Noname manuscript No.
(will be inserted by the editor)

Building Accurate Radio Environment Maps from Multi-Fidelity Spectrum Sensing Data

Selvakumar Ulaganathan · Dirk Deschrijver · Mostafa Pakparvar · Ivo Couckuyt · Wei Liu · David Plets · Wout Joseph · Tom Dhaene · Luc Martens · Ingrid Moerman

Received: date / Accepted: date

Abstract In cognitive wireless networks, active monitoring of the wireless environment is often performed through advanced spectrum sensing and network sniffing. This leads to a set of spatially distributed measurements which are collected from different sensing devices. Nowadays, several interpolation methods (e.g. Kriging) are available and can be used to combine these measurements into a single globally accurate radio environment map that covers a certain geographical area. However, the calibration of multi-fidelity measurements from heterogeneous sensing devices, and the integration into a map is a challenging problem. In this paper, the auto-regressive co-Kriging model is proposed as a novel solution. The algorithm is applied to model measurements which are collected in a heterogeneous wireless testbed environment, and the effectiveness of the new methodology is validated.

Keywords Radio environment maps · Wireless experimentation · Kriging · Multi-fidelity modeling

1 Introduction

A reliable connectivity for wireless services that have stringent QoS requirements is often compromised by the saturation of the wireless radio spectrum. Due to the uncoordinated use of the unlicensed frequency bands (2.4 GHz and 5 GHz) by various wireless technologies and devices, the adverse effects of interference are becoming increasingly important and can no longer be neglected.

S. Ulaganathan, D. Deschrijver, M. Pakparvar, I. Couckuyt, W. Liu, D. Plets, W. Joseph, T. Dhaene, L. Martens, I. Moerman
Ghent University - iMinds, Department of Information Technology,
Gaston Crommenlaan 8 (Bus 201), B-9050, Ghent, Belgium
Tel.: +32-93-314940
Fax: +32-93-314899
E-mail: selvakumar.ulaganathan@ugent.be

Over the past years, cognitive radio techniques have been developed to ensure efficient interoperability of heterogeneous systems through advanced monitoring of the wireless environment, and the optimization of network configurations through cognitive decision making [1]. To this end, Radio Environment Maps (REMs) contain a lot of information as they represent an integrated database that provides real-time information concerning, e.g., spectrum availability, regulations or policies, and the degree of channel utilization [2]. In terms of spectrum utilization, REMs have been proposed to measure power spectral density (PSD) in order to determine the degree of spectrum utilization in a certain geographical area. These models are typically calculated from a set of distributed measurements and spatial interpolation techniques are applied to build an approximation model that estimates the corresponding values at arbitrary spatial coordinates. Several algorithms to calculate a REM have been studied previously, such as splines, Kriging, probabilistic models, Shepard's interpolation and Inverse Distance Weighting [3]. Kriging is reported to achieve a good performance in terms of overall prediction accuracy and generality [4]. In addition to PSD maps, other authors present so-called channel gain maps, that capture information about the propagation medium [5]. In [6], the use of medium utilization is proposed as a metric to be included in the REMs dedicated for wireless LANs. In [7], REMs have been used for intelligently guiding spectrum access for deployment of a prototype of a Long Term Evolution (LTE) system that opportunistically exploits the spectral white spaces in the upper Ultra High Frequency (UHF) TV bands.

This paper will focus on building REMs for measurements that originate from heterogeneous spectrum sensing devices on a wireless testbed. It is noted that each type of device may have its own processing and hardware capabilities in terms of detection mechanisms, sweeping time and data accuracy [8]. Also in [9], it is mentioned that the calibration of such measurements and integration into a REM involves a lot of challenges. The different sensing devices that will be considered in this work are subdivided into two categories. The USRP and IMEC sensing engines are considered to be high-fidelity (HF) devices which come at a higher production cost, but offer superior energy detection capabilities leading to very accurate measurements. The low-fidelity (LF) sensing devices such as Wi-Fi are considered to be low-cost and provide less accurate measurements because their sensing solutions are not as advanced. In order to combine these different types of data into a single globally accurate REM, the use of the autoregressive co-Kriging model [10] is presented as a novel approach. First, the densely sampled LF data is used to determine a trend function which is then corrected by the sparsely sampled HF data. An independent data set, based on additional measurements, is then used to validate the accuracy of the model and to assess how well its predictions cover the overall environment.

The paper is structured as follows: in Section 2, a spatial modeling algorithm is presented which explains how a single model can be calculated from measurement data with different levels of fidelity. In Section 3, a brief description of the w-iLab.t testbed is given - a large scale wireless testbed where

all the measurements are performed. In Section 4, more details are provided about the different spectrum sensing devices installed in the testbed. In Section 5, a practical application example of the novel technique is considered and the approach is validated experimentally. Finally, conclusions are provided in Section 6.

2 Spatial Modeling Algorithms

In order to build spatially-interpolated REMs, several mathematical algorithms can be considered [3] to model the measurement data. This paper will focus on a Kriging-based approach, which it is reported to achieve good performance in terms of prediction accuracy and generality [4]. First, a brief recap of the Kriging algorithm is presented in Section 2.1. Then, the use of the co-Kriging algorithm will be advocated in Section 2.2 as a novel approach to build REMs by combining data from sensing devices having different levels of fidelity.

