaluation of various wireless systems without the need to model their properties explicitly. We have provided model parameters of multiple wireless systems for generation of synthetic simulation scenarios and further analytical work. The combination of both approaches has shown that differences in empirical models can be well explained using the results of the deterministic evaluation.

Applications of spatial statistics in the area of wireless communication and in particular in the dynamic spectrum access (DSA) context are manifold. The semivariogram enables cognitive radios (CRs) to build up a detailed understanding of their surrounding radio environment. The model parameters can be computed with manageable overhead and efficiently shared within the network. A common view on the radio environment enables synchronized network improvements and parameter adaptations in a distributed as well as centralized fashion. For example, cooperative sensing systems could be enhanced by selecting less correlated sensors for cooperation or by optimizing sensing parameters such as the sensing rate. Additionally, spatial models can also be used on a more macroscopic scale as basis for discussion and further development of spectrum regulation and policies in a more quantitative manner.

In our future work, we plan to investigate the time-aspect of spectrum modelling. We will examine the required update rate for the semivariogram parameters and work on methods that reduce the involved overhead. Possible approaches fuse only differential information instead of complete sets of average PSD samples. Additionally, we will study the limit how close we can push the cooperation gain of a cooperative sensing system based on semivariogram-based modelling towards the optimum, which would require perfect knowledge on the correlation between single sensor pairs. In this context, we plan to extend our measurement setup, evaluate more appropriate measurement grids, and further improve our models. In parallel, we will investigate the behaviour of primary users over longer time spans. Finally, we plan to work on the integration of time and spatial modelling to a full space–time model for spectrum use.

Acknowledgments

The authors would like to thank RWTH Aachen University and the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) for providing financial support through UMIC research center. We would also like to thank the European Union for providing partial funding of this work through the

ARAGORN project. Finally, we would like to thank the Deutsche Messe AG for providing us access to our measurement location at the CeBIT in Hanover and the kind support throughout the campaign.

