Distributed Radio Map Reconstruction for 5G Automotive

Vinay-Prasad Chowdappa, *Member, IEEE,* Carmen Botella, *Senior Member, IEEE*, J. Javier Samper and
Rafael J. Martínez

Abstract—Radio maps are expected to be an essential tool for
the resource optimization and management of 5G automotive. In
this paper, we consider the problem of radio map reconstruction
using a wireless sensor network formed by sensor nodes in
vehicles, access nodes from a smart city infrastructure, etc.
Due to limited resource constraints in sensor networks, it is
crucial to select a small number of sensor measurements for
field reconstruction. In this context, we present a novel distributed
incremental clustering algorithm based on the regression Kriging
method for radio map reconstruction in terms of average received
power at locations where no sensor measurements are available.
The path-loss and shadowing components of the wireless channel
are separately estimated. For shadowing estimation, clusters of
sensors are adaptively formed and their size is optimized in
terms of the least number of sensors by minimizing the ordinary
Kriging variance. The complexity of the proposed algorithm is
analyzed and simulation results are presented to showcase the
algorithm efficacy to field reconstruction.

Index Terms—Radio maps, distributed channel prediction,
Kriging.

I. INTRODUCTION

5G will be a key enabler for the Internet of Things,
providing the platform to the next generation of connected
and autonomous vehicles through vehicle-to-everything (V2X)
communications [1]. Vehicles will need to communicate with
each other and with the network incorporating a traffic man-
agement system, in real-time, for more efficient and safer
use of existing road infrastructure [2], [3]. In this context,
the connectivity or in other words, wireless channel quality,
is critical in fulfilling the vision of 5G automotive [4]. The
wireless channel quality varies significantly as vehicles moves
from one location to another, specially in urban scenarios. The
location of vehicles, specifically, the mobility pattern including
the routes and stops of public transportation vehicles is known,
and constitutes a major portion of traffic [5]. Studies have
also shown that is possible to predict the locations of cars
[6]. Assuming that knowledge of the mobility patterns and
radio maps were available, the channel quality of the vehicle
along the trip could be predicted. Such awareness would guide
the network to learn if a vehicle is heading towards a poor
coverage area, thus adapting its resources to maintain the
quality of services. Moreover, all this process needs to be

performed in a distributed way. To tackle this problem, this
paper proposes an algorithm to reconstruct radio maps in a
distributed way that could support vehicular communications
in the smart city context, addressing the challenges related to
5G automotive.

Radio maps can provide a precise awareness of the radio
environment in the spatial domain by processing the geo-
localized spectrum use. Such information can be gathered by
measurement capable devices like on-board sensor nodes of
vehicles or from access nodes deployed in a smart city context
[7], etc. Updating the radio map frequently is expensive and
practically inefficient as measurements are collected from
various devices. Therefore, it is cost and resource effective to
update the radio map depending on the mobile data traffic and
time in a day. Radio maps also find their potential applications
in 5G heterogeneous networks, where their availability could
be crucial in spectrum sensing in cognitive radios [8]–[10],
interference management [11], coverage analysis [12]–[14],
device to device communications [15], formation control and
connectivity maintenance in multi-agent system [16], and
proactive resource allocation in anticipatory networks [6],
[17], [18]. A radio map contains information such as radio
signal strength, delay spread and interference levels over a
finite geographical region. By definition, radio maps require
to know the field value at every point on the whole area
of interest. However, in practical scenarios it is difficult to
have all such measurements. Therefore, we rely on spatial
interpolation methods to predict the field values at spatial
locations where no measurements are available, based on a
set of available measurements.

Spatial interpolation methods combine the available sensor
node measurements with geo-location information to construct
a complete map. One of the key challenging tasks in radio map
reconstruction is choosing an appropriate interpolation method
offering good quality and complexity trade-off. Methods that
have been specifically proposed are: nearest neighbor [19], thin
plate splines [20], [21], natural neighbor [20], inverse distance
weighting [22]–[25], Kriging [26] and GPR [16], [27]. Kriging
is one of the most frequently applied methods for radio map
reconstruction [8], [9], [13], [14], [28]–[34]. It is a stochastic
method capable of modeling deterministic variations (large-
scale), spatially autocorrelated variations (small-scale) and
uncorrelated noise. The advantage of Kriging is that estimation
error maps can be obtained, i.e., uncertainties associated with
predictions. The two-step process of Kriging begins with a
semivariogram modeling and Kriging prediction. The semi-
variogram model characterizes the spatial correlation in the

data and the Kriging method employs this model to generate the best linear unbiased estimates. Some notable work in this line is for example, [14], where fixed rank Kriging is applied to cellular coverage analysis with the aim of lowering the complexity of the spatial interpolation. In [12], [13], bayesian Kriging is implemented to construct the radio map for the purpose of coverage hole detection in cellular networks. In [23], universal Kriging is utilized for estimating the radio environment. In [35], a distributed Kriging algorithm based on ordinary Kriging (OK) is proposed in wireless sensor networks (WSNs) in order to interpolate the physical phenomenon inside the coverage holes. Another tool from spatial statistics is the kriged kalman filter, for modeling spatio-temporal variations [36]. Note, this paper deals with purely spatial models and temporal variations are not considered.

In this paper, we consider a cellular V2X standard which includes direct communication (between vehicles, vehicle-to-pedestrian, and vehicle-to-infrastructure) and cellular communications with networks [37]. In the smart city context, we consider a heterogeneous network combining one transmitter and a network of receivers, which consists of sensors placed in vehicles, specially from the public transportation network, pedestrian phones and fixed access nodes. In this framework, we employ a hybrid variant of Kriging interpolation known as regression Kriging (RK) [38], [39] to predict the wireless channel in terms of the transmitter's average received power. The received signal power can be probabilistically modeled as a multi-scale dynamical system consisting of path-loss, shadowing and small-scale fading [40]. Small-scale fading decorrelates within tens of centimeters, making it infeasible to predict even with highly accurate position information [27]. RK can harness both path-loss and shadowing using well established models [41]. Basically, RK is a non-stationary geo-statistical method that combines a regression of the dependent variable on auxiliary variables with Kriging of the regression residuals [38]. The auxiliary information (known as drift in geo-statistics) is assumed to be available at all the sensor nodes and it is the path-loss in our work. The key assumption of the RK is that there is no spatial dependence between the auxiliary variable and the residual of the linear regression. This makes it simpler to implement compared to its mathematical equivalent universal Kriging [39]. In RK, the drift is accounted by the regression model, while the model residuals by OK. The OK method works with assumption of constant mean and presents several interesting features such as: (1) It is a local interpolator, which operates within a small area around the estimation location and captures the short-range variations [42], [43]. (2) Along with the estimates, it also quantifies Kriging variance. Thus, Kriging combined with regression and thereby incorporating auxiliary information has proven to improve the precision of the prediction, when compared with plain OK, co-Kriging and plain regression [38].

Most of the existing work in the radio map reconstruction literature has focused on centralized estimation techniques, where all the measurements are forwarded to a central node or sink that performs all the field estimation. However, the centralized solution shows several drawbacks. Firstly, high

energy is required to communicate between the sensor nodes and the central link. Secondly, the centralized schemes are neither robust to central node failure nor scalable to network size increments. Lastly, the most important limitation is the large computational complexity resulting from calculating prediction using all sensor measurements. Further for Kriging method, due to the inversion of a $N \times N$ semivariance matrix, the computation cost scales as the cube of the number of sensor measurements N, resulting in cubic time complexity $\mathcal{O}(N^3)$. The scaling problem of cubic time complexity with respect to the sensor measurements prevents practical applications of centralized Kriging.

To overcome these limitations, in this paper, we apply the RK interpolation method, with the aim of reconstructing radio maps through a novel distributed incremental clustering algorithm (DICA-RK). The objective of this paper is two fold: (a) to propose a distributed reconstruction algorithm and (b) to reduce the computational complexity through an adaptive clustering of sensor nodes. Our algorithm employs the least possible number of measurements $n \ll N$ without compromising the accuracy of Kriging interpolation. As a result, the complexity is significantly reduced to $\mathcal{O}(n^3)$. Thus, driven by the advantages of combining regression with OK (i.e., RK) and also by the key features mentioned above that shows OK towards clustering, we extend our previous work [34], [44] on standard OK and propose a novel distributed algorithm based on RK for spatial prediction. Our main contributions can be summarized as follows:

- The problem of distributed spatial field reconstruction of radio maps that could support 5G automotive in a smart city context is formulated and analyzed, in terms of algorithm design and complexity analysis.
- A novel DICA-RK is presented, which consists of distributed ordinary least square estimation (D-OLS) and distributed cluster based OK prediction (DC-OK). The D-OLS estimates the path-loss while the DC-OK estimates the shadowing. In DC-OK, initial clusters of sensors are built first to perform semivariogram estimation and Kriging prediction locally in a distributed way. Later, sensor nodes which minimize the Kriging variance are added to the initial clusters to improve the quality of estimation. The final prediction estimate is obtained by summing the estimates of D-OLS and DC-OK.
- We detail the cluster formation technique in the DC-OK for shadowing estimation and present an update model to instantly calculate the Kriging weights and Kriging estimates. Our method is highly local in the sense that it operates within a small neighborhood around the estimation point and captures the local or short-range variations.
- Performance assessment results and interpolated maps are presented to interpret the reconstruction quality. In addition, the results obtained with the DICA-RK algorithm are compared with classical interpolation methods such as splines and natural neighbor approaches. We also investigate the impact of location uncertainty on the performance of the proposed algorithm.
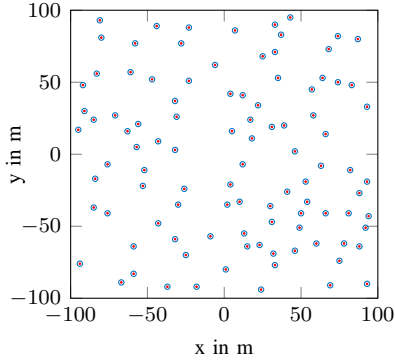
Fig. 1. Random deployment of 225 sensor nodes.