2.1 Kriging Interpolation

A well-known technique in surrogate modeling is Kriging [11,12]. Kriging surrogate models are also known as Gaussian Processes (GP) [13] or Gaussian Random Fields [14]. Originally proposed by Krige [15], Kriging was popularized for the Design and Analysis of Computer Experiments (DACE) by Sacks et al. [16], where it has proven to be very useful for tasks such as optimization [17,18], design space exploration, visualization, prototyping, and sensitivity analysis [19,20]. For a full survey of Kriging the reader is referred to [12] and [13]. In this section a summary is given of the most important aspects of Kriging, and a brief explanation is given on how to build a REM.

Let $X = (\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^n)$ be a base set of n spatial coordinates and $f_e(X)$ the associated expensive measurements (HF). Kriging first fits a constant regression function on the data and, subsequently, constructs a GP through the residuals. The idea is that the regression function captures the largest variance in the data, while the GP takes care of the finer details and the final interpolation. This is reflected in the Kriging interpolant which is derived as,

$$\hat{f}(\mathbf{x}) = \alpha + r(\mathbf{x}) \cdot \Psi^{-1} \cdot (f_e(X) - \mathbf{1}\alpha), \tag{1}$$

where $\mathbf{1}$ is a column vector of ones. The coefficients of the regression function, i.e., the vector α, are determined by generalized least squares,

$$\alpha = (X'\Psi^{-1}X)^{-1}X\Psi^{-1}f_e(X). \tag{2}$$

$r(\mathbf{x}) = \begin{pmatrix}\psi(\mathbf{x},\mathbf{x}^1) \dots \psi(\mathbf{x},\mathbf{x}^n)\end{pmatrix}$ is an $1 \times n$ vector of correlations between the point $\mathbf{x}$ and the base set X, and Ψ is a $n \times n$ correlation matrix given by,

$$\Psi = \begin{pmatrix}\psi(\mathbf{x}^1,\mathbf{x}^1) & \dots & \psi(\mathbf{x}^1,\mathbf{x}^n) \\ \vdots & \ddots & \vdots \\ \psi(\mathbf{x}^n,\mathbf{x}^1) & \dots & \psi(\mathbf{x}^n,\mathbf{x}^n)\end{pmatrix} \tag{3}$$

Moreover, Kriging also predicts the approximation error (prediction variance) at each location in the environment, also where no measurements have been performed. The approximation error is zero in the data points themselves, as Kriging interpolates all data.

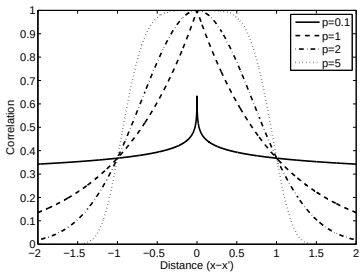
The regression function actually operates as the mean of the GP: predictions too far from existing measurement points (e.g., outside the sampled region) will revert to the mean (=regression function). As the behavior of the response is usually unknown, a popular choice that works well is the constant regression function as explained further on. However, in this case Kriging is purely an interpolation technique (in contrast to extrapolation). By using prior knowledge or other techniques (e.g., blind Kriging [21]) one may identify basis functions (linear, quadratic, etc.) to use in the regression function, enabling Kriging to extrapolate outside the sampled environment. This is especially useful for problems with missing data, i.e., in cases where there exist large gaps in the environment because no sensing devices are installed or available.

The choice of correlation function is crucial to create an accurate Kriging surrogate model. The popular (squared) exponential correlation functions are defined by $\psi(\mathbf{x},\mathbf{x}') = \exp(-\sum_{i=1}^{d}\theta_i|x_i - x_i'|^p)$. These correlation functions are called stationary because the correlation function only depends on the distance between the two points $\mathbf{x}$ and $\mathbf{x}'$. The smaller the distance between two points, the higher the correlation and, hence, the more the prediction of one point is influenced by the other. Similarly, if the distance increases the correlation drops to zero.

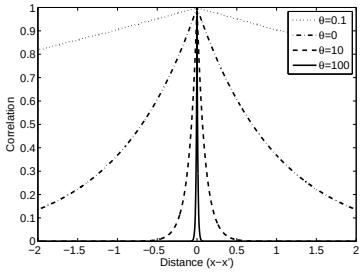
The rate and the manner at which this happens are governed by several parameters. In essence, the parameter p determines the 'smoothness' of the prediction, see Figure 1a. A value of $p = 2$ leads to a smooth prediction, but also has strict smoothness requirements on the response $f_e(\mathbf{x})$. With a smaller value of p the correlation decreases much faster as the two points move farther from each other, which is suitable for more sharp (discontinuous) responses.

