Citation for the original published paper:
V. Chowdappa, C. Botella, S. Santos Sáez, J. J. Samper and R. J. Martínez, "A low complexity distributed cluster based algorithm for spatial prediction," 2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC), 2017, pp. 2158-2162, doi: 10.1109/IWCMC.2017.7986617.

A Low Complexity Distributed Cluster based Algorithm for Spatial Prediction

Vinay-Prasad Chowdappa, Carmen Botella, Sara Santos Sáez, J. Javier Samper, Rafael J. Martínez
Instituto de Robótica y Tecnologías de la Información y Comunicación (IRTIC), Universidad de Valencia, Spain
Email: {vinay.chowdappa, Carmen.Botella, jose.j.samper, Rafael.Martinez}@uv.es, sansasa@alumni.uv.es

***Abstract*—Radio Environment Maps (REM) can be an essential tool for numerous applications in future 5G wireless networks. In this paper, we employ a popular geo-statistical method called ordinary kriging to estimate the REM of an area covered by a eNodeB equipped with multiple antennas. Wireless sensors are distributed over the area of interest and adaptive clusters of sensors are arranged in order to improve the quality of the estimation. In this paper, we modify the distributed clustering algorithm proposed in a previous work to reduce the complexity of kriging prediction. Simulations are performed to detail the cluster formation technique and to analyze the performance in comparison with centralized and classical interpolation methods. The computational complexity is verified in terms of the number of message exchanges among the sensor nodes. Simulation results demonstrate that clusters are formed by an average of 5 sensor nodes.**

***Index Terms*—Radio environment maps, Distributed channel prediction, Ordinary Kriging, Wireless sensor networks.**

I. Introduction

Radio Environment Map (REM) is an advanced framework that enhances contextual awareness of radio environment in the spatial domain through geo-location aware spectrum measurements [1] [2]. Such measurements are gathered by devices such as Wireless Sensor Nodes (WSNs). A REM indicates the radio signal strength, delay spread and interference levels over a finite geographical region. Exploiting this knowledge of REM is relevant for resource allocation [3], coverage hole prediction and detection [5], interference management [4] and in Anticipatory Networks [6]. In this paper, a REM is formed by the power radiated by the eNodeB. The power received by the WSNs can be probabilistically modeled as multi-scale dynamical system with three major components: path-loss, shadowing and small scale fading [7].

In order to estimate a REM, we rely on mathematical models from geo-statistics. In this paper, we employ the Ordinary Kriging (OK) interpolation method [8]- [13], because of its robustness and its best trade-off between complexity and performance. In addition, OK does not require the knowledge of the mean of the field. It works with the assumption of a constant mean and only requires a variogram function and data values for implementation. The ordinary kriging method presents several key features: (1) It is a local interpolator, which operates within a small area around the estimation location and captures the short-range variations [10]. (2) It employs semivariance as the function to represent spatial dependence instead of covariance. Hence, the mean estimation is not necessary. (3) Along with the estimates, it also quantifies kriging variance. The kriging variance can be used for data screening in order to choose the best set of measurements in the neighborhood [11]. The local nature of the ordinary kriging and its ability to quantify the kriging variance are the key elements that make clustered interpolation feasible in WSNs. The prediction of kriging increases with the number of measurements and its correlation with the unmeasured location. Kriging is a local predictor and only, the closest sensors carry significant weights. Furthermore, the kriging method implies a computational cost that scales as the cube of the number of sensor measurements N, resulting in cubic time complexity $\mathcal{O}(N^3)$.

In this paper, we build on the work of [8] and [9], and modify [8] to reduce the complexity of kriging to $\mathcal{O}(n^3)$, where $n \ll N$. In [8], the Distributed Clustering Algorithm (DCA) for spatial field reconstruction in WSNs was proposed. The algorithm operates in two phases. Firstly, an initial cluster of sensor nodes is built and the kriging operations such as semivariogram estimation and kriging prediction are locally performed in a distributed way. Secondly, the sensor nodes that reduce the kriging variance are added to the initial set of nodes to enhance the estimation quality. However, in [8] the size of initial cluster was not fixed and was estimated from a minimum square error analysis by comparing various initial cluster sizes. In this paper, our main contributions compared to our previous work are as follows: (1) an analytical expression to begin the clustering phase is provided, (2) the DCA performance is analyzed by comparing its performance with classical interpolation methods and (3) The computational complexity in terms of number of messages exchanged is analyzed for both centralized and distributed clustering algorithms.