II. MODEL AND PROBLEM STATEMENT

A. Network model

We consider a heterogeneous network formed by a transmitter TX and a network of wireless receivers RXs. The receiver network is modeled as a connectivity graph $\mathbf{G}(\mathbf{V}, \mathbf{E})$, composed by a set of N sensor nodes $\mathbf{V} = \{1, 2, ..., N\}$ and a set of links $\mathbf{E}$. The sensor nodes are deployed over a square area $\mathbf{s} \in \mathbb{R}^2$ to monitor a 2-D spatial physical field denoted by $z(\mathbf{s}_i)$, where $i = \{1, 2, ..., N\}$. Each sensor node i is located at a spatial location denoted by $\mathbf{s}_i = (x_i, y_i)$. In matrix notation, $\mathbf{z} = [z(\mathbf{s}_1), z(\mathbf{s}_2), ..., z(\mathbf{s}_N)]$ and $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_N]$. We denote the spatial location where the field needs to be estimated as $\mathbf{s}_0$. Due to sensor network power constraints, the transmission range of each sensor node is limited to a distance R. As a result, the communication between sensor nodes i and j is feasible only if the euclidean distance $d(\mathbf{s}_i, \mathbf{s}_j) = ||\mathbf{s}_i - \mathbf{s}_j||$ is less than R. We assume that sensor nodes have perfect knowledge of their location information, allowing them to estimate the inter-node distance $d(\mathbf{s}_i, \mathbf{s}_j)$ with one hop neighbors and also, calculate the distance $d(\mathbf{s}_i, \mathbf{s}_0)$ with spatial location $\mathbf{s}_0$. N and n represent the network size and cluster size, respectively. The sensor node deployment and transmission range define the WSN topology, which in this paper is a random distribution of sensor nodes as shown in Fig. 1.

To fulfill the challenges posed by 5G automotive, we assume a completely distributed architecture. Each sensor node acts as a cluster head in the fully distributed architecture. Sensor nodes within the communication range collaborate and form clusters. Sensor nodes in each cluster perform local field reconstruction by exchanging measurements in each cluster within their neighborhood. Sensor nodes in the network operate co-operatively in the sense that they are allowed to contribute to multiple location estimations depending on their position.

B. Channel model

Consider a transmitter TX located at $\mathbf{s}_t$ in a square area $\mathbf{s} \in \mathbb{R}^2$ and assume that P_{TX} is the power transmitted through a wireless propagation channel. The power measured by N sensor nodes can be modeled as:

$$z(\mathbf{s}_i) = P_{\text{RX}}(\mathbf{s}_i) + n_i, \qquad i = 1, 2, \ldots, N, \qquad (1)$$

where $P_{\text{RX}}(\mathbf{s}_i)$ is the power received by a sensor node at location $\mathbf{s}_i$ from a single-antenna transmitter TX located at $\mathbf{s}_t$. The term n_i accounts for measurement noise and can be modeled as a zero mean additive white Gaussian noise with variance σ_n^2 .

The received signal power in dB scale can be expressed as the sum of path-loss and shadowing components:

$$P_{\text{RX}}(\mathbf{s}_i) = G_0 - 10\eta \log_{10} ||\mathbf{s}_t - \mathbf{s}_i|| + y(\mathbf{s}_i). \qquad (2)$$

Constant $G_0 = P_{\text{TX}} + K_{dB} + 10\eta \log_{10} d_0$, where K_{dB} is the constant path-loss factor, η is the path-loss exponent and d_0 is the reference distance. The term $||\mathbf{s}_t - \mathbf{s}_i||$ is the distance between transmitter at location $\mathbf{s}_t$ and sensor node at location $\mathbf{s}_i$, and $y(\mathbf{s}_i)$ is the location dependent shadowing. We assume that the shadowing follows a log-normal distribution i.e., $y(\mathbf{s}_i) \sim \mathcal{N}(0, \sigma_\Psi^2)$, where σ_Ψ^2 is the shadowing variance. The channel model (2) has been empirically tested by [45]–[48], to accurately model the variations of the received signal power in a wireless propagation channel. The path-loss component is clear from equation (2) whereas for modeling the spatial correlation of shadowing, we employ the Gudmundson model [41]:

$$C(\mathbf{s}_i, \mathbf{s}_j) = \sigma_\Psi^2 \exp\left(-\frac{||\mathbf{s}_i - \mathbf{s}_j||}{d_c}\right), \qquad (3)$$

where d_c is the correlation distance of the shadowing.

C. Problem statement

The objective of this paper is to obtain a high-quality radio map for future 5G automotive in smart city deployments. In other words, our aim is to perform

1) *Distributed semivariogram/ parameter estimation:* estimate the channel parameters $\theta = [\eta, G_0, \sigma_\Psi^2, d_c, \sigma_n^2]^{\text{T}}$ from measurements $\{\mathbf{S}, \mathbf{z}\}$.
2) *Distributed prediction:* obtain an estimate of the field value $\hat{z}(\mathbf{s}_0)$ at location $\mathbf{s}_0$ where values are not known, using the least number of geo-location aware sensor node measurements $z(\mathbf{s}_i)$.

III. SPATIAL STATISTICS

In this section, we present the techniques developed in the context of geo-statistics, a subfield of spatial statistics, for capturing the spatial structure from sensor measurements and for the reconstruction of a realistic radio map. We modify the regression Kriging presented in [38] to be suitable for wireless channel prediction assuming first a centralized setting, which means that all the sensor measurements are available at a central entity. Geo-statistics treats the various processes in the environment as the realizations of random processes [26], [43]. Since we are are dealing with a two-dimensional space, we will treat the measurements as a random field.

A. Random field

Sensor node measurements obtained at various sensor node locations represent a continuous spatial phenomenon that can be modeled as a random field. The random field in two

dimensions is denoted as $\{z(\mathbf{s}_i) : \mathbf{s}_i \in D \subset \mathbb{R}^2\}$, where $i = \{1, 2, ..., N\}$ and D is the domain of interest. $z(\mathbf{s}_i)$ is decomposed into [49]:

$$z(\mathbf{s}_i) = \mu(\mathbf{s}_i) + y(\mathbf{s}_i), \qquad i = 1, 2, \ldots, N, \qquad (4)$$

where $\mu(\mathbf{s}_i) = \mathbb{E}[z(\mathbf{s}_i)]$ is a deterministic mean function:

$$\mu(\mathbf{s}_i) = P_{\text{TX}} + G_0 - 10\eta \log_{10} \|\mathbf{s}_t - \mathbf{s}_i\|, \qquad (5)$$

and $y(\mathbf{s}_i)$ is a zero mean stationary random process with variogram function $\gamma(\mathbf{h})$, called residual variogram function of $y(\mathbf{s}_i)$:

$$\begin{aligned} \gamma(\mathbf{h}) &= \frac{1}{2} \text{Var}[y(\mathbf{s}_i) - y(\mathbf{s}_i + \mathbf{h})] \\ &= \frac{1}{2} \mathbb{E}[(y(\mathbf{s}_i) - y(\mathbf{s}_i + \mathbf{h}))^2], \end{aligned} \qquad (6)$$

where $y(\mathbf{s}_i)$ and $y(\mathbf{s}_i + \mathbf{h})$ are the values of shadowing at location $\mathbf{s}_i \in D$ and $\mathbf{s}_i + \mathbf{h} \in D$, respectively, $\mathbb{E}$ is the expectation and $\mathbf{h}$ (or $\|\mathbf{h}\|$) is the lag distance representing the separation between two spatial locations. The random field is further assumed to be isotropic and thus, $\gamma(\mathbf{h})$ depends only on $\mathbf{h}$. In terms of semivariance, the Gudmundson model (3) is known as exponential model in the geo-statistics literature and it is given by:

$$\gamma(\mathbf{h}) = \begin{cases} \sigma_0^2 + \sigma_\Psi^2 \left(1 - \exp\left(-\frac{\|\mathbf{s}_i - \mathbf{s}_j\|}{d_c}\right)\right), & \|\mathbf{s}_i - \mathbf{s}_j\| > 0, \\ 0, & \mathbf{s}_i = \mathbf{s}_j, \end{cases} \qquad (7)$$

where the nugget σ_0^2 specifies the uncertainty of the semivariogram at a distance close to zero, σ_Ψ^2 is the spatially correlated variance and the range d_c is the spatial correlation limit. The quantity $\sigma_0^2 + \sigma_\Psi^2$ is known as sill and it is the value at which the semivariogram attains the range.

B. Semivariogram analysis

For predicting random fields at unmeasured locations based on observations of the random field, we use the Kriging interpolation method. Moreover, for choosing an appropriate Kriging type, we need to first perform a semivariogram analysis.

The semivariogram describes the spatial variability of a random field from a set of observations. It is a structural and descriptive tool that measures the spatial correlation as a function of distance. Basically, a semivariogram analysis consists of estimating the experimental semivariogram (EV) followed by semivariogram modeling. The spatial statistics of the random field can be obtained from the set of observations $z(\mathbf{s}_i)$, where $i = \{1, 2, ..., N\}$, by estimating the EV $\hat{\gamma}(\mathbf{h})$. The EV, which is defined as half the average squared difference between the points separated by a lag distance $\mathbf{h}$, is computed using Matherons method of moments estimator [49]:

$$\hat{\gamma}(\mathbf{h}) \equiv \frac{1}{2|\mathcal{N}(\mathbf{h})|} \sum_{\mathcal{N}(\mathbf{h})} (z(\mathbf{s}_i) - z(\mathbf{s}_j))^2, \qquad (8)$$

where $z(\mathbf{s}_i)$ and $z(\mathbf{s}_j)$ are field values at locations $\mathbf{s}_i$ and $\mathbf{s}_j$, respectively. $\mathcal{N}(\mathbf{h}) = \{(\mathbf{s}_i, \mathbf{s}_j) : \mathbf{s}_i - \mathbf{s}_j \in \mathbf{h} \text{ for } i, j =$

$1, \ldots, N\}$ denotes the set of all location pairs separated by the particular lag distance $\mathbf{h}$, whereas $|\mathcal{N}(\mathbf{h})|$ denotes the number of distinct pairs in $\mathcal{N}(\mathbf{h})$.

Semivariogram modeling is an important step between spatial description and spatial prediction. The EV provides semivariance estimates only at a finite set of lags. However, the Kriging method requires correlation between samples where no measurements are available. In order to obtain the estimates at arbitrary lags, the EV is replaced by a parametric semivariogram model (SV). A SV model is a simple mathematical expression that models the trend in the EV. The typical choices in geo-statistics are spherical, Gaussian and exponential models. In this paper, we consider the exponential model, denoted as $\overline{\gamma}(\mathbf{h})$, to model the spatial correlation, as it is equivalent to the Gudmundson model (3) proposed for cellular networks. Note, $\overline{\gamma}(\mathbf{h})$ is presented earlier in equation (7).