Often, the parameter p is set to two, also known as the Gaussian correlation function, which is suitable for many problems. However, the Matérn class of correlation functions is a more realistic choice for real-life problems [22] and is considered in this work over the commonly used Gaussian correlation function. The Matérn class of correlations functions is observed to model rough surfaces more accurately than the Gaussian correlation function on various occasions. Two instances of the Matérn correlation functions are defined by,

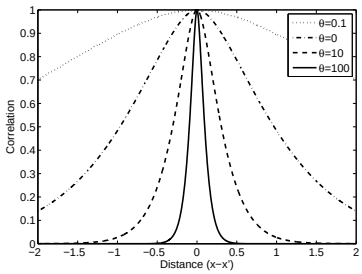
$$\psi(\mathbf{x},\mathbf{x}')_{\nu=\frac{3}{2}} = \left(1+\sqrt{3}l\right)\exp\left(-\sqrt{3}l\right), \tag{4}$$



(a)



(b)



(c)

Fig. 1: Examples of one-dimensional correlation functions: a) the exponential correlation function with varying parameter p for $\theta = 0.1$ b) the exponential correlation function with varying parameter θ for $p = 1$ c) the Matérn correlation function with varying parameter θ for $\nu = \frac{3}{2}$.

$$\psi(\mathbf{x}, \mathbf{x}')_{\nu=\frac{5}{2}} = \left(1 + \sqrt{5}l + \frac{5l^2}{3}\right) \exp\left(-\sqrt{5}l\right), \tag{5}$$

with $l = \sqrt{\sum_{i=1}^{d} \theta_i (x_i - x_i')^2}$. The parameter ν of the Matérn correlation functions has a similar role as the p parameter. Usually both parameters are fixed and in this work ν is set to $\frac{3}{2}$.

The second set of parameters, $\left(\theta_1 \ldots \theta_d\right)$, describes the influence sphere of a point on nearby points for each dimension, see Figure 1b and 1c. This is useful as it describes the linearity of the response and, hence, can be used

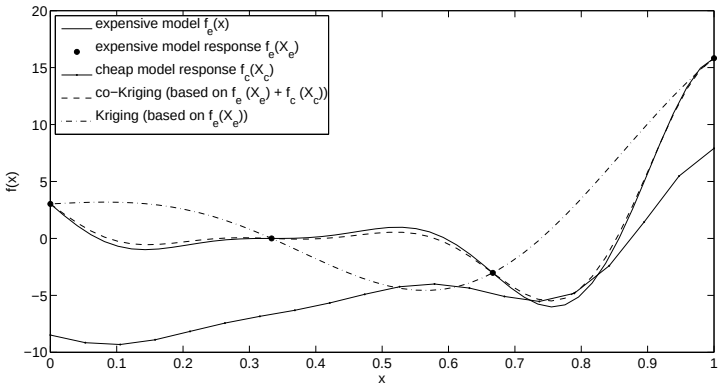


Fig. 2: Kriging and co-Kriging applied to a 1-dimensional mathematical example $f(x)$ having a normalized input coordinate x. Co-Kriging interpolates the expensive model response and is further corrected by the cheap model response.

to identify relevant variables. The parameters $\left(\theta_1 \ldots \theta_d\right)$ are identified using Maximum Likelihood Estimation (MLE). In particular, we minimize the negative concentrated log-likelihood,

$$-\ln(\mathcal{L}) \cong -\frac{n}{2}\ln(\hat{\sigma}^2) - \frac{1}{2}\ln(|\Psi|), \tag{6}$$

where $\hat{\sigma}^2 = (f_e(X) - \mathbf{1}\alpha)'\Psi^{-1}(f_e(X) - \mathbf{1}\alpha)/n$. Note that the application of Kriging is primarily limited by the size of the dataset. The number of samples has a direct impact on the correlation matrix which grows quickly as the number of samples increases. As the inverse of the correlation matrix needs to be computed many times during the MLE, Kriging is typically used for datasets with less than 1000 samples. Because the number of sensing devices in a testbed is usually limited (< 100 samples), the computation time is merely a matter of seconds which makes it a suitable aid for visualization and real-time decision making.

2.2 co-Kriging Interpolation

The popularity of Kriging has generated a large body of research, including several extensions to Kriging to handle different problem settings, e.g., by adding gradient information in the prediction [23], or by approximating stochastic simulations [24], etc. Co-Kriging is a natural multi-response extension to Kriging and allows to incorporate both expensive (i.e. high-fidelity, HF) and cheap (i.e. low-fidelity, LF) measurements from heterogeneous sensing devices in order to build accurate REMs [25]. In this paper the autoregressive co-Kriging model of Kennedy et al. [10] is adopted.

Creating a co-Kriging model can be interpreted as constructing two Kriging models in sequence. First a Kriging model $\hat{f}_c$ of the cheap data $(X_c, f_c(X_c))$ is constructed. Subsequently, the second Kriging model $\hat{f}_r$ is constructed on the residuals of the expensive and cheap data (X_e, f_r), where $f_r = f_e(X_e) - \rho \cdot f_c(X_e)$. The parameter ρ is included in the MLE of the second Kriging model. If the response values $f_c(X_e)$ are not available, they can be approximated by the first Kriging model $\hat{f}_c$, namely, $f_c(X_e) \approx \hat{f}_c(X_e)$.

Note that the configuration (choice of correlation function, regression function, etc.) of both Kriging models can be adjusted separately for the cheap data and the residuals, respectively.

