Citation for the original published paper:
V. Chowdappa, M. Frohöle, H. Wymeersch and C. Botella, "Distributed channel prediction for multi-agent systems," 2017 IEEE International Conference on Communications (ICC), 2017, pp. 1-6, doi: 10.1109/ICC.2017.7997044.

Distributed Channel Prediction for Multi-Agent Systems

Vinay-Prasad Chowdappa*, Markus Fröhle†, Henk Wymeersch†, Carmen Botella*

*Group of Information and Communication Systems, Universidad de Valencia, Valencia, Spain
Email: {vinay.chowdappa, carmen.botella}@uv.es

†Department of Signals and Systems, Chalmers University of Technology, Gothenburg, Sweden
Email: {frohle, henkw}@chalmers.se

***Abstract*—Multi-agent systems (MAS) communicate over a wireless network to coordinate their actions and to report their mission status. Connectivity and system-level performance can be improved by channel gain prediction. We present a distributed Gaussian process regression (GPR) framework for channel prediction in terms of the received power in MAS. The framework combines a Bayesian committee machine with an average consensus scheme, thus distributing not only the memory, but also computational and communication loads. Through Monte Carlo simulations, we demonstrate the performance of the proposed GPR.**

I. Introduction

While on mission, Multi-agent systems (MAS) employ the wireless channel to coordinate among themselves and to report back to the command center (CC). By its nature, MAS have potential to aid humans in situations such as natural and urban disasters, bomb disposal, surveillance, remote surgery, etc., where steady wireless connectivity with the command center is crucial for communication and control. Hence, the ability to predict the channel gain in terms of received power can improve system performance. Typical application scenarios in MAS include formation control [1] and connectivity maintenance [2]. Channel prediction is also applicable to resource allocation in anticipatory networks [3], [4], routing in ad-hoc networks [5], interference management in HetNets [6] and spectrum sensing in cognitive radio [7]. The received power in a wireless channel is mainly affected by deterministic path-loss, random shadowing and random small-scale fading [8]. Since the channel gain is location-dependent (i.e., dependent on the location of the the transmitters (TX), receivers (RX), and environment), standard regression tools can be used for channel prediction, provided a training database of channel measurements is available.

Gaussian process regression (GPR) [9] has been adopted as a tool for channel prediction [10], even in the presence of location uncertainty [11], [12]. GPR can harness both the deterministic components of the channel (path-loss), as well as the spatially correlated random components (shadowing) [13], using well-established correlation models [14]. A drawback of GPR is a computing complexity that is cubic in the number of measurements in the training database. Furthermore, channel prediction is centralized, meaning that full knowledge of the training database is required to perform a channel prediction at an unvisited test location. To remedy the first deficiency, sparse GPR has been proposed to limit computing complexity with increasing size of the training dataset [15]. To address the second deficiency, a different approach was taken in [16], [17], where prediction is distributed among many computation units, all part of a (wired) computation network. With the help of a specialized covariance function, [18] and [19] are able to perform distributed GPR based on a consensus scheme and a sparse GPR approximation, respectively, allowing their application in a MAS setting. In [20], a distributed GPR is proposed for the problem of area coverage, where each robot's predictions are centrally fused to obtain a global prediction.

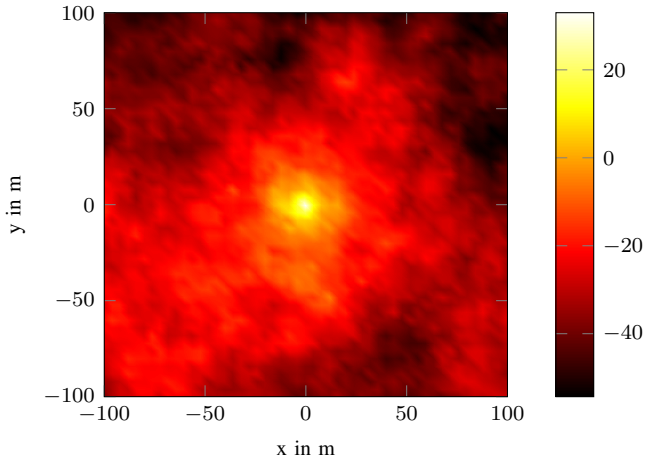

Fig. 1. Example realization of the received power (in dBm) with respect to a transmitter located in the origin.

In this paper, we propose a distributed GPR framework to: (i) perform distributed channel prediction when there is no pre-existing communication infrastructure and (ii) reduce the computational complexity of GPR. We build on the work of [16], [17] to develop a distributed GPR, suitable for wireless communication between training databases, as is the case in MAS. The proposed GPR is capable of handling large training databases, by distributing the computations among the independent mobile agents. In particular, each agent performs

predictions independently on its own local database. The individual predictions are then combined using distributed consensus to obtain a global prediction. In doing so, we ensure that computation complexity at every agent is kept low. Furthermore, the proposed method is neither restricted to any type of covariance function, nor the structure of the MAS network itself. Moreover, it operates on the full dataset in contrast to sparse GPR [16]. As part of our analysis, we assess the impact of how the training database is distributed and how location uncertainty impacts predictions.