After the model selection, the model parameters $\hat{\theta} = [\hat{\sigma}_\Psi^2, \hat{d}_c]$ can be determined by fitting the semivariogram model to the EV using a given analytical method. The most employed methods are maximum likelihood estimation (MLE), weighted least squares estimation (WLSE) and generalized least squares estimation (GLSE) [50], [51]. The drawback is that the MLE relies heavily on a Gaussian distribution and the estimates are biased, and that the GLSE is computationally more demanding. Hence, we employ WLSE because of its true compromise between simplicity and statistical efficiency [26]. One of the crucial steps in the fitting process is the initialization of the model parameters. We use the following equations to compute the initial values from the EV [50]:

$$\sigma_{0_{ini}}^2 = max\left[0, \hat{\gamma}(\mathbf{h}_1) - \frac{\mathbf{h}_1}{\mathbf{h}_2 - \mathbf{h}_1}\hat{\gamma}(\mathbf{h}_2) - \hat{\gamma}(\mathbf{h}_1)\right],$$

$$d_{c_{ini}} = \frac{\mathbf{h}_N}{2}, \qquad (9)$$

$$\sigma_{0_{ini}}^2 + \sigma_{\psi_{ini}}^2 = \frac{\hat{\gamma}(\mathbf{h}_{N-2}) + \hat{\gamma}(\mathbf{h}_{N-1}) + \hat{\gamma}(\mathbf{h}_N)}{3},$$

where $\sigma_{0_{ini}}^2$, $d_{c_{ini}}$ and $\sigma_{0_{ini}}^2 + \sigma_{\psi_{ini}}^2$ are the initial values for nugget, range and sill, respectively.

An EV and SV for a given set of sensor measurements $z(\mathbf{s}_i)$ are shown in Fig. 2 in red color. In general, it can be seen that the semivariance increases when increasing the lag distance, which is a characteristic of global trend in the measurements. However, the figure shows that the EV of the measurements is no longer a function of a single random variable and the assumption of intrinsic stationarity on which the variogram function is defined no longer holds. As a result, $z(\mathbf{s}_i)$ should be modeled as a combination of a deterministic mean function (known as drift in geo-statistics) and a random component. For dealing with the trend in the measurements, we employ a variant of universal Kriging known as regression Kriging.

C. Centralized Regression Kriging

RK provides the solution to the problem of field estimation based on the spatial statistical model of a random field. Let $\mathbf{z} = [z(\mathbf{s}_1), z(\mathbf{s}_2), ..., z(\mathbf{s}_N)]$ be the data observed using N sensor nodes at measured spatial locations $(\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_N)$. We assume
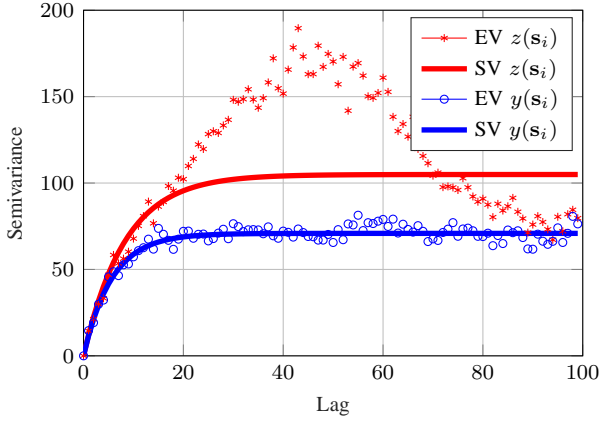
Fig. 2. Semivariogram modeling of $z(\mathbf{s}_i)$ versus $y(\mathbf{s}_i)$

that the sensor nodes know G_0 and σ_n. Then, RK provides an estimate of channel parameters $\hat{\theta} = [\hat{\eta}, \hat{\sigma}_\psi^2, \hat{d}_c]$ and predicts $\hat{z}(\mathbf{s}_0)$ at unmeasured spatial locations $\mathbf{s}_0$ using N sensor nodes, by summing the estimated mean and the residuals:

$$\hat{z}(\mathbf{s}_0)|_N = \hat{\mu}(\mathbf{s}_0)|_N + \hat{y}(\mathbf{s}_0)|_N, \tag{10}$$

where the mean $\hat{\mu}(\mathbf{s}_0)$ is obtained using linear regression analysis and the residual $\hat{y}(\mathbf{s}_0)$ is interpolated using OK. Comparing the model (10) and the channel model (2), we can deduce that the mean is the path-loss and the residual is the shadowing.

1) Path-loss Estimation: Assuming that the correlation between the path-loss component and the random shadowing is zero, the ordinary least square (OLS) estimate of the path-loss exponent $\hat{\eta}$ can be computed as

$$\hat{\eta} = (\mathbf{h}^{\mathrm{T}}\mathbf{h})^{-1}\mathbf{h}^{\mathrm{T}}(\mathbf{z} - \mathbf{1}^{\mathrm{T}}G_0), \tag{11}$$

where $\mathbf{h}^{\mathrm{T}} = -10[\log_{10}(||\mathbf{s}_1 - \mathbf{s}_t||), \log_{10}(||\mathbf{s}_2 - \mathbf{s}_t||), ..., \log_{10}(||\mathbf{s}_N - \mathbf{s}_t||)]^{\mathrm{T}}$. Once $\hat{\eta}$ is estimated, the mean of the path-loss estimate at $\mathbf{s}_0$ is given by,

$$\hat{\mu}(\mathbf{s}_0)|_N = \mathbf{1}^{\mathrm{T}}G_0 + \eta\mathbf{h}_0, \tag{12}$$

where $\mathbf{h}_0^{\mathrm{T}} = -10[\log_{10}(||\mathbf{s}_1 - \mathbf{s}_0||), \log_{10}(||\mathbf{s}_2 - \mathbf{s}_0||), ..., \log_{10}(||\mathbf{s}_N - \mathbf{s}_0||)]^{\mathrm{T}}$.

2) Shadowing Estimation: The random residual for shadowing estimation is obtained by de-trending the measurements $z(\mathbf{s}_i)$:

$$y(\mathbf{s}_i) = z(\mathbf{s}_i) - \mu(\mathbf{s}_i), \qquad i = 1, 2, \ldots, N. \tag{13}$$

Then, $y(\mathbf{s}_i)$ becomes a zero-mean random variable. As a result, the stationarity condition holds and OK can be used. Field estimation using OK follows two key steps: semivariogram estimation and Kriging prediction. The objective of semivariogram estimation is to obtain the parameters $\hat{\theta} = [\hat{\sigma}_\Psi^2, \hat{d}_c]$. For the sake of readability, we refer readers to section III-B for details on semivariogram modeling and present the results of EV and SV for $y(\mathbf{s}_i)$ in Fig. 2.

After estimating $\hat{\theta}$, semivariogram model can be employed in the Kriging system of equations for prediction. Kriging is

a statistical interpolation method for obtaining an estimation of the shadowing $\hat{y}(\mathbf{s}_0)$ at an unmeasured spatial location $\mathbf{s}_0$, from the weighted linear combinations of available data. This is achieved by allocating weights to each sensor node in such way that the Kriging variance is minimized. Let $(y(\mathbf{s}_1), y(\mathbf{s}_2), ..., y(\mathbf{s}_N))$ be the N shadowing data at measured spatial locations $(\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_N)$. Then, the kriged estimate $\hat{y}(\mathbf{s}_0)$ at unmeasured spatial location $\mathbf{s}_0$ using N sensor nodes is the weighted average of the data in its neighborhood given by [26]:

$$\hat{y}(\mathbf{s}_0)|_N = \sum_{i=1}^N w_i(\mathbf{s}_0)|_N \, y(\mathbf{s}_i), \tag{14}$$

where N is the number of available sensor nodes, $\hat{y}(\mathbf{s}_0)|_N$ is the shadowing estimate using N sensor nodes and $w_i(\mathbf{s}_0)|_N$ is the weight allocated for sensor node i from an estimation performed using N sensor nodes. These weights fulfill the unbiased conditions of the estimator, that is:

$$\sum_{i=1}^N w_i(\mathbf{s}_0)|_N = 1, \tag{15}$$

and the expected error between the estimated value and the actual value at location $\mathbf{s}_0$, $\mathbb{E}[\hat{y}(\mathbf{s}_0) - y(\mathbf{s}_0)]$, is 0. The weights in equation (14) can be obtained by solving a set of linear equations known as the Kriging system, which contains the semivariance drawn from an analytical model. The Kriging system is given by:

$$\sum_{i=1}^N w_i(\mathbf{s}_0)|_N \overline{\Gamma}(\mathbf{s}_i, \mathbf{s}_j) + \mathcal{L}(\mathbf{s}_0) = \sum_{i=1}^N \overline{\gamma}(\mathbf{s}_i, \mathbf{s}_0), j = 1, 2, \ldots, N, \tag{16}$$

where $\overline{\Gamma}(\mathbf{s}_i, \mathbf{s}_j)$ is the semivariogram between measurements from sensor node locations $\mathbf{s}_i$ and $\mathbf{s}_j$, $\overline{\gamma}(\mathbf{s}_i, \mathbf{s}_0)$ is the semivariogram between samples from sensor node location $\mathbf{s}_i$ and target location $\mathbf{s}_0$ and $\mathcal{L}(\mathbf{s}_0)$ is the the Lagrange multiplier. Note that $\overline{\Gamma}(\mathbf{s}_i - \mathbf{s}_j)$ and $\overline{\gamma}(\mathbf{s}_i - \mathbf{s}_0)$ are obtained from the theoretical exponential model (7).