The final co-Kriging model is built upon the two Kriging models. Namely, the co-Kriging interpolant is defined similarly as (1),

$$\hat{f}(\mathbf{x}) = \dot{M}\alpha + \dot{r}(\mathbf{x}) \cdot \dot{\Psi}^{-1} \cdot (f_r - \dot{F}\alpha), \tag{7}$$

where the block matrices $\dot{M}$, $\dot{F}$, $\dot{r}(\mathbf{x})$ and $\dot{\Psi}$ can be written as a function of the two underlying Kriging models $\hat{f}_c$ and $\hat{f}_r$:

$$\dot{r}(\mathbf{x}) = \left[\rho \cdot \hat{\sigma}_c^2 \cdot r_c(\mathbf{x}),\, \rho^2 \cdot \hat{\sigma}_c^2 \cdot r_c(\mathbf{x}, X_e) + \hat{\sigma}_r^2 \cdot r_r(\mathbf{x})\right], \tag{8}$$

$$\dot{\Psi} = \begin{bmatrix} \hat{\sigma}_c^2 \cdot \Psi_c & \rho \cdot \hat{\sigma}_c^2 \cdot \Psi_c(X_c, X_e) \\ \mathbf{0} & \rho^2 \cdot \hat{\sigma}_c^2 \cdot \Psi_c(X_e, X_e) + \hat{\sigma}_r^2 \cdot \Psi_r \end{bmatrix}, \tag{9}$$

$$\dot{F} = \begin{bmatrix} \mathbf{1} & \mathbf{0} \\ \rho \cdot \mathbf{1} & \mathbf{1} \end{bmatrix}, \dot{M} = \left[\rho\ 1\right], \tag{10}$$

where $(\hat{\sigma}_c^2, \Psi_c)$ and $(\hat{\sigma}_r^2, \Psi_r)$ are matrices obtained from the Kriging models $\hat{f}_c$ and $\hat{f}_r$, respectively (see Section 2.1). In particular, $\hat{\sigma}_c^2$ and $\hat{\sigma}_r^2$ are process variances, while $\Psi_c(\cdot,\cdot)$ and $\Psi_r(\cdot,\cdot)$ denote correlation matrices of two datasets with the optimized $\left(\theta_1 \ldots \theta_d\right)$ parameters and correlation function of the Kriging models $\hat{f}_c$ and $\hat{f}_r$, respectively. The block matrix is the crucial part of co-Kriging, as it is here that the correlation between the cheap and expensive model data is taken into account.

Similarly to Section 2.1, we choose the Matérn correlation function with $\nu = \frac{3}{2}$ for the underlying Kriging models, $\hat{f}_c$ and $\hat{f}_r$. For illustration purposes, Kriging and co-Kriging are applied to a mathematical example, see Figure 2. Using the same expensive data, co-Kriging is able to capture the behavior of the environment better than Kriging, which is attributed to the use of the additional cheap sensing data.

3 Wireless Testbed Facility (w-iLab.t)

The w-iLab.t testbed is a generic and heterogeneous wireless testbed that is used for experimental testing and validation. It consists of two sub-testbeds: the w-iLab.t office and w-iLab.t Zwijnaarde. The w-iLab.t office is deployed in a real office environment whereas the testbed Zwijnaarde is located in a

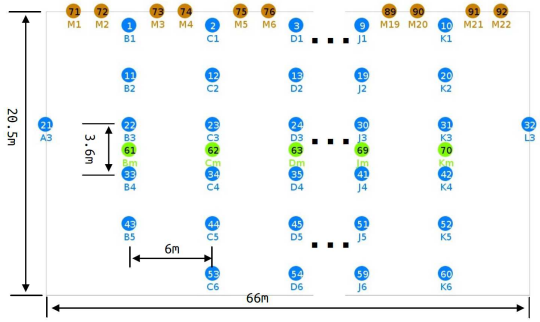


Fig. 3: Topology of the w-iLab.t Zwijnaarde testbed.

utility room. There is little external interference at the Zwijnaarde testbed as no regular human activity is present and most of its walls and ceiling are covered with metal. Since Zwijnaarde testbed was deployed more recently, the devices in this testbed are more advanced in terms of processing power, memory and storage [26,27].

In this work, all experiments are performed at the Zwijnaarde testbed. There are 60 Zotac nodes installed, each having two Wi-Fi interfaces, one sensor node, one Bluetooth dongle and a wired control interface. Furthermore, the testbed is equipped with several types of spectrum sensing devices. These include 6 USRP N210 Software Defined Radios [28] and 7 IMEC Sensing Engines [29]. The testbed is also equipped with mobile nodes which are suited for mobility experiments. The OMF (cOntrol and Management Framework) was adopted, as it allows experimenters to describe their experiments systematically. It provides easy data logging services and the ability to configure multiple devices. The topology of the testbed is presented in Fig. 3.

4 Heterogeneous Sensing Devices

In this section, the capabilities of three different sensing devices is briefly presented. As explained below, the Wi-Fi nodes are considered to be LF nodes whereas USRP and IMEC Sensing Engines are seen as HF information sources.

4.1 Wi-Fi nodes (LF nodes)

The monitor mode of the IEEE 802.11 standard [30] enables the wireless terminal to trace the spectrum activity on the channel it is configured to. More precisely, the receiver of the wireless terminal sniffs all IEEE 802.11 packets that are detectable, regardless of the packet source and destination address. In this way, the terminal can gather information on the present Wi-Fi traffic by means of packet sniffing applications such as tcpdump [31] and libtrace [32]. The radio tap header of the Wi-Fi packets contains link layer information of

the received packets, such as Received Signal Strength Indicator (RSSI), Link Quality Indicator (LQI), packet length, receiving antenna, transmission rate, and other parameters.

Thanks to the increasing utilization of Wi-Fi technology in smart devices, the cost of Wi-Fi interfaces has been declining. This makes Wi-Fi sniffing a promising solution for spectrum monitoring in cognitive radio networks. However, this type of spectrum monitoring is limited to homogeneous traffic, i.e. technologies other than Wi-Fi are excluded from the detection results. Moreover, if the processing capacity of the sniffer terminal is not commensurate with the rate of sniffed traffic, it is likely that the terminal drops some packets which will introduce uncertainties to the fidelity of the measurements.