References

[1] M. A. McHenry, NSF Spectrum Occupancy Measurements Project Summary, Tech. rep., Shared Spectrum Company (August 2005).
[2] S. W. Ellingson, Spectral occupancy at VHF: implications for frequency-agile cognitive radios, in: Proc. of IEEE Vehicular Technology Conference (VTC), Dallas, TX, USA, 2005, pp. 1379–1382.
[3] R. I. C. Chiang, G. B. Rowe, K. W. Sowerby, A Quantitative Analysis of Spectral Occupancy Measurements for Cognitive Radio, in: Proc. of IEEE Vehicular Technology Conference (VTC), Dublin, Ireland, 2007, pp. 3016–3020.
[4] M. Wellens, J. Wu, P. Mähönen, Evaluation of spectrum occupancy in indoor and outdoor scenario in the context of cognitive radio, in: Proc. of IEEE Second International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM), Orlando, FL, USA, 2007, pp. 420–427.
[5] M. Islam, C. Koh, S. Oh, X. Qing, Y. Lai, C. Wang, Y.-C. Liang, B. Toh, F. Chin, G. Tan, W. Toh, Spectrum Survey in Singapore: Occupancy Measurements and Analyses, in: Proc. of International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM), Singapore, 2008.
[6] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, S. Mohanty, Next generation/dynamic spectrum access/cognitive radio wireless networks: a survey, Elsevier Computer Networks Journal 50 (13) (2006) 2127–2159.
[7] Q. Zhao, B. Sadler, A Survey of Dynamic Spectrum Access, IEEE Signal Processing Magazine 24 (3) (2007) 79–89.
[8] R. Brodersen, A. Wolisz, D. Cabric, S. Mishra, D. Willkomm, White Paper: CORVUS - A Cognitive Radio Approach for Usage of Virtual Unlicensed Spectrum, Tech. rep., University of California, Berkeley, available online http://bwrc.eecs.berkeley.edu/Research/MCMA/ CR_White_paper_final1.pdf [Cited on 16th September 2008] (2004).
[9] E. Visotsky, S. Kuffner, R. Peterson, On Collaborative Detection of TV Transmissions in Support of Dynamic Spectrum Sharing, in: Proc. of IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN), Baltimore, MD, USA, 2005, pp. 338–345.
[10] A. Ghasemi, E. S. Sousa, Collaborative Spectrum Sensing for Opportunistic Access in Fading Environments, in: Proc. of IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN), Baltimore, MD, USA, 2005, pp. 131–136.
[11] S. M. Mishra, A. Sahai, R. W. Brodersen, Cooperative Sensing among Cognitive Radios, in: Proc. of IEEE International Conference on Communications (ICC), Vol. 4, Istanbul, Turkey, 2006, pp. 1658–1663.
[12] A. Ghasemi, E. S. Sousa, Asymptotic performance of collaborative spectrum sensing under correlated log-normal shadowing, IEEE Communications Letters 11 (1) (2007) 34–36.
[13] M. Wellens, J. Riihijärvi, M. Gordziel, P. Mähönen, Evaluation of Cooperative Spectrum Sensing based on Large Scale Measurements, in: Proc. of IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN), Chicago, IL, USA, 2008.
[14] D. Cabric, A. Tkachenko, R. W. Brodersen, Experimental Study of Spectrum Sensing based on Energy Detection and Network Cooperation, in: Proc. of Workshop on Technology and Policy for Accessing Spectrum (TAPAS), Boston, MA, USA, 2006.
[15] D. Cabric, A. Tkachenko, R. W. Brodersen, Spectrum sensing measurements of pilot, energy, and collaborative detection, in: Proc. of IEEE Military Communications Conference (MILCOM), Washington, D.C., USA, 2006.
[16] M. Gudmundson, Correlation model for shadow fading in mobile radio systems, Electronic Letters 27 (23) (1991) 2145–2146.
[17] K. Sawa, E. Kudoh, F. Adachi, Impact of shadowing correlation on spectrum efficiency of a power controlled cellular system, IEICE Transactions on Communications 87 (7) (2004) 1964–1969.
[18] J. C. Liberti, T. S. Rappaport, Statistics of shadowing in indoor radio channels at 900 and 1900 MHz, in: Proc. of IEEE Military Communications Conference (MILCOM), Vol. 3, San Diego, CA, USA, 1992, pp. 1066–1070.
[19] F. Graziosi, F. Santucci, A general correlation model for shadow fading in mobile radio systems, IEEE Communications Letters 6 (3) (2002) 102–104.
[20] N. Cressie, Statistics for Spatial Data, revised Edition, Wiley, 1993.
[21] A. Jindal, K. Psounis, Modeling spatially correlated data in sensor networks, ACM Transactions on Sensor Networks (TOSN) 2 (4) (2006) 466–499.
[22] J. Riihijärvi, P. Mähönen, M. Rübsamen, Characterizing Wireless Networks by Spatial Correlations, IEEE Communications Letters 11 (1) (2007) 37–39.
[23] P. Mähönen, M. Petrova, J. Riihijärvi, Applications of Topology Information for Cognitive Radios and Networks, in: Proc. of IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN), Dublin, Ireland, 2007, pp. 103–114.
[24] J. Riihijärvi, P. Mähönen, Exploiting Spatial Statistics of Primary and Secondary Users towards Improved Cognitive Radio Networks, in: Proc. of International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM), Singapore, 2008.
[25] J. Riihijärvi, P. Mähönen, M. Wellens, M. Gordziel, Characterization and Modelling of Spectrum for Dynamic Spectrum Access with Spatial Statistics and Random Fields, in: Proc. of 1st International Workshop on Cognitive Radios and Networks (CRNETS), in conjunction with IEEE PIMRC 2008, Cannes, France, 2008.