II. System Model

A. Network and Channel Model

Consider a geographical area $\mathcal{A} \subset \mathbb{R}^2$ with a single transmitter (TX) and N mobile agents (RXs) located at $\mathbf{x}_t$ and $\mathbf{x}_i$, respectively, where $i = 1, 2, \ldots, N$. The mobile agents are modeled as a connected undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, with vertices $\mathcal{V} = \{1, 2, \ldots, N\}$ representing N agents and edges $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ representing the links among the agents. We assume that the number of agents N in the network is known to every agent. Let A be the adjacency matrix of graph $\mathcal{G}$. Agent i at location $\mathbf{x}_i$ can communicate with agent j at location $\mathbf{x}_j$ only if $[A]_{ij} = 1$.

Let P_{TX} denote the power transmitted through a wireless propagation channel. The channel gain can be decomposed into path-loss, shadowing and small-scale fading [8]. Assuming that the small-scale fading is averaged out by the RX, the received power in dBm can be expressed as

$$P_{\mathrm{RX}}(\mathbf{x}_i) = P_{\mathrm{TX}} + G_0 - 10\gamma \log_{10} \|\mathbf{x}_t - \mathbf{x}_i\| + \psi(\mathbf{x}_i), \quad (1)$$

where G_0 is a constant that captures antenna and other propagation gains, γ is the path-loss exponent and $||\mathbf{x}_t - \mathbf{x}_i||$ is the Euclidean distance between TX and RX positions, and $\psi(\mathbf{x}_i)$ is the spatially correlated shadow fading component (in dB), which is Gaussian distributed with mean zero and variance σ_ψ^2. The channel model (1) has been empirically confirmed by [13], [21]–[23], allowing to model the variations of received signal power in a wireless channel. Due to the imperfect RX characteristics, each agent is assumed to obtain a noisy version of the received power $y_i = P_{\mathrm{RX}}(\mathbf{x}_i) + n_i$, where $n_i \sim \mathcal{N}(0, \sigma_n^2)$. An example of a realization of the received power with respect to a TX located at the origin is shown in Fig. 1.

B. Training Database Model

During the measurement phase, each agent i visits a set of locations, aggregated in a matrix $\mathbf{X}_i$ and collects the corresponding RX powers, aggregated in a vector $\mathbf{y}_i$. The agent thus builds up a training database $\mathcal{D}_i = \{\mathbf{X}_i, \mathbf{y}_i\}$, as visualized in Fig. 2. The number of visited locations will be denoted by $|\mathcal{D}_i|$. We further introduce $\mathbf{y} = [\mathbf{y}_1^{\mathrm{T}}, \mathbf{y}_2^{\mathrm{T}}, \ldots, \mathbf{y}_N^{\mathrm{T}}]^{\mathrm{T}}$, $\mathbf{X} = \{\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_{\mathrm{N}}\}$, and the complete database $\mathcal{D} = \bigcup_{i=1}^{N} \mathcal{D}_i$, which is not available to any agent. We assume no overlap between the local databases, so that $|\mathcal{D}| = \sum_{i=1}^{N} |\mathcal{D}_i|$.

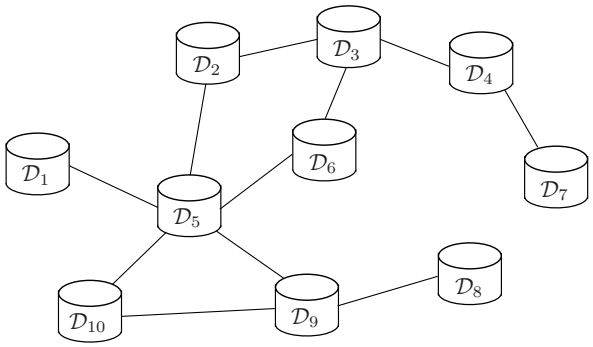

Fig. 2. Each agent i has a local database $\mathcal{D}_i$, containing $|\mathcal{D}_i|$ entries.

III. Problem Statement

Our aim is to perform:

1) *Distributed Learning*: Estimate the channel parameters (say θ, to be defined later) from $\mathcal{D}$, through distributed processing.
2) *Distributed Prediction*: Determine the predictive distribution $p\big(P_{\mathrm{RX}}(\mathbf{x}_*) \mid \mathcal{D}, \theta, \mathbf{x}_*\big)$ of the received power at an unvisited test location $\mathbf{x}_*$ through distributed processing.

Before detailing the distributed approaches to learning and prediction, we first review standard centralized learning and prediction.