The OK system can be represented in matrix form as:

$$\mathbf{A}\boldsymbol{\lambda} = \mathbf{b}, \tag{17}$$

where

$$\mathbf{A} \equiv \begin{cases} \overline{\Gamma}(\mathbf{s}_i, \mathbf{s}_j), & i, j = 1, \ldots, N, \\ 1, & i = N + 1, j = 1, \ldots, N, \\ & j = N + 1, i = 1, \ldots, N, \\ 0, & i, j = N + 1, \end{cases}$$

$$\boldsymbol{\lambda} = [w_1, w_2, \ldots, w_{\mathrm{N}}, \mathcal{L}]^{\mathrm{T}},$$

$$\mathbf{b} = [\overline{\gamma}(\mathbf{s}_0, \mathbf{s}_1), \overline{\gamma}(\mathbf{s}_0, \mathbf{s}_2), \ldots, \overline{\gamma}(\mathbf{s}_0, \mathbf{s}_N), 1]^{\mathrm{T}}.$$

The minimized estimation variance for N sensor nodes, referred to as the OK variance, can be calculated as:

$$\sigma^2(\mathbf{s}_0)|_N = \sum_{i=1}^N w_i(\mathbf{s}_0)\overline{\gamma}(\mathbf{s}_i, \mathbf{s}_0) + \mathcal{L}(\mathbf{s}_0). \tag{18}$$

IV. Algorithm description

In centralized estimation, the Kriging equations (10) and (18) use all the available sensor measurements to calculate the spatial interpolations. In this paper, our objective is to minimize the number of measurements through DICA-RK. The DICA-RK estimates the path-loss and shadowing separately, and subsequently, combines them to obtain the final wireless channel prediction. These operations are performed in a distributed way using a D-OLS and a DC-OK algorithms. The complete algorithm is represented in the flowchart (see Fig. 3). DICA-RK consists of the following phases:

A. Neighbor discovery

We employ a broadcasting protocol in order to find the set of sensor nodes in the neighborhood of $\mathbf{s}_0$ i.e., $\mathcal{N}_{\mathbf{s}_0}$, and the node closest to $\mathbf{s}_0$. Note that the neighborhood of $\mathbf{s}_0$ includes all the nodes that are in-range and at one-hop distance from $\mathbf{s}_0$. Neighbor discovery is performed by each node to gather information about its multi-hop neighborhood. During this stage, each node broadcast its one-hop neighborhood information to its neighbors. At the end of multiple message exchanges, each node is aware about its multi-hop neighborhood. The broadcast messages are received by multiple sensor nodes that are within the communication range R. These messages include information about node measurements and their location. All the information gathered during this stage is later utilized for local message exchanges for mean estimation and Kriging prediction.

B. Distributed OLS (D-OLS)

Once the sensor nodes are aware about the information of their multi-hop neighborhood, they can estimate $\hat{\eta}$ and thereby compute $\hat{\mu}(\mathbf{s}_0)$. OLS equation (11) can be solved using standard distributed methods, such as distributed recursive least-squares algorithm from [52]. Each node can compute the path-loss component and subtract it with its own measurement to obtain the shadowing component $y(\mathbf{s}_i)$ using (13).

C. Distributed Cluster based OK (DC-OK)

For shadowing estimation, equations (14) and (16) require the knowledge of the semivariogram between the unmeasured location and all the sensor nodes in the WSN. However, this estimation technique is practically inefficient, since OK is a local interpolator method and the closest sensor nodes to the unmeasured location carry more significant weight than the distant ones. Moreover, the influence of distant sensor measurements on the estimates is minimal [43]. Due to this, we can restrict the assumption of stationarity of the mean of the shadowing $\mathbf{y}(\mathbf{s})$ to the local neighborhood of the unmeasured location. The semivariogram can be estimated and modeled only at a few lag distances by capturing only local variations. This means that the Kriging system (16) can be formed with a small number of sensor nodes, $n \ll N$. Consequently, inverting matrix $\mathbf{A}$ will be rapid and computational time can be saved. Therefore, we present a DC-OK algorithm to minimize the computation complexity of the shadowing estimation, by
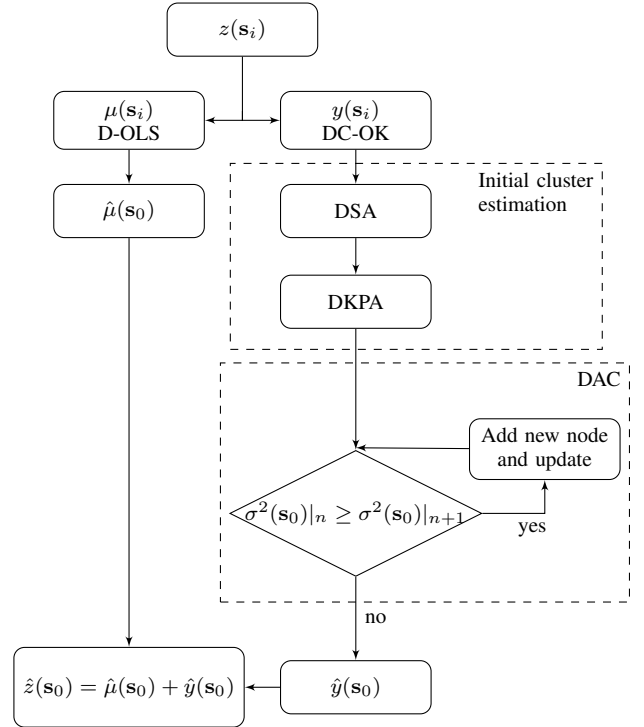

Fig. 3. DICA-RK algorithm

using only a small subset of sensor nodes. The objective is to improve the quality of shadowing estimation by forming adaptive clusters with the least number of sensor nodes. Note that each cluster can be formed by a different number of sensors. This is achieved by progressively incorporating the most relevant sensor node to each cluster and simultaneously updating the Kriging weights and variances. The size of the cluster is optimized by using the Kriging variance as a metric. DC-OK consists of:

1) Initial cluster estimation: An initial set of one-hop, in-range, p sensor nodes, that are closest to unmeasured spatial location $\mathbf{s}_0$, forms an initial cluster and begins the initial estimation process. In this paper, the WLSE analytical fitting method is employed. Due to this, the initialization of the semivariogram model parameters using equation (9) require a minimum of three sensor node measurements. Therefore, in our algorithm, the value of the initial cluster size is set to $p = 3$. The initial cluster of p sensor nodes computes the spatial statistics of the data and the Kriging estimate through a distributed OK operation, which consists of two steps:

a). Distributed semivariogram estimation: The EV between all the sensor nodes in the cluster is obtained by the iterative process of computation and exchange of information with the sensor nodes within the cluster. Due to this iterative process, an initial cluster is built, the field similarity is known and the EV is estimated between all the sensor nodes in the cluster. To compute the EV, we present an iterative algorithm described in Algorithm 1, which solves equation (8) in a distributed way. Sensor node i closest to $\mathbf{s}_0$, $i \in \mathcal{N}_{\mathbf{s}_0}$, activates and sends a packet containing its field measurement value and geo-location information to the next closest sensor node j to $\mathbf{s}_0$, $j \in \mathcal{N}_{\mathbf{s}_0}$,

Algorithm 1 Distributed semivariogram algorithm (DSA)

$\mathcal{N}_{\mathbf{s}_0}$: Set of sensor nodes in $\mathbf{s}_0$ neighborhood // p: Number of nodes in the initial cluster // $\mathbf{h}$: Lag distance //
$d(\mathbf{s}_i, \mathbf{s}_j)$: Euclidean distance between sensor nodes (i,j) //
$\hat{\gamma}(\mathbf{h})$: Experimental semivariogram // $\mathcal{N}_p$: Set of sensor nodes included in the initial cluster.

1: Sensor node i closest to $\mathbf{s}_0$, $i \in \mathcal{N}_{\mathbf{s}_0}$, activates and looks for next closest sensor node j to $\mathbf{s}_0$, $j \in \mathcal{N}_{\mathbf{s}_0}$.
2: Sensor node i sends a packet containing $y(\mathbf{s}_i)$, location $\mathbf{s}_i$ and $\hat{\gamma}(\mathbf{h})$ to sensor node j.
3: Sensor node j receives the packet and calculates the $d(\mathbf{s}_i, \mathbf{s}_j)$ and $\hat{\boldsymbol{\gamma}}(\mathbf{h})$ using (8).
4: Sensor nodes $i, j \in \mathcal{N}_p$ look for the next closest sensor node k to $\mathbf{s}_0$.
5: **while** $|N_p| < p$ **do**
6:    **if** exists a sensor node k closest to $\mathbf{s}_0$, $k \in \mathcal{N}_{\mathbf{s}_0}$ **then**
7:       A new node is added to the cluster.
8:    **end if**
9: **end while**
10: An initial cluster is formed by p nodes and semivariogram is calculated.
11: Each sensor node in $\mathcal{N}_p$ performs semivariogram modeling.

Algorithm 2 Distributed Kriging prediction algorithm (DKPA)

$\mathbf{r}_i$: row of Kriging system (16) created by sensor node i //
$\overline{\Gamma}(\mathbf{s}_i, \mathbf{s}_j)$: semivariogram between sensor nodes i and j //
$\mathcal{L}$: Lagrange multiplier.

1: Each sensor node $i \in \mathcal{N}_p$ creates row $\mathbf{r}_i$.
2: Each sensor node i computes $\overline{\Gamma}(\mathbf{s}_i, \mathbf{s}_j)$, $\forall j \in \mathcal{N}_p$ using stored distances and model parameters.
3: Each sensor node $i \in \mathcal{N}_p$ assigns values for all elements in row $\mathbf{r}_i$:
$\mathbf{r}_i(1:p) \leftarrow [\overline{\gamma}(\mathbf{s}_i, \mathbf{s}_1), \overline{\gamma}(\mathbf{s}_i, \mathbf{s}_2), \ldots, \overline{\gamma}(\mathbf{s}_i, \mathbf{s}_p)]$
$\mathbf{r}_i(p+1:p+2) \leftarrow [\overline{\gamma}(\mathbf{s}_i, \mathbf{s}_0), 1]$
4: Closest sensor node, $m \in \mathcal{N}_p$, to $\mathbf{s}_0$ initiates the iteration and creates an additional row:
$\mathbf{r}_1(1:p+2) \leftarrow [1, 1, \ldots, 1]$
5: Sensor node m sends row $\mathbf{r}_1$ to each sensor node $j \in \mathcal{N}_p$.
6: Each sensor node j receive row $\mathbf{r}_1$ and update its row $\mathbf{r}_j$ by:
$\mathbf{r}_j \leftarrow \mathbf{r}_j - \mathbf{r}_j(1) \times \mathbf{r}_1$
7: Sensor node $i \in \mathcal{N}_p$ sends its row $\mathbf{r}_i$ to each sensor node $j \in \mathcal{N}_p$.
8: Each sensor node $j \in \mathcal{N}_p$ updates its $\mathbf{r}_j$ by:
$\mathbf{r}_j \leftarrow \mathbf{r}_j - \mathbf{r}_i / \mathbf{r}_j(i)$
and sensor node m also updates row $\mathbf{r}_1$ by:
$\mathbf{1} \leftarrow \mathbf{1} - \mathbf{r}_i / \mathbf{1}(i)$
9: Steps 7-10 are repeated for $p+1$ iterations. Then, sensor node m sends row $\mathbf{r}_m$.
10: A back-substitution is performed from sensor node m. Hence, weights $w_i(\mathbf{s}_0)|_p$ for each sensor node i and $\mathcal{L}$ are obtained.