4.2 USRP Sensing Engine (HF nodes)

The Universal Software Radio Peripheral (USRP) [28] is a Software Defined Radio (SDR) platform maintained by National Instruments. It consists of two parts - a fixed motherboard and a removable daughterboard. The motherboard contains an Analog-to-digital converter (ADC) a digital-to-analog converter (DAC), a field-programmable gate array (FPGA) for digital down sampling and an interface connected to a host computer. The daughterboard provides the RF front-end functionality. There are many third-party software platforms, such as GNU Radio [33] and Iris SDR platform [34], which can communicate with the USRP. Thus, spectrum sensing applications can be implemented in many ways. In our case, the sensing algorithm is FFT-based energy detection, implemented directly above the USRP hardware driver (UHD) [35]. The application uses multi-threading to increase the processing speed on the host machine. On the hexa-core server in w-iLab.t, seamless FFT operation of 25 Msps can be achieved in real time [36], which ensures that no transient signal is missed from the detection.

4.3 IMEC Sensing Engine (HF nodes)

The IMEC Sensing Engine is an integrated sensing device developed by IMEC, an interuniversity micro-electronics center [29]. The design of IMEC Sensing Engine targets on low-power and hand-held devices. Hence it is powered and configured over a single USB connection. Similar to USRP, it has configurable gain settings and a separate PCB for the RF front-end functionality. The imec sensing engine has a very wide RF frequency range (from 100 MHz up to 6 GHz) and a programmable instantaneous bandwidth between 1 MHz and 40 MHz. Additionally, it uses a dedicated IC for signal processing instead of using the host computer. There are several pre-defined modes in the IC, including sensing based on FFT and sensing based on fast sweeping over a set of consecutive RF frequencies.

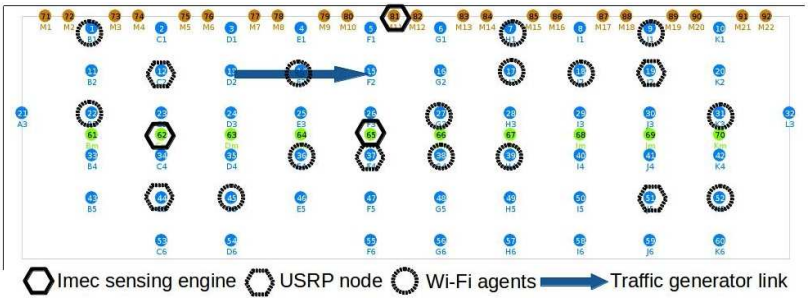


Fig. 4: Topology of the w-iLab.t testbed and sensing devices (training data).

5 Application of co-kriging for building REMs

In this example, it is shown how a comprehensive REM can be built from received signal strength indicator (RSSI) values, which are measured by a heterogeneous set of sensing devices. The setup that was considered is a typical scenario of two IEEE 802.11 standard compliant nodes, operating in infrastructure mode with 802.11g standard and generating active traffic on Uplink. We refer to these two nodes as the System Under Test (SUT). The sender node 13 will directly transmit iPerf data to the receiver node 15 on IEEE 802.11g channel 6 over a short time period of 10 seconds. During the course of traffic generation, all sensing devices collect their measurements into a centralized database. These measurements originate from Wi-Fi nodes, hereby acting as LF information sources, operating on monitor mode and on channel 6. These Wi-Fi agents store the RSSI field of all packets they sniff. Aside form this, the other sensing devices (USRP nodes and IMEC sensing engines) collect HF measurement data at a sparser set of locations. Note that these devices store the power spectral density of all IEEE 802.11 channels into the database. The topological setup of the sensing devices is shown in Fig. 4. Once all measurements are collected, the maximum value of stored values is queried at their corresponding node locations on channel 6. By performing these testbed experiments, a total amount of 8 HF and 14 LF data samples have been collected to build a spatially-interpolated REM using co-Kriging with the Matérn correlation function.

All the HF data samples are shown in Fig. 5 as solid black dots ($\bullet$), whereas the LF data samples are marked as black asterisks ($*$). The surface inbetween these data samples represents the RSSI values which are predicted by the co-Kriging model at arbitrary coordinates in the plane. As can be seen, the model exactly interpolates the HF data whereas the trend of the function is determined by the LF data. This result is also confirmed by Fig. 6, which shows a scatter plot of the predicted and the observed RSSI values. It is seen that predictions at the coordinates of the HF sensing devices matches exactly the HF measurement data, whereas the LF data samples are biased within a

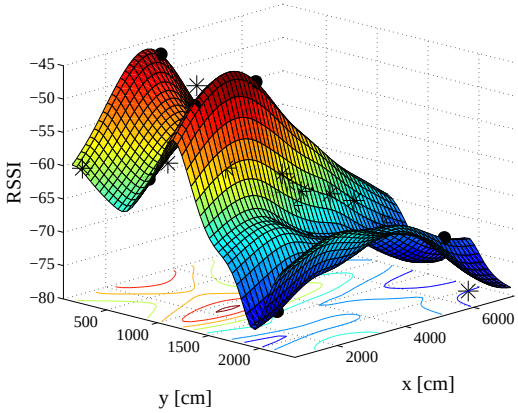


Fig. 5: Spatially-interpolated co-Kriging model of measured RSSI values.