IV. Centralized Gaussian Process Regression

In this section, we present GPR for capturing the spatial correlation of the received power and for predicting the channel at a test location. The received power $P_{\mathrm{RX}}(\mathbf{x}_i)$ at location $\mathbf{x}_i$ represents a continuous spatial process and is modeled as a Gaussian process:

$$P_{\mathrm{RX}}(\mathbf{x}_i) \sim \mathcal{GP}\big(\mu(\mathbf{x}_i), C(\mathbf{x}_i, \mathbf{x}_j)\big), \quad (2)$$

where the mean function $\mu(\mathbf{x}_i)$ and the covariance function between location $\mathbf{x}_i$ and location $\mathbf{x}_j$ denoted $C(\mathbf{x}_i, \mathbf{x}_j)$ are defined as

$$\begin{aligned} \mu(\mathbf{x}_i) &= \mathbb{E}[P_{\mathrm{RX}}(\mathbf{x}_i)] \\ &= P_{\mathrm{TX}} + G_0 - 10\gamma \log_{10} \|\mathbf{x}_t - \mathbf{x}_i\|, \end{aligned} \quad (3)$$

$$C(\mathbf{x}_i, \mathbf{x}_j) = \sigma_\psi^2 \exp\bigg(- \frac{\|\mathbf{x}_i - \mathbf{x}_j\|^\alpha}{d_c^\alpha} \bigg), \quad (4)$$

where d_c is the correlation distance of the shadowing. By setting $\alpha = 1$, we will obtain the Gudmundson model [14], which will be used in this paper to model the correlation properties of shadowing.

A. Learning

The objective of learning is to estimate the hyper-parameters of the GPR model from the channel database $\mathcal{D} = \{\mathbf{X}, \mathbf{y}\}$. From (3) and (4), the hyper-parameters of our GPR model are given by

$$\theta = [\gamma, G_0, \sigma_\psi, d_c, \sigma_n]^{\mathrm{T}}. \quad (5)$$

The vector θ can be learned by maximizing the log-likelihood with respect to θ, which is given by

$$\hat{\theta} = \arg\max_{\theta} \log p(\mathbf{y} \mid \mathbf{X}, \theta). \tag{6}$$

B. Prediction

Once the hyper-parameter vector θ is estimated, the predictive probability density function (pdf) of the received power at a test location $\mathbf{x}_*$ is obtained by conditioning $P_{\mathrm{RX}}(\mathbf{x}_*)$ on the dataset $\mathcal{D} = \{\mathbf{X}, \mathbf{y}\}$. It is denoted $p\big(P_{\mathrm{RX}}(\mathbf{x}_*) \mid \mathcal{D}, \theta, \mathbf{x}_*\big)$ and follows a Gaussian distribution with mean $\bar{P}_{\mathrm{RX}}(\mathbf{x}_*)$ and variance $\sigma^2_{\mathrm{RX}}(\mathbf{x}_*)$ given by [9]

$$\bar{P}_{\mathrm{RX}}(\mathbf{x}_*) = \mu(\mathbf{x}_*) + \mathbf{k}_*^{\mathrm{T}} \mathbf{K}^{-1} \big(\mathbf{y} - \mu(\mathbf{X})\big), \tag{7}$$

$$\sigma^2_{\mathrm{RX}}(\mathbf{x}_*) = k_{**} - \mathbf{k}_*^{\mathrm{T}} \mathbf{K}^{-1} \mathbf{k}_*, \tag{8}$$

where $\mathbf{K} = C(\mathbf{x}_i, \mathbf{x}_j) + \sigma_n^2$, so that $\mathbf{K}$ is the $N \times N$ covariance matrix of $\mathbf{y}$ at $\mathbf{X}$; $\mathbf{k}_* = C(\mathbf{x}_i, \mathbf{x}_*)$ for $\mathbf{x}_i \in \mathbf{X}$ is the $N \times 1$ covariance vector between measurements $\mathbf{y}$ and the measurement $\mathbf{y}_*$ at $\mathbf{x}_*$; $k_{**} = C(\mathbf{x}_*, \mathbf{x}_*) + \sigma_n^2$ is the prior variance; $\mu(\mathbf{X})$ is the $N \times 1$ mean vector at training locations $\mathbf{X}$ where the i-th entry is computed with the help of (3), i.e., $\mu(\mathbf{X}) = \mu(\mathbf{x}_i)$ for $\mathbf{x}_i \in \mathbf{X}$, and $\mu(\mathbf{x}_*)$ is the prior mean at test location $\mathbf{x}_*$. Note, the prior mean and prior variance can be obtained in the absence of the database $\mathcal{D}$.