where $\mathcal{N}_{\mathbf{s}_0}$ is the set of one-hop, in-range sensor nodes that can be found in $\mathbf{s}_0$ neighborhood. Sensor node j receives the packet and updates the EV value with the received data. As a result, sensor nodes $i, j \in \mathcal{N}_p$ form the initial cluster, where $\mathcal{N}_p$ is the set of sensor nodes included in the initial cluster. The packet is iteratively relayed between the neighborhood of $\mathcal{N}_p$, until the next closest sensor node to $\mathbf{s}_0$, $k \in \mathcal{N}_{\mathbf{s}_0}$, is found. This process continues until an initial cluster with p nodes is formed and the EV is computed between all the p sensor nodes. Once the EV is obtained, each sensor node performs semivariogram modeling to obtain the model parameters. In our case, since $p = 3$, the iterative process is terminated when sensor nodes $i, j, k \in \mathcal{N}_p$ are found.

b). Distributed Kriging prediction: After the parameter estimation, the sensor nodes in the initial cluster are able to predict $\hat{y}(\mathbf{s}_0)|_p$ by solving the Kriging estimator (14) and Kriging system (16) in a distributed way. Each sensor node in the cluster $i \in \mathcal{N}_p$ creates one row $\mathbf{r}_i$ of the Kriging system consisting of the semivariogram $\overline{\Gamma}(\mathbf{s}_i, \mathbf{s}_j)$ between all the sensor nodes in the cluster and the semivariogram $\overline{\gamma}(\mathbf{s}_i - \mathbf{s}_0)$ to the target location $\mathbf{s}_0$. The Kriging system is solved by the modified Gaussian elimination method as described in Algorithm 2. We adapted the algorithm from [33], to work with the cluster of sensor nodes. The closest sensor node to $\mathbf{s}_0$, $m \in \mathcal{N}_p$, is chosen to begin the iterative process. The sensor node m also constructs an additional array row of ones $\mathbf{r}_1$ of size $p + 2$ and sends it to each sensor node in the cluster $j \in \mathcal{N}_p$. Each sensor node j updates its row $\mathbf{r}_j$ by substracting its stored row with the received row $\mathbf{r}_1$. Note that the content of row $\mathbf{r}_1$ depends on the type of Kriging variant employed. Following to the initial iteration, each sensor node $i \in \mathcal{N}_p$ sends its row $\mathbf{r}_i$ to each sensor node $j \in \mathcal{N}_p$ and

updates its row elements by substracting the stored row from the received row. As a result of iterations, the weight $w_i(\mathbf{s}_0)|_p$ for each sensor node and Lagrange multiplier is obtained. The gaussian elimination method is implemented in a iterative way by local computation and the exchange of rows between the sensor nodes in the cluster.

At the end, the Kriging estimate $\hat{y}(\mathbf{s}_0)|_p$ can be obtained from the Kriging estimator (14) in a distributed way i.e., by multiplying the weight of each sensor node with its measurement and summing at each sensor node.

2) Adaptive Cluster Estimation: After the initial estimation phase, an initial cluster is formed by p sensor nodes within the range of the unmeasured location and, the local semivariogram and kriged estimate are obtained. With this approach, the sensor nodes in the cluster which are highly correlated to the estimation point are given all the weights whereas the non-cluster sensor nodes are neglected. The quality of the field estimation can be further improved by incorporating one or more sensor nodes to the initial cluster of p sensor nodes. However, because of the local nature of Kriging, adding more sensor nodes to the initial cluster does not guarantee the best estimation value. Furthermore, adding more sensor nodes also increases the computational complexity. Thus, a metric with better trade-off between the size and quality is necessary. To this end, a distributed adaptive clustering (DCA) is proposed, which employs the OK variance in combination with the Kriging system to optimize the cluster size. The

Algorithm 3 Distributed adaptive clustering (DAC)

p: Initial number of sensor nodes in the cluster // n: Total number of sensor nodes in the cluster after updating process // $\hat{y}(\mathbf{s}_0)|_n$: Final field estimate value obtained through clustering // $\mathbf{s}_0$: location where field value is not known // t: Number of sensor nodes added to the initial cluster // d_c: Range parameter of semivariogram model // $d(\mathbf{s}_i, \mathbf{s}_0)$: Euclidean distance between $\mathbf{s}_i$ and $\mathbf{s}_0$.

1: **for** all $\mathbf{s}_0$ **do**
2:    **for** $n = p$ **to** $p + t - 1$ **do**
3:       Compute $\hat{\gamma}(\mathbf{h})|_n$ and $\hat{\gamma}(\mathbf{h})|_{n+1}$.
4:       Estimate field values $\hat{y}(\mathbf{s}_0)|_n$ and $\hat{y}(\mathbf{s}_0)|_{n+1}$.
5:       Kriging variances $\sigma^2(\mathbf{s}_0)|_n$ and $\sigma^2(\mathbf{s}_0)|_{n+1}$ are computed by using equation (18).
6:       **if** $\sigma^2(\mathbf{s}_0)|_n \leq \sigma^2(\mathbf{s}_0)|_{n+1}$ **then**
7:          Terminate update process and cluster formation.
8:          Field estimate value at location $\mathbf{s}_0$ and Kriging variance are obtained.
            $\hat{y}(\mathbf{s}_0) = \hat{y}(\mathbf{s}_0)|_n$
            $\sigma^2(\mathbf{s}_0) = \sigma^2(\mathbf{s}_0)|_n$
9:       **else**
10:         A new sensor node is added to the cluster and process is restarted from line 3
            $n = n + 1$
11:       **end if**
12:    **end for**
13: **end for**

clustering procedure for shadowing estimation is explained in Algorithm 3.

Subsequent to the initial estimation of $\hat{y}(\mathbf{s}_0)|_p$, the Kriging variance $\sigma^2(\mathbf{s}_0)|_p$ is computed. Note that the Kriging variance given by equation (18) can be computed distributively by summing the local multiplications with the Lagrange multiplier. A new sensor node $q \in \mathcal{N}_{\mathbf{s}_0}$ is added to the initial cluster, and the estimate $\hat{y}(\mathbf{s}_0)|_{p+1}$ and the Kriging variance $\sigma^2(\mathbf{s}_0)|_{p+1}$ are obtained. The Kriging estimate and the weights are quickly updated when one sensor node is added to the cluster by the following set of equations [53]:

$$\hat{y}(\mathbf{s}_0)|_{p+1} = \hat{y}(\mathbf{s}_0)|_p - w_q(\mathbf{s}_0)_{p+1}[\hat{y}(\mathbf{s}_q)|_p - y(\mathbf{s}_q)], \quad (19)$$

$$w_i(\mathbf{s}_0)|_{p+1} = w_{i|p}(\mathbf{s}_0) - w_q(\mathbf{s}_0)_{p+1}w_i(\mathbf{s}_q)|_p, i = 1, 2, \ldots, p, \quad (20)$$

where $\hat{y}(\mathbf{s}_0)|_{p+1}$ is the estimate at $\mathbf{s}_0$ using $p + 1$ sensor nodes located at $\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_p, \mathbf{s}_q$, $w_q(\mathbf{s}_0)_{p+1}$ is the weight assigned to sensor node q when predicting $y(\mathbf{s}_0)|_{p+1}$, $\hat{y}(\mathbf{s}_q)|_p$ is the Kriging estimate at $\mathbf{s}_q$ from sensor nodes located at $\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_p$ and $y(\mathbf{s}_q)$ is the measurement of sensor node q at location $\mathbf{s}_q$.

The Kriging variance of clusters with p and $p + 1$ sensor nodes are compared. If $\sigma^2(\mathbf{s}_0)|_p$ is greater than $\sigma^2(\mathbf{s}_0)|_{p+1}$, the new sensor node q is added to the initial cluster to reduce the Kriging variance. This process is iteratively repeated with a total number of sensor nodes $t \geq 1$ until the resulting variance $\sigma^2(\mathbf{s}_0)|_{p+t+1}$ is higher or equal to the one with $\sigma^2(\mathbf{s}_0)|_{p+t}$. Note, that d_c is estimated from semivariogram modeling. As

a result, a cluster is formed with $n = p + t$ sensor nodes. Each sensor node is successively considered from closest to farthest to the unmeasured location $\mathbf{s}_0$ and included in the cluster if the above condition is satisfied. The cluster is not only formed by the least number of sensor nodes but also by the best set of sensor nodes giving the best estimation value in the area.

V. COMPLEXITY ANALYSIS

In this section, we analyze the computational complexity of the DICA-RK at each stage of its operation.

- *DSA* involves the exchange of information among the sensor nodes and local computations in each sensor node. The local computations are basically the distance and semivariogram calculations, which are either integer sums or multiplications. Such computations can be neglected since they have lower complexity when compared to transmission. Semivariogram complexity substantially depends on the number of iterations. Hence, the cost to implement DSA is $cost_{DSA} \in \mathcal{O}(n)$.
- *DKPA* consists of solving the Kriging system using the Gaussian elimination method by creating a row at each sensor node by applying the semivariogram model and the exchange of rows between the sensor nodes. The complexity of the Gaussian elimination method is widely known, which is $cost_{gauss} \in \mathcal{O}(n^3)$. Moreover, obtaining the Kriging estimate by equation (14) shows low complexity, since it is computed by local multiplication and a single packet transmission by each sensor node. Therefore, the cost of DKPA is $cost_{DKPA} \in \mathcal{O}(n^3)$.
- In *DAC*, a cluster is formed with $n = p + t$ sensor nodes, where p is the initial set of sensor nodes to begin the estimation and t is the number of additional sensor nodes incorporated to improve the quality. Spatial estimation $\hat{y}(\mathbf{s}_0)|_p$ using the initial cluster requires one operation. However, $\hat{y}(\mathbf{s}_0)|_t$ requires t operations since the estimation is performed for every new sensor node addition. Thus, the adaptive cluster estimation requires $t + 1$ times the local semivariogram and the Kriging operations.