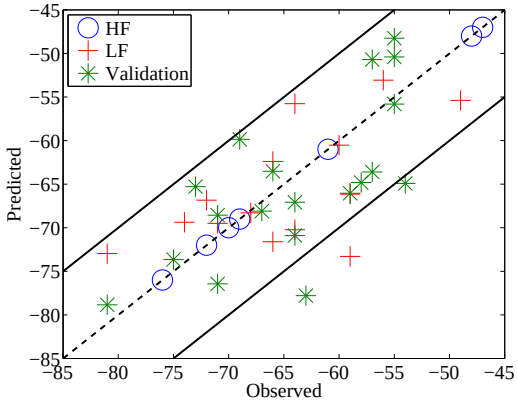


Fig. 6: Predicted data versus observed measurement data.

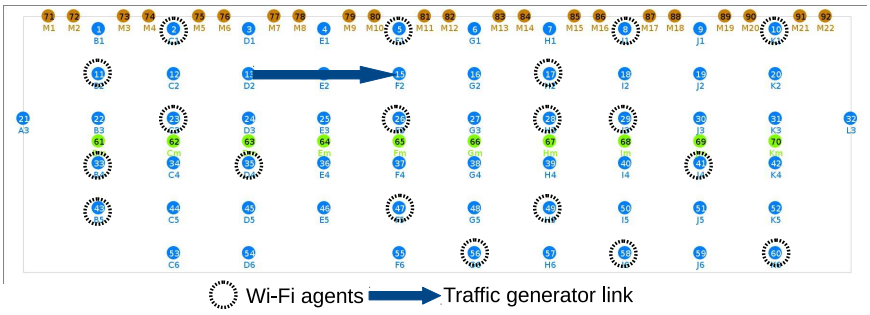


Fig. 7: Topology of the w-iLab.t testbed and sensing devices (validation data).

Table 1: A summary of accuracy metrics of the co-Kriging model.

Data set	NRMSE	SROCC	PLCC
HF training	0.0000	1.000	1.000
LF training	0.1996	0.507	0.624
LF validation	0.2454	0.677	0.688

range of approximately ±10 dB. This deviation matches with the stochastic variability (noise level) of the testbed environment.

In order to validate the accuracy of the model predictions on unseen data, an independent validation set of LF data was measured at different coordinates as shown in Fig. 7. This validation set consists of 19 additional LF measurements, based on another run of experiments which was performed at a later time. As shown in Fig. 6, the predicted RSSI values and the observed measurements in the validation set (+) show a good agreement. This shows that the co-Kriging model is quite accurate as the absolute fitting error lies within the range of ±10 dB for most of the validation data points. A more extensive summary of these results is presented in Table 1, where the Normalized Root Mean Square Error (NRMSE), Spearman's Rank-Order Correlation Coefficient (SROCC) and Pearson Linear Correlation Coefficient (PLCC) are compared for both the training and the validation set. The latter metrics quantify the statistical dependence between the observed and the predicted RSSI values, more specifically the correlation between both variables and their rank. A strong correlation is observed, as most coefficients are substantially higher than 0.5. Of course, a perfect correlation is observed for the HF data as it is interpolated exactly. Moreover, in order to show the feasibility of the variable fidelity data modeling with co-Kriging over the single fidelity data modeling with Kriging, both the HF and the LF data are independently modeled by Kriging. Table 2 shows a summary of accuracy metrics of the Kriging models built with either LF data or HF data only. It can be observed that more than 30% of accuracy reduction is exhibited in the Kriging models built with either LF data or HF data only as compared to the co-Kriging models built with both the LF and the HF data (see Table 1 and Table 2). This clearly shows the advantage of why one can go for co-Kriging over Kriging when a variable fidelity data is available.

Table 2: A summary of accuracy metrics of the Kriging model.

Data set		NRMSE	SROCC	PLCC
Training	Validation			
HF only	LF validation	0.4427	0.3502	0.3015
LF only	HF	0.3020	0.7306	0.6571

As the choice of the correlation function is important for successful modeling in co-Kriging, three different correlation functions which are more widely

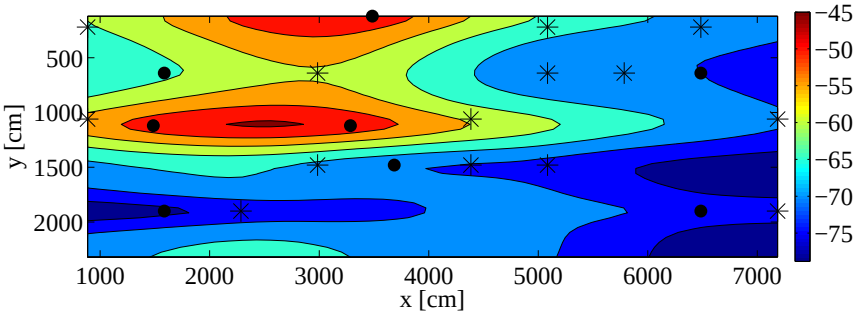


Fig. 8: Radio Environment Map of the predicted RSSI using co-Kriging.

used in the context of surrogate modeling are compared. They are the Gaussian correlation function, Matérn $\frac{3}{2}$ correlation function and Matérn $\frac{5}{2}$ correlation function. All the three correlation functions are observed to be equally capable of providing accurate REM models (see Table 3 which shows a summary of accuracy metrics of the co-Kriging models built with different correlation functions). But, with a close observation, one can see that the Matérn $\frac{3}{2}$ correlation function results in a slightly more accurate REM model than the other correlation functions employed. This is essentially due to the fact that the Gaussian correlation function is infinitely differentiable and is thus very smooth [22]. This smoothness assumption is considered to be unrealistic in real-life data and the Matérn class of correlation functions effectively models such not-so-smooth real-life data [22]. Readers are referred to [13] and [22] for more information on various correlation functions which are commonly used in Kriging-based surrogate modeling.