II. Model and problem statement

A. Network and Channel Model

Consider an heterogeneous networks in a square area $\mathbf{x} \in \mathbb{R}^2$, consisting of a base station t and N wireless sensor nodes. The eNodeB is located at the center of the square at location $\mathbf{x}_t$, whereas the sensor nodes are deployed randomly at locations $\mathbf{x}_i = (x, y)$. Fig. 1 depicts the considered scenario. We will model the WSNs as a connectivity graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ and $\mathcal{E}$ represent the N sensor nodes and links, respectively. We assume that the resource constraints of sensor nodes limits the transmission range of each node to a distance

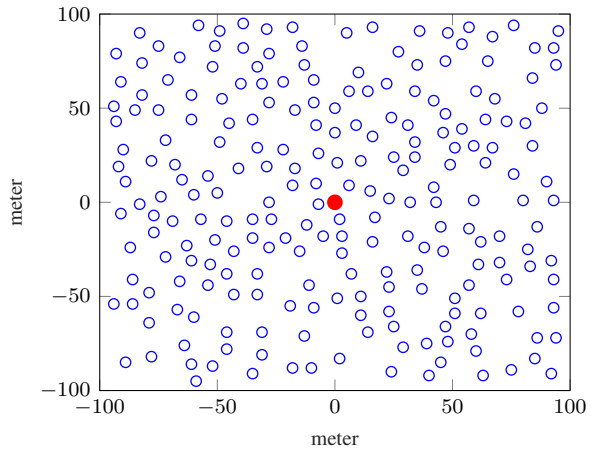


Fig. 1. Heterogeneous network. The red and blue circles indicate eNodeB and sensor nodes, respectively

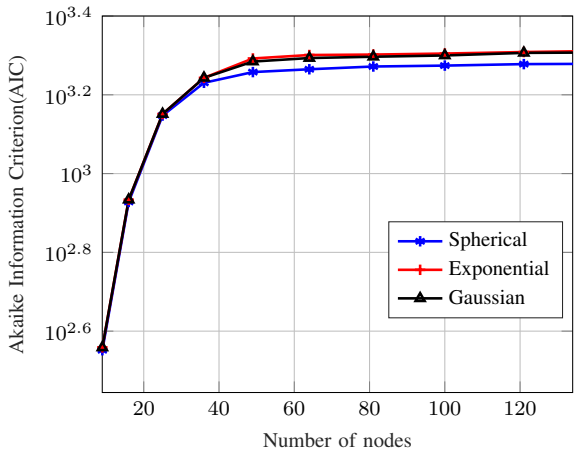


Fig. 2. AIC versus number of sensor nodes for typical semivariogram models

R. As a result, the link is established only if the euclidean distance $||\mathbf{x}_i - \mathbf{x}_j|| \leq R$. N and

Let P_t be the power transmitted through the wireless channel. The power sensed by the sensor nodes can be modeled as:

$$V(\mathbf{x}_i) = P_r(\mathbf{x}_i, \mathbf{x}_t) + n_i, \qquad i = 1, 2, \ldots, N, \quad (1)$$

where $P_r(\mathbf{x}_i, \mathbf{x}_t)$ is the power received at a sensor node located in $\mathbf{x}_i$ from transmitter node located at $\mathbf{x}_t$. The term n_i accounts for a zero mean additive white gaussian noise random variable. The power received at a sensor node, $P_r(\mathbf{x}_i, \mathbf{x}_t)$, averaged over small scale fading in time or frequency, can be expressed in dBm scale as [14]:

$$P_r(\mathbf{x}_i, \mathbf{x}_t) = P_t + P(\mathbf{x}_i, \mathbf{x}_t) + \psi(\mathbf{x}_i, \mathbf{x}_t), \quad (2)$$