V. Distributed Gaussian Process Regression

In centralized GPR, the CC performs the prediction computations (i.e., mean (7) and variance (8)) by collecting the measurements from all the agents. In this section, we present a distributed GPR where each agent performs the prediction computations independently in a distributed way. For this, we assume independence among the agent databases $\mathcal{D}_i$, similar to [16]. Each agent determines a local pdf conditioned on the local database, and subsequently combines the local predictions using a Bayesian Committe Machine (BCM) from [17] to obtain the overall predictive pdf $p\big(P_{\mathrm{RX}}(\mathbf{x}_*) \mid \mathcal{D}, \theta, \mathbf{x}_*\big)$ for a test location $\mathbf{x}_*$. We use the BCM to develop a distributed GPR in MAS by adapting a distributed average consensus algorithm [24]. Our focus will be on prediction, but for the sake of completeness, we briefly discuss learning as well.

A. Learning

Under the independence assumption among individual databases $\mathcal{D}_i$, the log-likelihood is approximated as

$$\log p(\mathbf{y}|\mathbf{X}, \theta) \approx \sum_{i=1}^{N} \log p(\mathbf{y}_i|\mathbf{X}_i, \theta), \tag{9}$$

so that maximizing with respect to θ can be written in the following form:

$$\underset{\theta_1, \theta_2, \ldots, \theta_N}{\text{maximize}} \quad \sum_{i=1}^{N} \log p(\mathbf{y}_i|\mathbf{X}_i, \theta_i) \tag{10}$$

$$\text{subject to} \quad \theta_i - z = 0, i \in \{1, \ldots, N\}, \tag{11}$$

where z is a common global variable. Problems of this form can be solved using standard distributed approaches, such as ADMM [25].

B. Prediction

Once an estimate of θ is obtained, we aim to determine $p\big(P_{\mathrm{RX}}(\mathbf{x}_*) \mid \mathcal{D}, \theta, \mathbf{x}_*\big)$ at an unvisited test location $\mathbf{x}_*$. For distributed prediction, each mobile agent relies on its own database. Hence, we perform the following GP approximation, similar to [17]

$$p\big(\mathcal{D}_i \mid \mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_{i-1}, \mathcal{D}_{i+1}, \ldots, \mathcal{D}_N, P_{\mathrm{RX}}(\mathbf{x}_*), \theta, \mathbf{x}_*\big) \approx p\big(\mathcal{D}_i \mid P_{\mathrm{RX}}(\mathbf{x}_*), \theta, \mathbf{x}_*\big). \tag{12}$$

According to [17], (12) provides a good approximation when the correlation between databases is small. This condition can be met if measurements from one database are spatially separated from the other. In other words, as pointed out in [17], by partitioning the database $\mathcal{D}$ and assigning the database of each cluster $\mathcal{D}_i$ to a separate agent i. We then have

$$p\big(P_{\mathrm{RX}}(\mathbf{x}_*) \mid \mathcal{D}, \theta, \mathbf{x}_*\big) \propto p\big(P_{\mathrm{RX}}(\mathbf{x}_*) \mid \theta, \mathbf{x}_*\big) \times \prod_{i=1}^{N} p\big(\mathcal{D}_i \mid P_{\mathrm{RX}}(\mathbf{x}_*), \theta, \mathbf{x}_*\big). \tag{13}$$

Applying Bayes' rule, we obtain

$$p\big(P_{\mathrm{RX}}(\mathbf{x}_*) \mid \mathcal{D}, \theta, \mathbf{x}_*\big) \propto \frac{\prod_{i=1}^{N} p\big(P_{\mathrm{RX}}(\mathbf{x}_*) \mid \mathcal{D}_i, \theta, \mathbf{x}_*\big)}{\big(p(P_{\mathrm{RX}}(\mathbf{x}_*)|\theta, \mathbf{x}_*)\big)^{N-1}}, \tag{14}$$

where $p(P_{\mathrm{RX}}(\mathbf{x}_*)|\theta, \mathbf{x}_*)$ is the a priori predictive pdf at test location $\mathbf{x}_*$ with mean $\mu(\mathbf{x}_*)$ and variance k_{**}. Since all distributions involved in (14) are Gaussian, and the product of Gaussians is also Gaussian, we can write

$$p\big(P_{\mathrm{RX}}(\mathbf{x}_*) \mid \mathcal{D}, \theta, \mathbf{x}_*\big) \propto \frac{\mathcal{N}_{P_{\mathrm{RX}}(\mathbf{x}_*)}\left(\mu_{\mathrm{c}}(\mathbf{x}_*), \sigma^2_{\mathrm{c}}(\mathbf{x}_*)\right)}{\mathcal{N}_{P_{\mathrm{RX}}(\mathbf{x}_*)}\left(\mu_{\mathrm{p}}(\mathbf{x}_*), \sigma^2_{\mathrm{p}}\right)}, \tag{15}$$

with

$$\mu_{\mathrm{c}}(\mathbf{x}_*) = \frac{\sum_{i=1}^{N} \sigma^{-2}_{\mathrm{RX},i}(\mathbf{x}_*)\ \bar{P}_{\mathrm{RX},i}(\mathbf{x}_*)}{\sum_{i=1}^{N} \sigma^{-2}_{\mathrm{RX},i}(\mathbf{x}_*)}, \tag{16}$$