Finally, the complexity of the DICA-RK is $|t + 1| \times (cost_{DSA} + cost_{DKPA} + cost_{DAC}) \in \mathcal{O}(N + n^3)$. If the estimates are required at m unmeasured locations, the complexity further increases to $\mathcal{O}(m(N + n^3))$.

VI. SIMULATION RESULTS

To assess the performance of the proposed DICA-RK, we consider a LTE-sensor network in a sub-urban environment in a square area of 200 m $\times$ 200 m. The LTE picocell eNB is placed at location $\mathbf{s} = [0,0]^{\mathrm{T}}$ and the WSN consisting of N sensor nodes with a minimum 2 m inter-node spacing are randomly and uniformly deployed. We simulate a realistic radio environment based on the propagation model (2) and compute the received signal power that could be sensed by the sensor nodes at any location. The key parameters for the simulation scenario are presented in Table I [54], [55]. The complete radio map is obtained by performing predictions in a fine grid of $N_L = 2601$ locations, when considering a 4 m
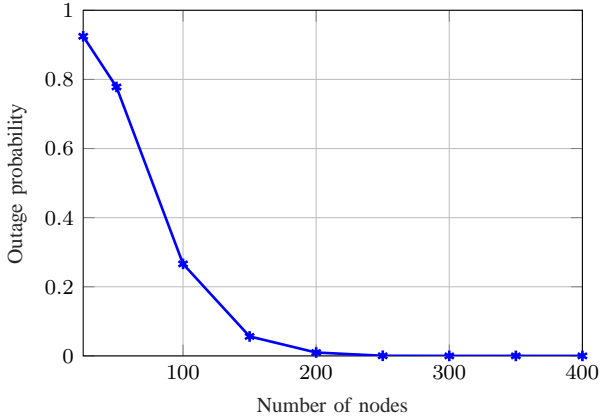
Fig. 4. Outage probability versus number of sensor nodes

resolution grid on the square area under study. The DICA-RK was tested for estimating locations sequentially and randomly. The generated spatial maps revealed that the estimation quality remained the same irrespective of the estimation sequence. Note that in the proposed DICA-RK, a single sensor node can contribute to the estimation at multiple points depending on its location. This means that the clusters can overlap.

TABLE I
SIMULATION PARAMETER VALUES

LTE standard parameter	value
Field dimension	200 m × 200 m
Picocell BS Transmission power	24 dBm
Picocell BS antenna height	5 m
Picocell BS Carrier frequency	2000 MHz
802.15.4 Transmission power	-10 dBm
802.15.4 RX sensitivity	-90 dBm
802.15.4 Base frequency	2400 MHz
Path-loss exponent	3
Shadow fading standard deviation	6 dB
Correlation distance of shadowing	10 m
Path-loss for 1 m distance	38 dB

Important factors for implementing the DICA-RK are the network size and the initial cluster size. Based on the model (9) presented in section III-B, our algorithm requires a minimum of three samples to begin the semivariogram modeling. This means that an average of $p = 3$ sensor nodes must cover each estimation point. Hence, we define the outage probability as the probability that each spatial location $\mathbf{s}_0$ is not in the communication range of three sensor nodes for initial cluster formation. Fig. 4 illustrates the outage probability for different WSNs sizes in the scenario under consideration.

In order to demonstrate the cluster formation procedure, we consider the unmeasured location $\mathbf{s}_0 = [0, 36]^{\mathrm{T}}$ of Fig. 5(a) as an example. An estimation is started by an initial set $p = 3$ composed by sensor nodes $\{V_1, V_2, V_3\}$, which are closest to $\mathbf{s}_0$. After applying the adaptive cluster estimation procedure based on minimizing the Kriging variance, sensor nodes $\{V_4\}$, $\{V_5\}$ and $\{V_6\}$ are added to the initial cluster. This procedure can be seen in Fig. 5(b), which illustrates how
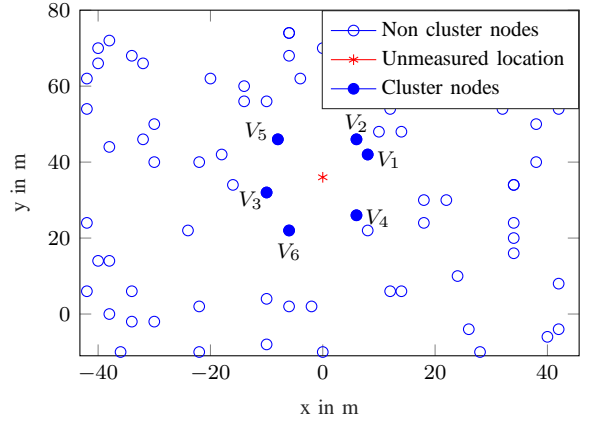
(a) Fig. 1 zoomed to show cluster formation
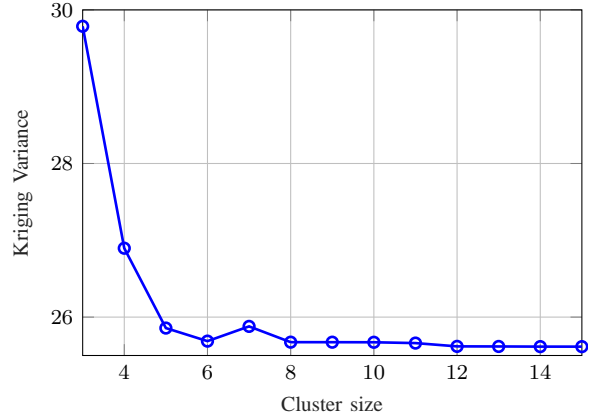
(b) Kriging variance versus cluster size

Fig. 5. An example of cluster formation procedure at $\mathbf{s}_0 = [0, 36]^{\mathrm{T}}$

Kriging variance changes as new sensor nodes are added to the initial cluster. As seen from the Fig. 5(b), adding new sensor nodes could reduce the Kriging variance further. However, the computational complexity also increases geometrically with the number of sensor nodes. Therefore, in this case, a cluster is formed with 6 sensor nodes without having a significant impact on the prediction quality and precision. Note that the above procedure is the same for estimating every unmeasured location. This implies that the size of each cluster changes depending on the field behavior and especially, the Kriging variance.

The accuracy of prediction is an important criteria to benchmark the performance of spatial interpolation methods. The prediction quality depends on the sensor node deployment, the quality of measurements and the number of sensor nodes. In order to analyze the accuracy, we consider the mean squared error (MSE) between the ground truth $z(\mathbf{s}_{0,l})$ and the predicted mean value $\hat{z}(\mathbf{s}_{0,l})$ for each spatial location l, where $l = \{1, 2, ..., N_L\}$. It is defined as,

$$\mathrm{MSE} = \frac{1}{N_L} \sum_{l=1}^{N_L} (z(\mathbf{s}_{0,l}) - \hat{z}(\mathbf{s}_{0,l}))^2 \qquad (21)$$

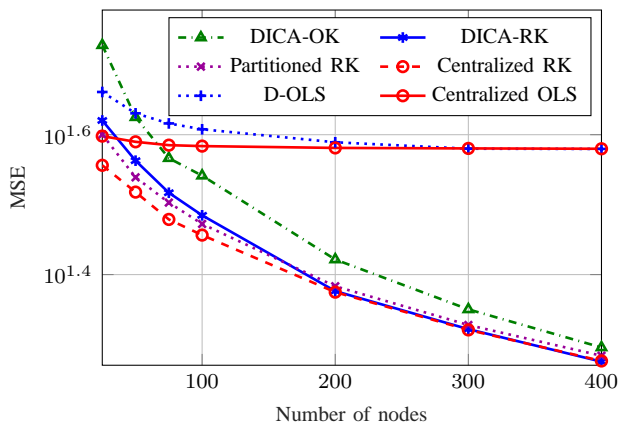In Fig. 6, the DICA-RK is compared with a centralized RK and a partitioned RK, and study their effect on the

Fig. 6. Comparison of DICA-RK with centralized RK and partitioned RK

Fig. 7. Comparison of DICA-RK with classical interpolation methods

reconstruction quality. In centralized RK, sensor nodes send their measurements and location information to a central node for global field reconstruction. On the other hand, in partitioned RK, the total area is divided into subregions and a subregion head performs local field reconstruction, using only the measurements and positions from the subregion. It is similar to the centralized case except that the global problem is divided into smaller problems. The solution to the prediction problem is known only by the subregion heads. DICA-OK stands for distributed incremental clustering algorithm based on OK and is proposed in our previous work [34]. The mean MSE is averaged over 50 realizations of the channel field. Note that this average accounts for the mobility of the sensor nodes located in vehicles. The trend of the MSE plot proves that the quality of prediction grows with the number of measurements. This is because the sensors are able to exploit the spatial correlation effectively, when there are more number of measurements. It is worth to notice from the Fig. 6, that the MSE for DICA-RK is lower than the DICA-OK and partitioned Kriging, and converges to the centralized case for networks with $N > 200$. On the contrary, for $N < 200$, the algorithm suffers from lack of sensor nodes within the communication range to build clusters and thereby, fails to exploit spatial correlation. In case of partitioned Kriging, since the inter-region sensor node sharing is not allowed, the cluster heads have less information at the borders. Therefore, the prediction quality deteriorates at the borders, which can be seen in the interpolated map of Fig. 8(c). Finally, we remark that for path-loss only prediction using OLS, i.e., centralized OLS or D-OLS, the MSE is significantly higher than other methods that consider path-loss plus shadowing, highlighting the importance of predicting shadowing correlation in wireless channels.