where $\psi(\mathbf{x}_i, \mathbf{x}_t)$ is the location dependent shadow fading between transmitter and receiver. The shadow fading component follows a log-normal distribution with zero mean and variance $\sigma^2_{\psi_{dB}}$. $P(\mathbf{x}_i, \mathbf{x}_t)$ depends on path-loss and is given by [14]:

$$P(\mathbf{x}_i, \mathbf{x}_t) = K_{dB} + 10\eta \log_{10} d_0 - 10\eta \log_{10} d(\mathbf{x}_i, \mathbf{x}_t), \quad (3)$$

where K_{dB} is the constant path-loss factor which depends on antenna characteristics and other propagation gains, η is the path-loss exponent, d_0 is a reference distance for antenna far field and $d(\mathbf{x}_i, \mathbf{x}_t) = ||\mathbf{x}_i - \mathbf{x}_t||$ is the distance between sensor node location $\mathbf{x}_i$ and transmitter location $\mathbf{x}_t$.

B. Problem Statement

The goal of this paper is to perform distributed cluster based ordinary kriging interpolation for constructing the REM to significantly reduce the complexity. This is achieved by predicting the received power $\hat{V}(\mathbf{x}_0)$ at all unmeasured locations $\mathbf{x}_0$, using the least number of geo-location aware sensor node measurements $V(\mathbf{x}_i)$.

III. Ordinary Kriging

Ordinary kriging (Ok) is based on the assumption that mean is not required for prediction. The two step process of OK consists of semivariogram estimation and kriging prediction. The semivariogram measures the spatial correlation in the data while kriging employs a semivariogram model to perform prediction. The Empirical Semivariogram (EV) is defined as,

$$\hat{\gamma}(\mathbf{h}) \equiv \frac{1}{2|\mathcal{N}(\mathbf{h})|} \sum_{\mathcal{N}(\mathbf{h})} (V(\mathbf{x}_i) - V(\mathbf{x}_j))^2, \quad (4)$$

where $V(\mathbf{x}_i)$ and $V(\mathbf{x}_i)$ are field values at locations $\mathbf{x}_i$ and $\mathbf{x}_j$, and $\mathcal{N}(\mathbf{h})$ is the number of paired comparisons at lag $\mathbf{h}$.

For kriging, due to computation reasons (refer [10]), The EV has to be replaced by a mathematical model. The most popular choices in geostatistics literature are spherical, gaussian and exponential models. In this paper, we use Akaike Information Criterion (AIC), a quantitative measure for selecting the model that best fits the computed EV values. AIC is estimated by [13]:

$$\widehat{AIC} = N \ln r + 2P, \quad (5)$$

where N is the number of points in the EV, r is the mean of the squared residuals between the experimental values and the fitted model, and P is the number of parameters in the semivariogram model. Fig. 2 shows the AIC performance evaluation for the three typical semivariogram models, which is computed from equation (5) by varying the WSN network size. The semivariogram model with the smallest AIC is the best model for kriging [13]. Hence, from the results, we choose the spherical model for fitting, which is given by the following mathematical function,

$$\overline{\gamma}(\mathbf{h}) = \begin{cases} c_0 + c_1 \left\{ \frac{3}{2}(\frac{\mathbf{h}}{a}) - \frac{1}{2}(\frac{\mathbf{h}}{a})^3 \right\}, & 0 < \mathbf{h} \leq a \ , \\ c_0 + c_1, & \mathbf{h} > a, \\ 0, & \mathbf{h} = 0, \end{cases} \quad (6)$$

where c_0, a and $c_0 + c_1$ specifies nugget, range and sill, respectively. The spherical model is fitted by weighted least squares estimation [12]. The following equations are used to compute the initial values from the EV [13]:

$$c_{0_{ini}} = max\left[0, \hat{\gamma}(\mathbf{h}_1) - \frac{\mathbf{h}_1}{\mathbf{h}_2 - \mathbf{h}_1}\hat{\gamma}(\mathbf{h}_2) - \hat{\gamma}(\mathbf{h}_1)\right],$$
$$a_{ini} = \frac{\mathbf{h}_{\mathrm{N}}}{2}, \quad (7)$$
$$c_{0_{ini}} + c_{1_{ini}} = \frac{\hat{\gamma}(\mathbf{h}_{\text{N-2}}) + \hat{\gamma}(\mathbf{h}_{\text{N-1}}) + \hat{\gamma}(\mathbf{h}_{\mathrm{N}})}{3},$$

where $c_{0_{ini}}$, a_{ini} and $c_{0_{ini}} + c_{1_{ini}}$ are the initial values for nugget, range and sill, respectively. Once the $\overline{\gamma}(\mathbf{h})$ is computed, it is used in the OK system of equations to obtain the weights,

$$\sum_{i=1}^{N} w_{i|N}(\mathbf{x}_0)\overline{\gamma}(\mathbf{x}_i - \mathbf{x}_j) + \mathcal{L}(\mathbf{x}_0) = \sum_{i=1}^{N} \overline{\gamma}(\mathbf{x}_i - \mathbf{x}_0), \quad j = 1, 2, \ldots, \mathrm{N}, \quad (8)$$

where $\overline{\gamma}(\mathbf{x}_i - \mathbf{x}_j)$ is the semivariogram between measurements from sensor node locations $\mathbf{x}_i$ and $\mathbf{x}_j$, $\overline{\gamma}(\mathbf{x}_i - \mathbf{x}_0)$ is the semivariogram between samples from sensor node location $\mathbf{x}_i$ and target location $\mathbf{x}_0$ and $\mathcal{L}(\mathbf{x}_0)$ is the Lagrange multiplier. Note that $\overline{\gamma}(.)$ is obtained from the model (6).

Finally, the OK estimate and variance is computed using,

$$\hat{V}(\mathbf{x}_0)|_N = \sum_{i=1}^{N} w_{i|N}(\mathbf{x}_0)V(\mathbf{x}_i), \quad (9)$$

$$\sigma^2(\mathbf{x}_0)|_N = \sum_{i=1}^{N} w_{i|N}(\mathbf{x}_0)\overline{\gamma}(\mathbf{x}_i - \mathbf{x}_0) + \mathcal{L}(\mathbf{x}_0). \quad (10)$$

IV. Low Complexity Distributed Clustering Algorithm

In this section, the DCA algorithm proposed in [8] is modified to reduce its computational complexity. The DCA consists of initial cluster estimation followed by adaptive cluster estimation, which is illustrated in the flowchart (from Fig. 3).

A. Initial Cluster Estimation

Each node employs a broadcast protocol in order to find its neighbors and acquire information about its multi-hop neighborhood. This information includes knowledge about neighbor measurements and geo-location. During this phase, an initial set of t nodes closest to an unknown location $\mathbf{x}_0$ begins the prediction process and estimates $\hat{\gamma}(\mathbf{h})|_t$ and $\hat{V}(\mathbf{x}_0)|_t$. In this paper, we modify the DCA of [8] in order to begin the estimation with an initial cluster size of $t = 3$. We make use of equation (7) to fix the size of the initial cluster. $\hat{\gamma}(\mathbf{h})|_t$ is computed using a distributed semivariogram algorithm by solving (4) in a distributed manner. And, $\hat{V}(\mathbf{x}_0)|_t$ is obtained using a distributed kriging algorithm, which performs a Gauß-jordan elimination method in an iterative way. We refer the reader to [9] for further details on distributed semivariogram and kriging estimation.