$$\sigma^2_{\mathrm{c}}(\mathbf{x}_*) = \left[\sum_{i=1}^{N} \sigma^{-2}_{\mathrm{RX},i}(\mathbf{x}_*)\right]^{-1}, \tag{17}$$

and

$$\mu_{\mathrm{p}}(\mathbf{x}_*) = \mu(\mathbf{x}_*), \tag{18}$$

$$\sigma^2_{\mathrm{p}} = \frac{k_{**}}{N-1}, \tag{19}$$

where $\bar{P}_{\mathrm{RX},i}(\mathbf{x}_*)$ and $\sigma^{-2}_{\mathrm{RX},i}(\mathbf{x}_*)$ are the predictions of agent i, based on $\mathcal{D}_i$, computed similarly to (7)–(8), respectively. Note that the prior mean $\mu(\mathbf{x}_*)$ and variance k_{**} are known to each agent. Using the results for the ratio of two Gaussian distributions, we find that $p\big(P_{\mathrm{RX}}(\mathbf{x}_*) \mid \mathcal{D}, \theta, \mathbf{x}_*\big)$ is proportional to a Gaussian distribution with mean

$$\mu_{\mathrm{post}}(\mathbf{x}_*) = \sigma^2_{\mathrm{post}}(\mathbf{x}_*) \left(\mu_{\mathrm{c}}(\mathbf{x}_*)/\sigma^2_{\mathrm{c}}(\mathbf{x}_*) - \mu_{\mathrm{p}}(\mathbf{x}_*)/\sigma^2_{\mathrm{p}}\right), \tag{20}$$

and variance

$$\sigma^2_{\mathrm{post}}(\mathbf{x}_*) = \frac{1}{1/\sigma^2_{\mathrm{c}}(\mathbf{x}_*) + 1/\sigma^2_{\mathrm{p}}(\mathbf{x}_*)}. \tag{21}$$

From the above derivations, it is clear that the agents should agree on $\mu_{\rm c}(\mathbf{x}_*)$ and $\sigma_{\rm c}^2(\mathbf{x}_*)$ in order to determine $\mu_{\rm post}(\mathbf{x}_*)$ and $\sigma_{\rm post}^2(\mathbf{x}_*)$. Both $\mu_{\rm c}(\mathbf{x}_*)$ and $\sigma_{\rm c}^2(\mathbf{x}_*)$ can be computed in a distributed manner through an average consensus scheme. To this end, we introduce state variables $\xi_i^{(l)}$ and $\lambda_i^{(l)}$ at agent i, where l represents the consensus iteration. We set

$$\xi_i^{(0)} = N\sigma_{{\rm RX},i}^{-2}(\mathbf{x}_*)\ \bar{P}_{{\rm RX},i}(\mathbf{x}_*) \tag{22}$$

and

$$\lambda_i^{(0)} = N\sigma_{{\rm RX},i}^{-2}(\mathbf{x}_*). \tag{23}$$

Letting $\boldsymbol{\zeta}_i^{(l)} = [\xi_i^{(l)}, \lambda_i^{(l)}]^{\rm T}$, agents then apply a consensus update rule [24]. For instance

$$\boldsymbol{\zeta}_i^{(l+1)} = \boldsymbol{\zeta}_i^{(l)} + \eta \sum_{j\in\mathcal{N}_i} \left(\boldsymbol{\zeta}_j^{(l)} - \boldsymbol{\zeta}_i^{(l)}\right), \tag{24}$$

where η is a small constant[1], and $\mathcal{N}_i$ is the neighborhood set of agent i, as determined by the adjacency matrix A. As $l \to +\infty$, it can be verified that

$$\xi_i^{(l)} \to \sum_{i=1}^{N} \sigma_{{\rm RX},i}^{-2}(\mathbf{x}_*)\ \bar{P}_{{\rm RX},i}(\mathbf{x}_*) \tag{25}$$

and

$$\lambda_i^{(l)} \to \sum_{i=1}^{N} \sigma_{{\rm RX},i}^{-2}(\mathbf{x}_*) = 1/\sigma_{\rm c}^2(\mathbf{x}_*), \tag{26}$$

from which $\mu_{\rm c}(\mathbf{x}_*)$ and $\sigma_{\rm c}^2(\mathbf{x}_*)$ are easily retrieved. By substituting $\mu_{\rm c}(\mathbf{x}_*)$ and $\sigma_{\rm c}^2(\mathbf{x}_*)$ in (20) and (21), the global predictive pdf (15) has mean $\mu_{\rm post}(\mathbf{x}_*)$ and variance $\sigma_{\rm post}^2(\mathbf{x}_*)$, which were computed in a distributed way. We note that in case predictions are made at multiple test locations $\mathbf{x}_{*,i}$, the corresponding consensus algorithms can run in parallel.