Vehicle mobility is subject to erroneous location measurements. Hence, we analyse the impact of location uncertainty on the prediction. Sensor nodes are assumed to obtain noisy location measurements $\mathbf{S}_i = \mathbf{S}_i + \mathbf{L}_i$, where $L_i \sim \mathcal{N}(0, \sigma^2)$ is zero mean additive white Gaussian noise with the location error standard deviation σ. We draw σ from an exponential distribution, i.e., $\sigma \overset{iid}{\sim} \exp(\lambda)$, where λ is the average error location standard deviation. We consider true location with
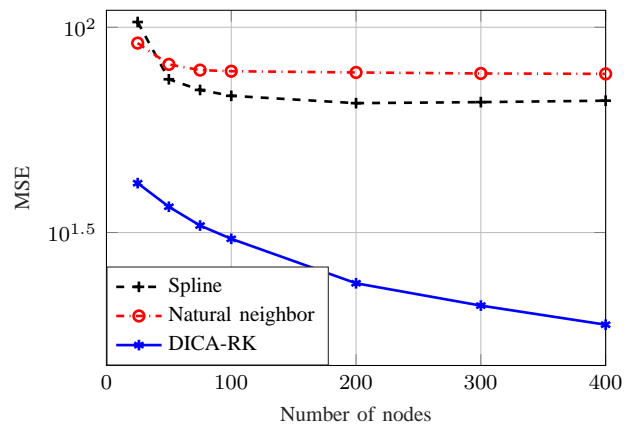
$\lambda = 6\,\mathrm{m}$ and $\lambda = 0\,\mathrm{m}$ for estimated location N_L. It can be seen in the Table II how the performance degrades under the impact of location noise.

TABLE II
PERFORMANCE UNDER LOCATION UNCERTAINTY.

WSN size	50	100	200	300	400
MSE DICA-RK	36.54	30.52	23.78	20.98	18.86
MSE DICA-RK under location uncertainty	38.60	33.95	28.49	26.35	25.13

We also compare the performance of DICA-RK with two classical interpolation methods such as natural neighbor and spline. Fig. 7 demonstrates that MSE for the classical interpolation methods is significantly higher than for the DICA-RK. To illustrate the excellent reconstruction quality of our algorithm, we present interpolated maps in Fig. 8, where all the interpolation methods presented are visually compared. An important aspect to notice is the performance of DICA-RK with $N = 50$, where predictions rely largely on path-loss only.

VII. CONCLUSION

In this paper, we proposed a novel distributed incremental clustering algorithm for radio map reconstruction as a powerful tool to realize the vision of 5G automotive. The algorithm minimizes the total number of sensor measurements required for radio map reconstruction through distributed processing and clustering of sensor nodes. The complexity of Kriging is significantly reduced while retaining its excellent prediction quality. The Kriging variance used for data screening has proved to offer a good trade-off between quality and complexity. Simulation results highlight the fact that both path-loss and shadowing components are important in wireless channel prediction. In terms of prediction quality, regression Kriging leads to a superior performance than the plain regression and standard OK. Our future work aims to apply the proposed algorithm to real field measurements. In addition, mobility patterns will be included to account for the temporal correlation of the field.
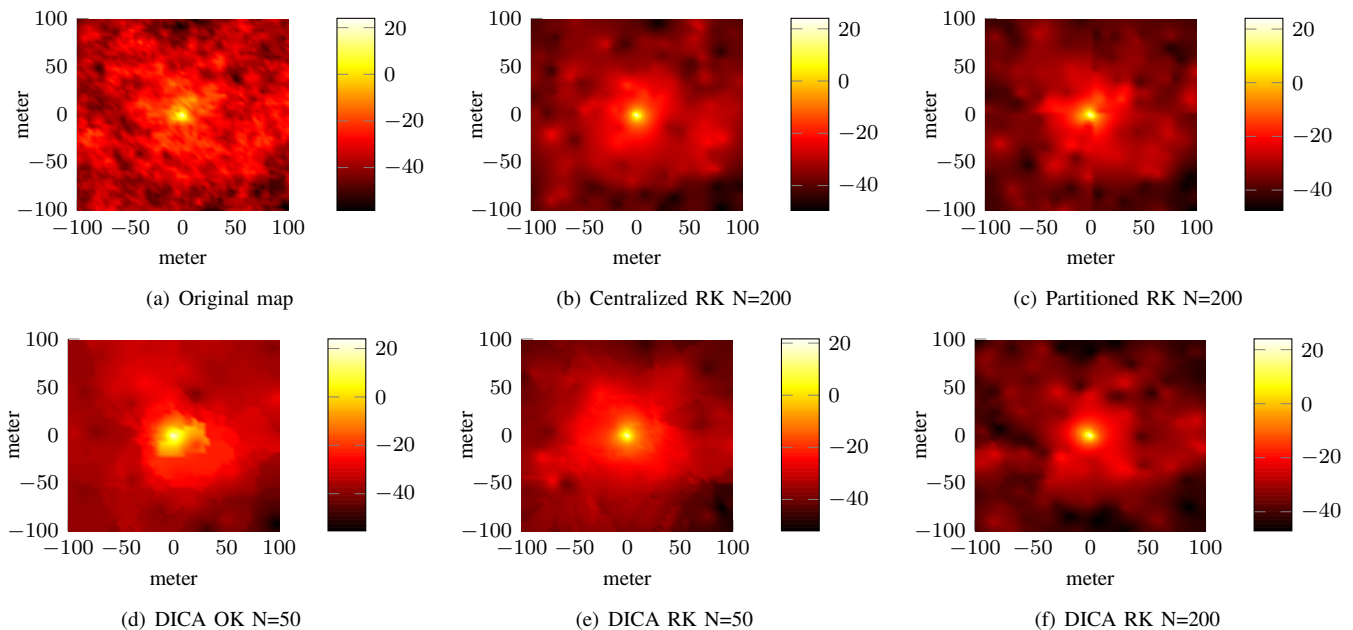
ACKNOWLEDGMENT

Fig. 8. Interpolated maps for different estimation frameworks

REFERENCES

[1] "5G Vision: The Next Generation of Communication Networks and Services," 5G Infrastructure Public Private Partnership (5G-PPP), Tech. Rep., Feb 2015.

[2] "The Road to 5G: Drivers, Applications, Requirements and Technical Development," Global mobile Suppliers Association (GSA), Tech. Rep., Nov 2015.

[3] "Understanding 5G: Perspectives on future technological advancements in mobile," GSMA Intelligence, Tech. Rep., Dec 2014.

[4] "5G Automotive Vision," 5G Infrastructure Public Private Partnership (5G-PPP), Tech. Rep., Oct 2015.

[5] H. Abou-zeid, H. S. Hassanein, Z. Tanveer, and N. AbuAli, "Evaluating mobile signal and location predictability along public transportation routes," in *2015 IEEE Wireless Communications and Networking Conference (WCNC)*, March 2015, pp. 1195–1200.

[6] N. Barman, S. Valentin, and M. G. Martini, "Predicting link quality of wireless channel of vehicular users using street and coverage maps," in *2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Sept 2016, pp. 1–6.

[7] "5G: New Air Interface and Radio Access Virtualization," Huawei white paper, Tech. Rep., April 2015.

[8] M. Höyhtyä, A. Mämmelä, M. Eskola, M. Matinmikko, J. Kalliovaara, J. Ojaniemi, J. Suutala, R. Ekman, R. Bacchus, and D. Roberson, "Spectrum Occupancy Measurements: A Survey and Use of Interference Maps," *IEEE Communications Surveys Tutorials*, vol. 18, no. 4, pp. 2386–2414, Fourthquarter 2016.

[9] H. B. Yilmaz, T. Tugcu, F. Alagz, and S. Bayhan, "Radio Environment Map as Enabler for Practical Cognitive Radio Networks," *IEEE Communications Magazine*, vol. 51, no. 12, pp. 162–169, Dec 2013.

[10] S. C. Lin and K. C. Chen, "Spectrum-Map-Empowered Opportunistic Routing for Cognitive Radio Ad Hoc Networks," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 6, pp. 2848–2861, July 2014.

[11] J. Perez-Romero, A. Zalonis, L. Boukhatem, A. Kliks, K. Koutlia, N. Dimitriou, and R. Kurda, "On the Use of Radio Environment Maps for Interference Management in Heterogeneous Networks," *IEEE Communications Magazine*, vol. 53, no. 8, pp. 184–191, August 2015.

[12] A. Galindo-Serrano, B. Sayrac, S. B. Jemaa, J. Riihijärvi, and P. Mähönen, "Harvesting MDT Data: Radio Environment Maps for Coverage Analysis in Cellular Networks," in *8th International Conference on Cognitive Radio Oriented Wireless Networks*, July 2013, pp. 37–42.

[13] B. Sayrac, A. Galindo-Serrano, S. B. Jemaa, J. Riihijärvi, and P. Mähönen, "Bayesian Spatial Interpolation as an Emerging Cognitive Radio Application for Coverage Analysis in Cellular Networks," *Transactions on Emerging Telecommunications Technologies*, vol. 24, no. 7-8, pp. 636–648, 2013.

[14] H. Braham, S. B. Jemaa, G. Fort, E. Moulines, and B. Sayrac, "Spatial Prediction Under Location Uncertainty in Cellular Networks," *IEEE Transactions on Wireless Communications*, vol. 15, no. 11, pp. 7633–7643, Nov 2016.

[15] G. Fodor, E. Dahlman, G. Mildh, S. Parkvall, N. Reider, G. Mikls, and Z. Turnyi, "Design Aspects of Network Assisted Device-to-Device Communications," *IEEE Communications Magazine*, vol. 50, no. 3, pp. 170–177, Mar 2012.

[16] V. P. Chowdappa, M. Fröhle and H. Wymeersch and C. Botella, "Distributed channel prediction for multi-agent systems," in *IEEE International Conference on Communications (ICC)*, May 2017.

[17] H. Abou-zeid, H. S. Hassanein, and S. Valentin, "Optimal Predictive Resource Allocation: Exploiting Mobility Patterns and Radio Maps," in *2013 IEEE Global Communications Conference (GLOBECOM)*, Dec 2013, pp. 4877–4882.

[18] N. Bui, M. Cesana, S. A. Hosseini, Q. Liao, I. Malanchini, and J. Widmer, "Anticipatory Networking in Future Generation Mobile Networks: a Survey," *CoRR*, vol. abs/1606.00191, 2016. [Online]. Available: http://arxiv.org/abs/1606.00191

[19] R. C. Dwarakanath, J. D. Naranjo, and A. Ravanshid, "Modeling of Interference Maps for Licensed Shared Access in LTE-advanced Networks Supporting Carrier Aggregation," in *2013 IFIP Wireless Days (WD)*, Nov 2013, pp. 1–6.

[20] S. Üreten, A. Yongaçoğlu, and E. Petriu, "A comparison of Interference Cartography Generation Techniques in Cognitive Radio Networks," in *2012 IEEE International Conference on Communications (ICC)*, June 2012, pp. 1879–1883.