Table 3: Performance of different correlation functions in co-Kriging model.

Correlation function	Data set	NRMSE	SROCC	PLCC
Gaussian	LF validation	0.2629	0.615	0.674
Matérn 3/2	LF validation	0.2454	0.677	0.688
Matérn 5/2	LF validation	0.2578	0.645	0.668

Finally, a heat map of the predicted RSSI values is visualized in Fig. 8 using the proper aspect ratio, which facilitates a direct comparison with the topology layout of the testbed in Fig. 4. As expected, it can be seen that the RSSI values are elevated in regions where the traffic generator link is situated.

6 Conclusions

This paper presents co-Kriging as a novel methodology to build REMs, based on measurements from heterogeneous sensing devices. It generalizes the existing work on Kriging, in a sense that the different fidelity levels of data can be taken into account. As demonstrated in the example, HF measurements are expensive to obtain so this data is modeled with a very high accuracy. The LF measurements are easier to collect and can be used to guide the trend of the approximation model at inbetween spatial coordinates. The resulting model can be used to build a heat map which visualizes spectrum information and can serve as a monitoring tool that facilitates cognitive decision making.

Acknowledgements The research activities that have been described in this paper were funded by Ghent University, iMinds, the Fund for Scientific Research in Flanders (FWO-V) project G.0325.11N and the Interuniversity Attraction Poles Programme BESTCOM initiated by the Belgian Science Policy Office. This paper is also the result of research carried out as part of the QoCON project funded by iMinds. QoCON is being carried out by a consortium of the industrial partners: Televic, Option and Barco in cooperation with iMinds research groups: IBCN (UGent), WiCa (UGent), SMIT (VUB), PATS (UA) and IMEC. D. Deschrijver and I. Couckuyt are post-doctoral research fellows of FWO-V.

References

1. J. Mitola III and G. Q. Maquire Jr., "*Cognitive Radio: Making Software Radios More Personal*", IEEE Personal Communications Magazine, vol. 6, no. 4, pp. 13-18, 1999.
2. Y. Zhao, "*Enabling Cognitive Radios Through Radio Environment Maps*", Ph.D. dissertation, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, 2007.
3. D. Denkovski, V. Atanasovski, L. Gavrilovska, J. Riihijarvi, "*Reliability of a Radio Environment Map: Case of Spatial Interpolation Techniques*", Proc. of the 7th International ICST Conference on Cognitive Radio Oriented Wireless Networks and Communications, pp. 248-253, 2012.
4. M. Pesko, T. Javornik, M. Stular, M. Mohorcic, "*The Comparison of Methods for Constructing the Radio Frequency Layer of Radio Environment Map using Participatory Measurements*", Proc. of the 4th Workshop of COST Action IC0902, Cognitive Radio and Networking for Cooperative Coexistence of Heterogeneous Wireless Networks, pp. 1-2, 2013.
5. S.-J. Kim, E. Dall'Anese, G. Giannakis, "*Cooperative Spectrum Sensing for Cognitive Radios using Kriged Kalman Filtering*", IEEE Journal of Selected Topics in Signal Processing, vol. 5, pp. 24-36, 2011.
6. M. Portoles-Comeras, C. Ibars, J. Nunez-Martinez, J. Mangues-Bafalluy, "*Characterizing WLAN Medium Utilization for Radio Environment Maps*", IEEE Vehicular Technology Conference (VTC Fall), pp. 1-5, 2011.
7. J. van de Beek, E. Lidstrom, T. Cai, Y. Xie, V. Rakovic, V. Atanasovski, L. Gavrilovska, J. Riihijarvi, P. Mahonen, A. Dejonghe, P. Van Wesemael, M. Desmet, "*REM-enabled Opportunistic LTE in the TV Band*, IEEE International Symposium on Dynamic Spectrum Access Networks (DYSPAN), pp. 272-273, 2012.
8. V. Atanasovski, "*Constructing Radio Environment Maps with Heterogeneous Spectrum Sensors*", Proc. of IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks, pp. 660-661, 2011.
9. L. Gavrilovska, V. Atanasovski, V. Rakovic, D. Denkovski, "*Integration of Heterogeneous Spectrum Sensing Devices Towards Accurate REM Construction*", Chapter 9 in Cognitive Communication and Cooperative HetNet Coexistence, M.-G. Di Benedetto and F. Bader (eds.), Springer International Publishing Switzerland, 2014.