C. Computation, Storage, and Communication Demands

The main benefit of distributed GPR over centralized GPR lies in the reduction of complexity and storage requirements, at a cost in terms of communication overhead and prediction performance. The storage requirement for distributed GPR relates to: (i) the size of the local database $\mathcal{D}_i$, where generally $|\mathcal{D}_i| \ll |\mathcal{D}|$ and (ii) the data structures for learning and prediction. These latter storage requirements are dominated by storing of the covariance matrices (e.g., $\mathbf{K}$ in (7)) of the training databases, and thus scale as $\mathcal{O}(|\mathcal{D}_i|^2)$ at agent i. The computational requirements are similarly dominated by the covariance matrices, which must be inverted during both the learning and prediction stages. This complexity scales as $\mathcal{O}(|\mathcal{D}_i|^3)$ at agent i. Clearly, the quadratic and cubic scaling of storage and complexity highlight the benefit of distributed GPR over centralized GPR.

In terms of communication overhead, both learning and prediction require iterative methods, whereby agents broadcast and update internal state variables. The number of broadcasts per agent depends on the tolerable disagreement and it is in general hard to quantify, as it depends on the connectivity of the network graph.

[1]A sufficient condition to reach consensus is $\eta < 1/\Delta$, where Δ is the maximum node degree of graph $\mathcal{G}$ [24].

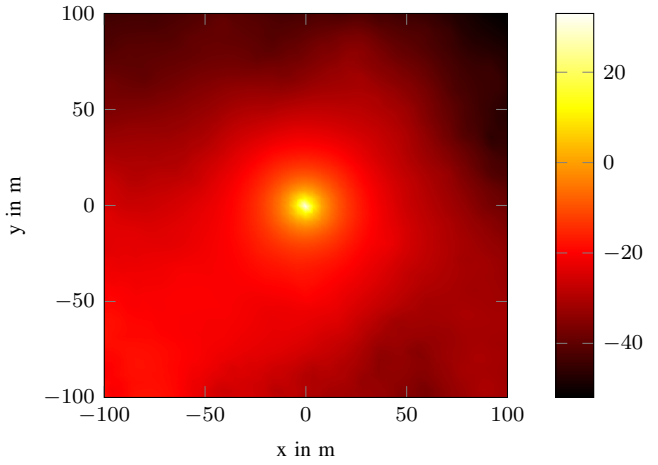

Fig. 3. Mean of GPR prediction (in dBm) for $N = 1000$ (i.e., fully distributed with 1 measurement per agent).

VI. Numerical Results

We present numerical results to illustrate the performance of predicting the received power $P_{\rm RX}$ of centralized GPR and distributed GPR.

A. Simulation Setup

We consider a MAS scenario in a square area $\mathcal{A}$ of $200\,{\rm m} \times 200\,{\rm m}$. The transmitting agent (TX) is placed at location $\mathbf{x}_t = [0, 0]^{\rm T}$ and transmits with a power of 33 dBm. The received power is generated according to a Gaussian field (2), with $\alpha = 1$ in (4), path-loss exponent 3.5, 8 dB of shadowing standard deviation, a shadowing correlation distance of $50\,{\rm m}$, and $\sigma_n = 0.01$. The N mobile agents are randomly and uniformly deployed. Communication links between pairs of agents are established independent and identically distributed (i.i.d.) with probability $p = 0.1$. We perform prediction in a fine grid of $N_p = 2061$ locations (corresponding to a resolution of 4 meters). For training measurements, we will fix the total size of database to $|\mathcal{D}| = 1000$, and let $|\mathcal{D}_i| = |\mathcal{D}|/N$, whereby we vary N. Hence, $N = 1$ corresponds to a centralized GPR, while $N = 1000$ corresponds to one measurement per agent.

We consider two distinct ways of assigning measurements to agents:

- *Random assignment:* measurements are randomly distributed in $\mathcal{A}$ and each measurement is assigned randomly to one of the N agents.
- *Clustered assignment:* measurements are first clustered geographically into N clusters of roughly equal size. Each cluster is then assigned to one of the N agents.

To assess the prediction quality using the centralized GPR and the distributed GPR, we consider the mean squared error (MSE) between the ground truth and the predicted mean. It is defined as

$${\rm MSE} = \frac{1}{N_p} \sum_{i=1}^{N_p} \left(\mu_{\rm post}^{(l)}(\mathbf{x}_{*,i}) - P_{\rm RX}(\mathbf{x}_{*,i})\right)^2, \tag{27}$$

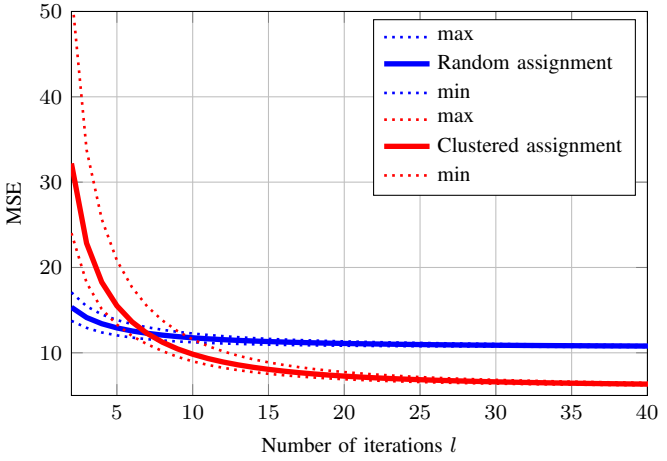

Fig. 4. Mean MSE for networks with different connectivity versus number of consensus iterations l for 20 agents with each having 50 measurements. The dashed lines indicate the minimum and maximum MSE, respectively.