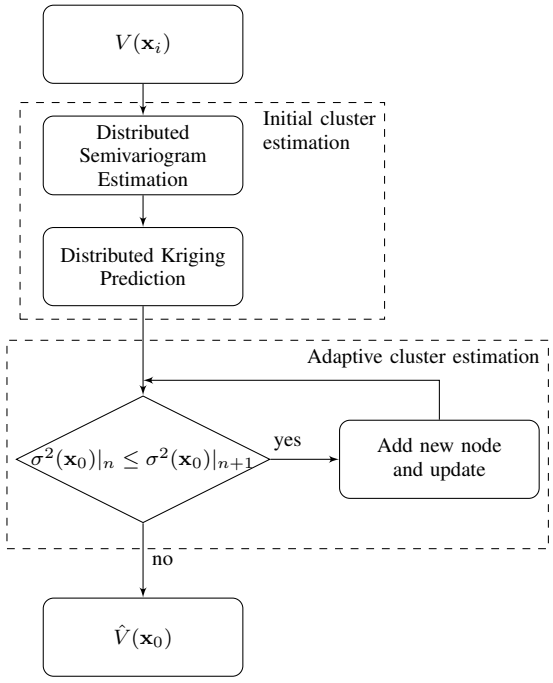


Fig. 3. DCA algorithm

B. Adaptive Cluster Estimation

In the adaptive clustering phase, the initial estimation using $t = 3$ nodes is improved by adding more sensor nodes that are in the communication range. Sensor nodes are successively added to the initial cluster until the incorporation of a new node no longer improves the estimation i.e., the kriging variance is no longer decreased. As a result, a cluster is formed with $n = t + k$ nodes, where n represents the total number of nodes in the cluster and k denotes the number of nodes added during the adaptive clustering phase. Note, the kriging variance can be seen as an approximation of the estimation error. Algorithm 1 in our previous work [8] shows this operation in detail. With DCA, nodes with high correlation to $\mathbf{x}_0$ are given all the weights while the non-cluster nodes are ignored. This ensures that all the information relevant to the spatial prediction is captured effectively.

V. Simulation Results

We consider a LTE-sensor network in a sub-urban environment for simulations. The WSNs consisting of N sensor nodes with 8 m inter-node spacing are randomly distributed in a square area of 190 m $\times$ 190 m, while the LTE picocell eNodeB is placed at the center. The key parameters for the simulation scenario are used from references [15] [16], with path-loss exponent $\eta = 3$, standard deviation of shadow fading $\sigma_{\psi_{dB}} = 6$ dB, shadowing correlation distance $d_c = 10$ m and $P_t = 24$ dBm. Based on propagation environment, we obtained $R = 21$ m. The complete REM is obtained by estimating the received power at 9216 locations, when considering a 2 m resolution grid on the square area under study.

The accuracy of field estimation and complexity are important aspects to gauge the performance of reconstruction algorithms in WSNs. In order to analyze the accuracy, we consider the Mean Squared Error (MSE) between the actual and the estimated field value. The DCA performance (see Fig. 4) is compared with centralized and partitioned cases. In a centralized case, sensor nodes send their measurements to a central node for REM estimation whereas, in partitioned case, sensor nodes send their measurements to a subregion head. Each subregion head estimates the REM by making use of local measurements. However, in DCA, each sensor node acts as a cluster head and functions according to the distributed architecture. The trend of the MSE plot proves that the quality of kriging estimation improves when increasing the number of measurements. This happens due to the decrease in the inter-node distance, which in fact increases the spatial correlation between the samples. It is worth to notice from the Fig. 4, that the prediction quality of DCA outperforms the partitioned kriging, and it is similar to the centralized case for networks with $N > 150$. On the other hand, for $N < 150$, the algorithm suffers from lack of sensor nodes within the communication range to build clusters. In case of partitioned kriging, since the inter-region sensor node sharing is not allowed, the cluster heads have less samples to interpolate at the subregion borders. As a result, the prediction quality degrades at the borders. This will be noticed in Fig. 6(c).