10. M. C. Kennedy, A. O'Hagan, "*Predicting the Output From Complex Computer Code when Fast Approximations are Available*", Biometrika, vol 87, pp. 1-13, 2000.
11. T. Simpson, J. D. Poplinski, P. N. Koch, and J. K. Allen, "*Metamodels for computer-based engineering design: Survey and recommendations*", Engeering with Computers (London), vol. 17, no. 2, pp. 129-150, 2001.
12. J.P.C. Kleijnen, "*Design and Analysis of Simulation Experiments*", Springer, 2008.
13. C. E. Rasmussen and C. K. I. Williams, "*Gaussian Processes for Machine Learning*", MIT Press, 2006.
14. M.T.M. Emmerich, K. Giannakoglou, and B. Naujoks, "*Single- and multiobjective evolutionary optimization assisted by gaussian random field metamodels*", IEEE Transactions on Evolutionary Computation, vol. 10, no. 4, pp. 421-439, 2006.
15. D.G. Krige, "*A statistical approach to some basic mine valuation problems on the Witwatersrand* ", Journal of the chemical, metallurgical and mining society of South Africa, vol. 52, pp. 119-139, 1951.
16. J. Sacks, W. J. Welch, T.J. Mitchell, and H. P. Wynn, "*Design and analysis of computer experiments*", Statistical science, vol. 4, no. 4, pp. 409-435, 1989.
17. J. Knowles and H. Nakayama, "*Meta-modeling in multiobjective optimization*", In Multiobjective Optimization: Interactive and Evolutionary Approaches, pp. 245-284. Springer-Verlag, Berlin, Heidelberg, 2008.
18. I. Couckuyt, D. Deschrijver, and T. Dhaene, "*Fast calculation of the multiobjective probability of improvement and expected improvement criteria for pareto optimization*", Journal of Global Optimization, pp. 1-22, 2013.
19. D. Gorissen, K. Crombecq, I. Couckuyt, P. Demeester, and T. Dhaene, "*A surrogate modeling and adaptive sampling toolbox for computer based design.* Journal of Machine Learning Research", vol. 11, pp. 2051-2055, 2010.
20. G. Wang and S. Shan, "*Review of metamodeling techniques in support of engineering design optimization*", Journal of Mechanical Design, vol. 129, no. 4, pp. 370-380, 2007.
21. I. Couckuyt, A. Forrester, D. Gorissen, F. De Turck, and T. Dhaene, "*Blind kriging: Implementation and performance analysis*", Advances in Engineering Software, vol. 49, pp. 1-13, 2012.
22. M.L. Stein, "*Interpolation of Spatial Data: Some Theory for Kriging*", Springer-Verlag, 1999.
23. M.D. Morris, T.J. Mitchell, and D. Ylvisaker, "*Design and analysis of computer experiments: use of derivatives in surface prediction*", Technometrics, vol. 35, no.3, pp. 243-255, 1993.
24. J. Staum, "*Better simulation metamodeling: The why, what, and how of stochastic kriging*", In Proceedings of the Winter Simulation Conference, 2009.
25. A.I.J. Forrester, A. Sobester, and A.J. Keane, "*Multi-fidelity optimization via surrogate modelling*", Royal Society, vol. 463, no. 2088, pp. 3251-3269, 2007.
26. S. Bouckaert et al., *Federating wired and wireless test facilities through Emulab and OMF: the iLab.t use case*, In Proceedings of TridentCom 2012.
27. W. Liu, S. Keranidis, M. Mehari, J. Vanhie-Van Gerwern, S. Bouckaert, O. Yaron, I. Moerman, "*Various Detection Techniques and Platforms for Monitoring Interference Condition in a Wireless Testbed*", Lecture Notes in Computer Science, vol. 7586, pp. 43-60, 2013.
28. *USRP N210 data sheet*, Ettus Research, Santa Clara (CA), USA.
29. S. Pollin, L. Hollevoet, P. V. Wesemael, M. Desmet, A. Bourdoux, E. Lopez, F. Naessens, P. Raghavan, V. Derudder, S. Dupont, and A. Dejonghe, "*An integrated reconfigurable engine for multi-purpose sensing up to 6 GHz*", IEEE International Symposium on Dynamic Spectrum Access Networks, pp. 656-657, 2011.
30. IEEE 802.11 standard, *Standard for Information Technology - Telecommunications and Information Exchange between Systems Local and Metropolitan Area Networks - Specific Requirements Part 11*, 2012.
31. TCPdump, *a command line packet analyzer*, online: http://www.tcpdump.org, accessed 19 Feb 2014.
32. S. Alcock, P. Lorier and R. Nelson, *Libtrace: A Packet Capture and Analysis Library*, SIGCOMM Comput. Commun. Rev., vol. 42, no. 2, pp. 42-48, 2012.
33. GNU Radio, *online: http://gnuradio.org/redmine/projects/gnuradio.* Last accessed on 2014/03/11.

34. P. D. Sutton, H. Lahlou, S. A. Fahmy, K. E. Nolan and B. Ozgul, T. W. Rondeau, J. Noguera, L. E. Doyle, "*Iris: An Architecture for Cognitive Radio Networking Testbeds*", IEEE Communications Magazine, vol. 48, pp. 114-122, 2010.
35. UHD, *online: http://code.ettus.com/redmine/ettus/projects/uhd/wiki*. Last accessed on 2014/03/11.
36. W. Liu, D. Pareit, Eli D. Poorter, I. Moerman, "*Advanced spectrum sensing with parallel processing based on software-defined radio*", EURASIP Journal on Wireless Communications and Networking, vol. 228, 15 pp, 2013.