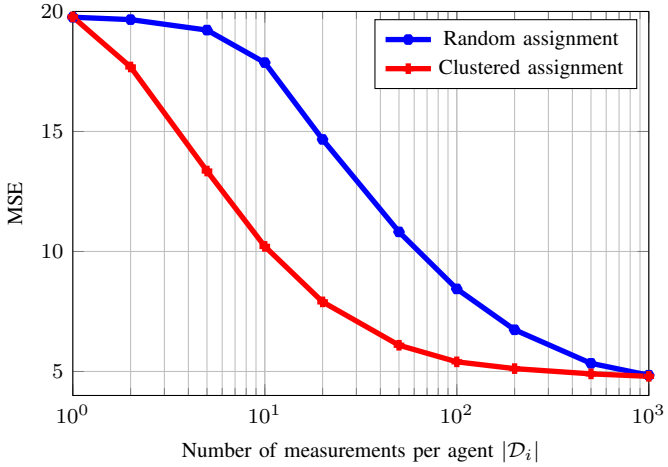

Fig. 5. MSE versus average size of the database per agent, for a total database size of $|\mathcal{D}| = 1000$.

where $\mu_{\text{post}}^{(l)}(\cdot)$ is defined in (20) evaluated at consensus iteration l.

B. Results and Discussion

Qualitative Comparison: To illustrate the reconstruction quality, we present the predicted mean of the received power $\bar{P}_{\text{RX}}$ for $N = 1000$ (i.e., fully distributed) in Fig. 3. We note that the result for $N = 1$ (i.e., fully centralized) matches closely to Fig. 1 and it is therefore omitted. If we compare the mean predicted received power with the true field of Fig. 1, we can notice that the quality of Fig. 3 is comparatively lower than Fig. 1. The main reason for this is that in the distributed GPR case, each agent has just one measurement, and hence (i) agents ($|\mathcal{D}_i| = 1$) cannot exploit the spatial correlation of the large-scale fading; (ii) predictions rely largely on the path-loss only.

Convergence Behavior: We now analyze the impact of the number of consensus iterations l on the MSE for the case of $N = 20$ and $|\mathcal{D}_i| = 50$. In Fig. 4, the mean MSE over the N agents together with the minimum and maximum MSE at each iteration l is plot for random and clustered assignment. We observe that the prediction error decreases with the number of iterations l. In addition, for the random assignment, each agent is able to perform a good prediction of the power at the test location $\mathbf{x}_*$, while for the clustered assignment, only one agent in the vicinity of $\mathbf{x}_*$ will be able to make a good prediction. This leads to a larger initial disagreement for the clustered case, and thus a slower convergence for the clustered assignment.

Quantitative Comparison: We assess the prediction quality with respect to the number of measurements per agent $|\mathcal{D}_i|$. In Fig. 5, the mean MSE is plotted for different number of measurements per agent, averaged over 100 realizations of the channel field. Note that as the number of agents N increases, the cardinality of the respective measurement set $|\mathcal{D}_i|$ decreases, so we transit from the centralized prediction to the completely distributed prediction. First of all, it can be observed that the prediction error in terms of the MSE reduces as the number of measurements at each agent $|\mathcal{D}_i|$ increases. This is because the agents are able to exploit spatial correlation more effectively when there are more entries in the local databases $\mathcal{D}_i$. Secondly, when the databases correspond to geographically clustered locations, the MSE is significantly lower, compared to a completely random assignment of measurements to agents. This is because in the random assignment, there is less opportunity to exploit spatial correlation, i.e., the location of measurements are far apart. Thirdly, reducing the database size from 1000 to 100 leads to a 60 % increase in MSE for the random assignment (resp. 10 % for the clustered assignment), while with fully distributed approach, both assignments lead to a 400 % increase in MSE. This highlight the fact that with distributed approach comes a performance penalty. Finally, we remark that for a simplified path-loss only model, we obtained a mean MSE of 58.34, which is significantly higher than the MSE of distributed GPR with any database size using model (1), highlighting the fact that considering shadowing correlation is important when predicting wireless channels.