[21] M. Molinari, M.-R. Fida, M. K. Marina, and A. Pescape, "Spatial Interpolation Based Cellular Coverage Prediction with Crowdsourced Measurements," in *Proceedings of the 2015 ACM SIGCOMM Workshop on Crowdsourcing and Crowdsharing of Big (Internet) Data*, ser. C2B(1)D '15. New York, NY, USA: ACM, 2015, pp. 33–38.

[22] H. B. Yilmaz and T. Tugcu, "Location Estimation-based Radio Environment Map Construction in Fading Channels," *Wireless Communications and Mobile Computing*, vol. 15, no. 3, pp. 561–570, 2015.

[23] J. Ojaniemi, J. Kalliovaara, A. Alam, J. Poikonen, and R. Wichman, "Optimal Field Measurement Design for Radio Environment Mapping," in *2013 47th Annual Conference on Information Sciences and Systems (CISS)*, Mar 2013, pp. 1–6.

[24] K. Connelly, Y. Liu, D. Bulwinkle, A. Miller, and I. Bobbitt, "A Toolkit for Automatically Constructing Outdoor Radio Maps," in *International Conference on Information Technology: Coding and Computing (ITCC'05) - Volume II*, vol. 2, April 2005, pp. 248–253 Vol. 2.

[25] D. Denkovski, V. Atanasovski, L. Gavrilovska and J. Riihijärvi and P. Mähönen, "Reliability of a radio environment map: Case of spatial interpolation techniques," in *2012 7th International ICST Conference*

on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM), June 2012, pp. 248–253.

[26] N. Cressie, *Statistics for Spatial Data*. Wiley Series in Probability and Statistics, 1993.

[27] R. D. Taranto, S. Muppirisetty, R. Raulefs, D. Slock, T. Svensson, and H. Wymeersch, "Location-Aware Communications for 5G Networks: How Location Information Can Improve Scalability, Latency, and Robustness of 5G," *IEEE Signal Processing Magazine*, vol. 31, no. 6, pp. 102–112, Nov 2014.

[28] M. Angjelicinoski, V. Atanasovski, and L. Gavrilovska, "Comparative Analysis of Spatial Interpolation Methods for Creating Radio Environment Maps," in *2011 19thTelecommunications Forum (TELFOR) Proceedings of Papers*, Nov 2011, pp. 334–337.

[29] C. Phillips, M. Ton, D. Sicker, and D. Grunwald, "Practical radio environment mapping with geostatistics," in *2012 IEEE International Symposium on Dynamic Spectrum Access Networks*, Oct 2012.

[30] A. B. H. Alaya-Feki, S. B. Jemaa, B. Sayrac, P. Houze, and E. Moulines, "Informed Spectrum Usage in Cognitive Radio Networks: Interference Cartography," in *2008 IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications*, Sept 2008, pp. 1–5.

[31] J. Riihijärvi and P. Mähönen, "Estimating Wireless Network Properties with Spatial Statistics and Models," in *2012 10th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt)*, May 2012, pp. 331–336.

[32] J. D. Naranjo, A. Ravanshid, I. Viering, R. Halfmann, and G. Bauch, "Interference Map Estimation using Spatial Interpolation of MDT reports in Cognitive Radio Networks," in *2014 IEEE Wireless Communications and Networking Conference (WCNC)*, Apr 2014, pp. 1496–1501.

[33] G. Hernández-Peñaloza and B. Beferull-Lozano, "Field Estimation in Wireless Sensor Networks using Distributed Kriging," in *2012 IEEE International Conference on Communications (ICC)*, June 2012.

[34] V. P. Chowdappa, C. Botella, and B. Beferull-Lozano, "Distributed Clustering Algorithm for Spatial Field Reconstruction in Wireless Sensor Networks," in *2015 IEEE 81st Vehicular Technology Conference (VTC Spring)*, May 2015, pp. 1–6.

[35] M. Umer, L. Kulik, and E. Tanin, "Spatial interpolation in wireless sensor networks: Localized algorithms for variogram modeling and kriging," *GeoInformatica*, vol. 14, no. 1, p. 101, 2009.

[36] E. Dall'Anese, S. J. Kim, and G. B. Giannakis, "Channel Gain Map Tracking via Distributed Kriging," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 3, pp. 1205–1211, Mar 2011.

[37] E. Uhlemann, "Initial steps toward a cellular vehicle-to-everything standard [connected vehicles]," *IEEE Vehicular Technology Magazine*, vol. 12, no. 1, pp. 14–19, March 2017.

[38] T. Hengl, G. B. Heuvelink, and A. Stein, "A Generic Framework for Spatial Prediction of Soil Variables based on Regression-Kriging," *Geoderma*, vol. 120, no. 12, pp. 75 – 93, 2004.

[39] H. G. Hengl, T and A. Stein, *Comparison of Kriging with External Drift and Regression Kriging*. ITC, 2003.

[40] M. Malmirchegini and Y. Mostofi, "On the Spatial Predictability of Communication Channels," *IEEE Transactions on Wireless Communications*, vol. 11, no. 3, pp. 964–978, Mar 2012.

[41] M. Gudmundson, "Correlation Model for Shadow Fading in Mobile Radio Systems," *Electronics Letters*, vol. 27, no. 23, pp. 2145–2146, 1991.

[42] Jin Li and Andrew D. Heap, "Spatial Interpolation Methods Applied in the Environmental Sciences: A Review," *Environmental Modelling and Software* , vol. 53, pp. 173 – 189, 2014.

[43] M.A. Oliver and R. Webster, "A Tutorial Guide to Geostatistics: Computing and Modelling Variograms and Kriging," *CATENA* , vol. 113, pp. 56 – 69, 2014.

[44] V. P. Chowdappa, C. Botella, Sara Santos Sáez, J. Javier Samper, Rafael J. Martínez, "Low complexity distributed cluster based algorithm for spatial prediction," in *International Wireless Communications and Mobile Computing Conference (IWCMC)*, June 2017.

[45] V. Erceg, L. J. Greenstein, S. Y. Tjandra, S. R. Parkoff, A. Gupta, B. Kulic, A. A. Julius, and R. Bianchi, "An Empirically Based Path Loss Model for Wireless Channels in Suburban Environments," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 7, pp. 1205–1211, Jul 1999.

[46] S. S. Ghassemzadeh, L. J. Greenstein, A. Kavcic, T. Sveinsson, and V. Tarokh, "UWB Indoor Path Loss Model for Residential and Commercial Buildings," in *2003 IEEE 58th Vehicular Technology Conference. VTC 2003-Fall*, vol. 5, Oct 2003, pp. 3115–3119.

[47] J. M. Lebreton, N. M. Murad, and R. Lorion, "Radio Frequency Mapping using an Autonomous Robot: Application to the 2.4 GHz Band," *IOP Conference Series: Materials Science and Engineering*, vol. 120, 2016.

[48] N. Jalden, "Analysis and Modelling of Joint Channel Properties from Multi-site, Multi-Antenna Radio Measurements," Ph.D. dissertation, KTH, 2010.

[49] O. Schabenberger and C. A. Gotway, *Statistical Methods for Spatial Data Analysis*. Chapman & Hall/CRC, 2004.

[50] X. Jian, R. A. Olea, and Y.-S. Yu, "Semivariogram Modeling by Weighted Least Squares," *Computers and Geosciences*, vol. 22, no. 4, pp. 387–397, Feb 1996.

[51] N. Cressie, "Fitting Variogram Models by Weighted Least Squares," *Journal of the International Association for Mathematical Geology*, vol. 17, no. 5, pp. 563–586, Jul 1985.

[52] F. S. Cattivelli, C. G. Lopes, and A. H. Sayed, "Diffusion Recursive Least-Squares for Distributed Estimation overAadaptive Networks," *IEEE Transactions on Signal Processing*, vol. 56, no. 5, pp. 1865–1877, May 2008.

[53] X. Emery, "The Kriging Update Equations and their Application to the Selection of Neighboring Data," *Computational Geosciences*, vol. 13, no. 3, pp. 269–280, 2009.

[54] "Universal Mobile Telecommunications System (UMTS); Base Station radio transmission and reception (FDD)," 3rd Generation Partnership Project (3GPP), Tech. Rep. 3GPP TS 25.104 V12.5.0 Release 12, 2015.

[55] "IEEE 802.15.4 TelosB Mote with sensor suit," Crossbow Technology, Tech. Rep. TPR2420CA, 2005.

Vinay-Prasad Chowdappa received his MSc degree with distinction honors from University of Southampton, UK in 2011. He is currently working towards his PhD degree in the Institute of Robotics and Information & Communication Technologies (IRTIC) at the University of Valencia, Spain. In 2016 and 2017, he was a visiting researcher in the Communications Systems group, Chalmers University of Technology, Sweden. His research interests include algorithm design, statistical inference and machine learning in wireless networks.

Carmen Botella (S03, M09, SM17) received her M.Sc. and Ph.D. degrees in Telecommunications engineering from Universidad Politécnica de Valencia, Spain, in 2003 and 2008, respectively. In 2009 and 2010, she was a postdoctoral researcher in the Communications Systems and Information Theory group, Chalmers University of Technology, Sweden. In 2011, she joins the University of Valencia where she is currently an Associate Professor. Her research interests include the general areas of coordination and cooperation in wireless systems, with special focus on solutions for 5G and beyond.

J. Javier Samper received the Ph.D degree in Computer Science Engineering from Universidad de Valencia, Spain, in 2005. He is currently an Associate Professor at the School of Engineering, carrying out his research at LISITT laboratory at the Institute of Robotics and TICs (IRTIC). He has participated in the EU projects from the beginning of its research work. He works in the areas of software engineering and consulting in the Intelligent Transportation Systems (ITS) sector. He is currently the director of the Master of Telecommunications Engineering (UVEG) and president of EATIS.org (Euro-American Conference on Telematics and Information Systems organization).

Rafael J. Martínez holds a degree in Computer Systems from the Polytechnic University of Valencia (1991) and a PhD in Computer Engineering from the Universitat de Valncia (1997). Since 2000, he has been a Professor in the Area of Architecture and Technology of Computers and teaches in degrees and masters of the Technical School of Engineering (ETSE), in subjects related to Operating Systems, Fault Tolerant Systems, Computer Graphics and Simulation.
Rafael Martínez is director of the Laboratory of Simulation and Modeling (LSyM), both belonging to the University of Valencia. His main interests are the study, development and integration of simulators for operator training. He is a member of ACM and Eurographics.