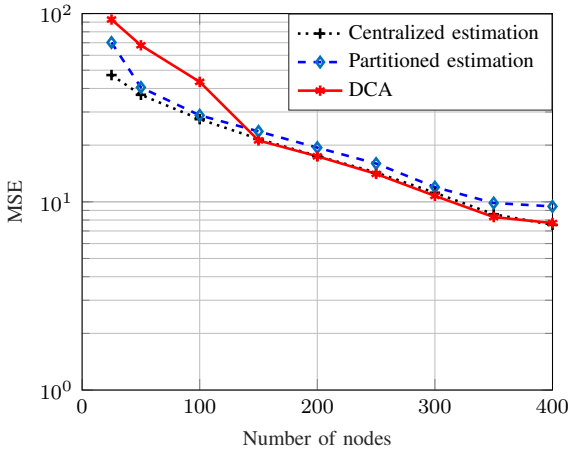


Fig. 4. Comparison of DCA with centralized and partitioned estimation

In order to model the complexity of the algorithms, the number of message exchanges has been evaluated for the centralized and DCA cases. For discovering nodes, a shortest path algorithm is used in the centralized case whereas for DCA, a broadcasting protocol is used. From table I, it can be seen that DCA requires significantly lower message exchanges than centralized kriging. Thus, reducing the complexity and also improving the network lifetime. We also compute the average size of clusters required to build REMs for various WSNs sizes. From Table II, it is clear that DCA requires an average of 5 nodes, which is prominently lower than the centralized case.

TABLE I
COMPLEXITY IN TERMS OF NUMBER OF MESSAGE EXCHANGES

WSNs size	Centralized kriging	DCA
100	3060376	83352
150	5115990	83587
200	7706248	83906
250	8655702	83920
300	10665226	84365
350	12637878	84620
400	13790128	85098

TABLE II
AVERAGE SIZE OF CLUSTER FOR VARIOUS WSNS SIZES

WSNs size	50	100	200	300	400
Average size of the cluster	4.03	4.91	4.95	5.06	5.27

In Fig. 5, the MSE of DCA is compared with two classical interpolation methods such as natural neighbor and spline. The performance evaluation plot clearly illustrates that the DCA outperforms the classical interpolation methods. To illustrate the excellent reconstruction quality of our algorithm, we present interpolated maps in Fig. 6.

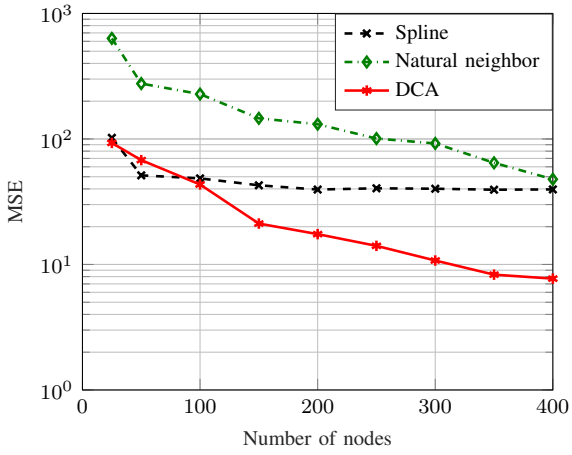
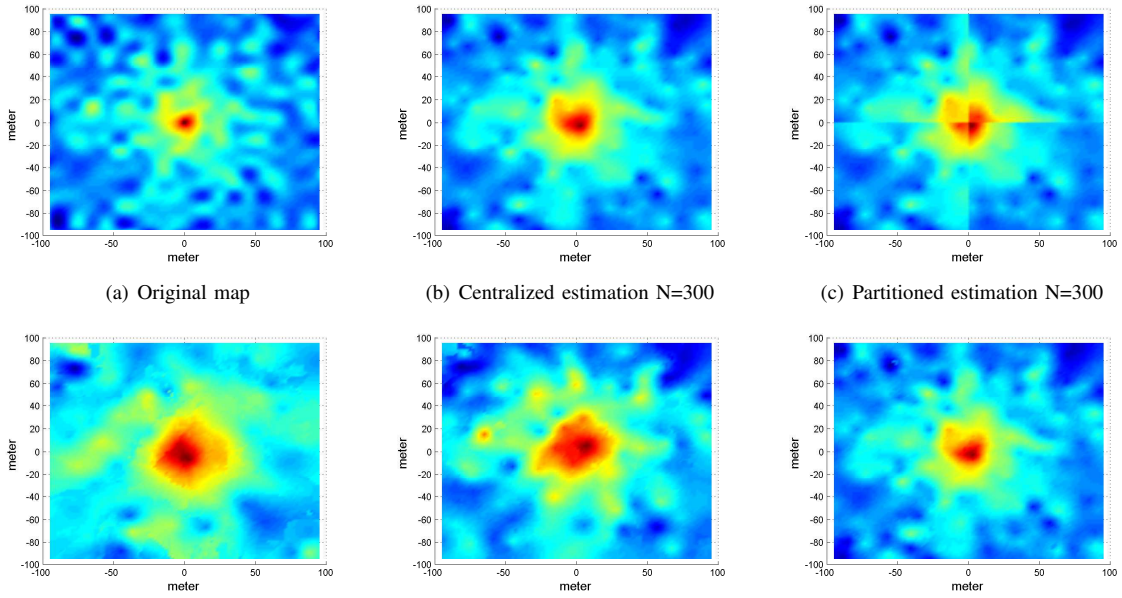