[26] D. Willkomm, S. Machiraju, J. Bolot, A. Wolisz, Primary Users in Cellular Networks: A Large-scale Measurement Study, in: Proc. of IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN), Chicago, IL, USA, 2008.
[27] M. Wellens, J. Riihijärvi, M. Gordziel, P. Mähönen, Spatial Statistics of Spectrum Usage: From Measurements to Spectrum Models, in: Proc. of IEEE International Conference on Communications (ICC), Dresden, Germany, 2009.
[28] J. Riihijärvi, P. Mähönen, S. Sajjad, Influence of Transmitter Configurations on Spatial Statistics of Radio Environment Maps, in: Proc. of IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), Tokyo, Japan, 2009.
[29] B. L. Mark, A. O. Nasif, Estimation of Interference-Free Transmit Power for Opportunistic Spectrum Access, in: Proc. of IEEE Wireless Communications and Networking Conference (WCNC), Las Vegas, NV, USA, 2008, pp. 1679–1684.
[30] B. L. Mark, A. O. Nasif, Estimation of Maximum Interference-Free Transmit Power Level for Opportunistic Spectrum Access, Tech. rep., George Mason University, Department of Electrical and Computer Engineering, Network Architecture and Performance Laboratory (NAPL) (August 2008).
[31] A. O. Nasif, B. L. Mark, Collaborative Opportunistic Spectrum Sharing in the Presence of Multiple Transmitters, in: Proc. of Global Telecommunications Conference (GLOBECOM), New Orleans, LA, USA, 2008.
[32] R. Tandra, S. M. Mishra, A. Sahai, What is a spectrum hole and what does it take to recognize one?, accepted for publication in Proceedings of the IEEE.
[33] J. Mitola III, Cognitive radio: an integrated agent architecture for software defined radio, Ph.D. Thesis, KTH (Royal Institute of Technology), 2000.
[34] Y. Zhao, L. Morales, J. Gaeddert, K. K. Bae, J.-S. Um, J. H. Reed, Applying Radio Environment Maps to Cognitive Wireless Regional Area Networks, in: Proc. of IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN), Dublin, Ireland, 2007, pp. 115–118.
[35] Y. Zhao, B. Le, J. H. Reed, Cognitive Radio Technology, Elsevier, 2006, Ch. Network Support – The Radio Environment Map, pp. 337–363.
[36] B. D. Ripley, Spatial Statistics, Wiley, 2004.
[37] M. D. Ecker, A. E. Gelfand, Bayesian Variogram Modeling for an Isotropic Spatial Process, Journal of Agricultural, Biological and Environmental Statistics 2 (4) (1997) 347–369.
[38] D. Stoyan, W. S. Kendall, J. Mecke, Stochastic Geometry and its Applications, 2nd Edition, John Wiley & Sons, 1995.
[39] R. M. Lark, Optimized spatial sampling of soil for estimation of the variogram by maximum likelihood, Geoderma 105 (1–2) (2002) 49–80.
[40] N. Cressie, Fitting Variogram Models by Weighted Least Squares, Mathematical Geology 17 (5) (1985) 563–586.
[41] P. J. Diggle, P. J. Ribeiro, O. F. Christensen, Spatial statistics and computational methods, Springer Verlag, Lecture Notes in Statistics, vol. 173, 2003, Ch. An Introduction to Model-based Geostatistics.
[42] G. Matheron, Principles of geostatistics, Economic Geology 58 (8) (1963) 1246–1266.
[43] A. Baddeley, R. Turner, Spatstat: an R package for analyzing spatial point patterns, Journal of Statistical Software 12 (6) (2005) 1–42.
[44] R Development Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria (2007).
[45] AWE Communications, WinProp Software Suite, `http://www.awe-communications.com` [Cited on 18th September 2008] (2007).
[46] R. Timo, K. Blackmore, L. Hanlen, Strong Stochastic Stability for MANET Mobility Models, in: Proc. of IEEE International Conference on Networks (ICON), Adelaide, Australia, 2007, pp. 13–18.
[47] M. Hata, Empirical Formula for Propagation Loss in Land Mobile Radio Services, IEEE Transactions on Vehicular Technology 29 (3) (1980) 317–325.
[48] Y. Okumura, E. Ohmori, T. Kawano, K. Fukuda, Field Strength and its Variability in VHF and UHF Land-Mobile Radio Service, Review Electrical Communication Laboratory 16 (9–10) (1968) 825–873.
[49] S. Shellhammer, G. Chouinard, Spectrum sensing requirements summary, IEEE 802.22-05/22-06-0089-05-0000 (July 2006).
[50] M. Schlather, Simulation of stationary and isotropic random fields, R News 1 (2) (2001) 18–20.
[51] Y. Selén, H. Tullberg, J. Kronander, Sensor Selection for Cooperative Spectrum Sensing, in: Proc. of IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN), Chicago, IL, USA, 2008.