Fig. 5. Comparison of DCA with classical interpolation methods

VI. CONCLUSION

In this paper, a distributed clustering algorithm is modified to reduce the complexity of prediction in REM. The proposed method employs an average of 5 nodes to perform the estimation for various network sizes. The kriging variance has proved to be a good metric for minimizing the computational complexity. Simulation results indicate that the distributed



(a) Original map (b) Centralized estimation N=300 (c) Partitioned estimation N=300

(d) DCA N=100 (e) DCA N=200 (f) DCA N=300

Fig. 6. Interpolated maps for different estimation frameworks

clustering algorithm outperforms the centralized and classical interpolation methods in terms of prediction and complexity, making it more suitable for practical applications. As a part of future work, we aim to investigate the influence of location uncertainty on the prediction quality. Also, we aim to investigate the performance of our algorithm in sparse wireless networks.

ACKNOWLEDGMENT

The authors would like to thank the regional government of Comunidad Valenciana for providing funding through the GVA fellowship S. Grisolía GRISOLIA/2012/028 and the government of Spain for the Spanish MINECO Grant RACHEL TEC2013-47141-C4-4-R. C. Botella belongs to the Polibienestar Unit.

REFERENCES

[1] 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, December 2013.

[2] Mueck M, Jiang W, Sun G, Cao H, Dutkiewicz E and Choi S, New Spectrum Usage Paradigms for 5G [White paper]. *IEEE TCCN SIG CR in 5G* 2014.

[3] 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, March 2011.

[4] J. Perez-Romero et al., 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.

[5] A. Galindo-Serrano, B. Sayrac, S. Ben 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*, Washington, DC, pp. 37-42, 2013.

[6] 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)*, Atlanta, GA, pp. 4877-4882, 2013.

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

[8] 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)*, Glasgow, pp. 1-6, 2015.

[9] G. Hernández-Peñaloza and B. Beferull-Lozano, Field estimation in wireless sensor networks using distributed kriging, in *2012 IEEE International Conference on Communications (ICC)*, Ottawa, ON, pp. 724-729, 2012.

[10] M.A. Oliver, R. Webster, A tutorial guide to geostatistics: Computing and modelling variograms and kriging, *CATENA*, Vol. 113, pp. 56-69, February 2014.

[11] X.Emery, The kriging update equations and their application to the selection of neighboring data, *Computational Geosciences* , vol. 13, pp. 269-280, 2009.

[12] N.Cressie, Fitting Variogram Models by Weighted Least Squares. *Mathematical Geology*, vol. 17, no. 5, 1985.

[13] Jian X, Olea R, Yu Y. Semivariogram modeling by weighted least squares. *Computers and Geosciences*, vol. 22, no. 4, 1996.

[14] Goldsmith A. *Wireless Communications*. Cambridge University press, 2005.

[15] Universal Mobile Telecommunications System (UMTS); Base Station (BS) radio transmission and reception (FDD). *3rd Generation Partnership Project; Technical Specification* [3GPP TS 25.104 version 12.5.0 Release 12], 2015.

[16] Crossbow Technology, *IEEE 802.15.4 TelosB Mote with sensor suite*, TPR2420CA datasheet, 2005.