Prediction Under Location Uncertainty: In practical MAS applications, the agents' locations may not have knowledge of their true locations. We consider the case where, for the purpose of prediction, measurements of agent i are taken at locations $\mathbf{X}_i$, but the agent assumes they are taken at $\hat{\mathbf{X}}_i = \mathbf{X}_i + \mathbf{W}_i$, where each entry in $\mathbf{W}_i$ is i.i.d. zero-mean Gaussian with a standard deviation drawn from an exponential distribution with mean $1/\lambda$. We consider uncertain training with $\lambda = 6\,\text{m}$ for a fraction of the agents and $\lambda = 0\,\text{m}$ for the remaining agents. Test locations have no location uncertainty. For $N = 5$, i.e., $|\mathcal{D}_i| = 200$ of Fig. 5, the impact of localization errors on prediction quality is presented in Table I. It can be seen how the MSE value increases with the increase of number of agents with the location uncertainty. Moreover,

the random assignment is less sensitive to location errors. In either case, since location errors are generally less than the shadowing correlation distance, the MSE impact is relatively limited.

TABLE I
IMPACT OF LOCATION UNCERTAINTY

Number of agents with uncertainty	0	1	2	3	4	5
MSE for random assignment	6.73	7.40	8.08	8.82	9.55	10.38
MSE for clustered assignment	5.11	6.88	8.39	9.91	11.35	12.84

VII. CONCLUSION AND OUTLOOK

In this paper, we presented a distributed Gaussian process regression framework for channel prediction in MAS. The proposed method reduces computation and memory requirements compared to a centralized GPR, thus allowing to increase the size of the training dataset. For prediction, this is achieved by first performing local prediction on each mobile agent and then combining the local information using a consensus scheme to obtain a global prediction. In terms of prediction quality, the path-loss and shadowing model employed in the distributed GPR provide superior performance over a simple path-loss only model. Numerical results show that the performance of the proposed method in terms of the mean squared error depends on (i) the number of measurements per agent, (ii) the geographic spread of measurements for each agent. When measurements assigned to an agent are clustered geographically, this leads to better performance, though at a cost of slower convergence. Our future work aims to investigate the size of the agent database that provides a good trade-off between the prediction quality and computation complexity versus communication overhead.

ACKNOWLEDGMENT

This work is supported, in part, by the EU FP7 Marie Curie Initial Training Network MULTI-POS (Multi-Technology Positioning Professionals) under Grant No. 316528, by the EU-H2020 project HIGHTS (High Precision Positioning for Cooperative ITS Applications) under grant no. MG-3.5a-2014-636537, The COPPLAR CampusShuttle cooperative perception & planning platform project, funded under Strategic Vehicle Research and Innovation Grant No. 2015-04849 from VINNOVA. V.P. Chowdappa's work has been supported by the Spanish MINECO Grant RACHEL TEC2013-47141-C4-4-R.

REFERENCES

[1] T. Shima and S. J. Rasmussen, *UAV Cooperative Decision and Control: Challenges and Practical Approaches*. SIAM, 2009.

[2] M. M. Zavlanos, M. B. Egerstedt, and G. J. Pappas, "Graph-theoretic Connectivity Control of Mobile Robot Networks," *Proceedings of the IEEE*, vol. 99, no. 9, pp. 1525–1540, Sept 2011.

[3] L. S. Muppirisetty, J. Tadrous, A. Eryilmaz, and H. Wymeersch, "On Proactive Caching with Demand and Channel Uncertainties," in *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Sept 2015, pp. 1174–1181.

[4] 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

[5] 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.

[6] 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.

[7] 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.

[8] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.

[9] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. MIT Press, 2006.

[10] J. Fink, "Communication for Teams of Networked Robots," Ph.D. dissertation, Univ. of Pennsylvania, 2011.

[11] L. S. Muppirisetty, T. Svensson, and H. Wymeersch, "Spatial Wireless Channel Prediction under Location Uncertainty," *IEEE Transactions on Wireless Communications*, vol. 15, no. 2, pp. 1031–1044, Feb 2016.

[12] M. Fröhle, L. S. Muppirisetty, and H. Wymeersch, "Channel Gain Prediction for Multi-Agent Networks in the Presence of Location Uncertainty," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016, pp. 3911–3915.

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

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

[15] J. Quiñonero-Candela and C. E. Rasmussen, "A Unifying View of Sparse Approximate Gaussian Process Regression," *Journal of Machine Learning Research*, vol. 6, pp. 1939–1959, Dec 2005.

[16] M. Deisenroth and J. W. Ng, "Distributed Gaussian Processes," in *Proceedings of the 32nd International Conference on Machine Learning*, 2015, pp. 1481–1490.

[17] V. Tresp, "A Bayesian Committee Machine," *Neural Computation*, vol. 12, no. 11, pp. 2719–2741, 2000.

[18] S. Choi, M. Jadaliha, J. Choi, and S. Oh, "Distributed Gaussian Process Regression under Localization Uncertainty," *Journal of Dynamic Systems, Measurement, and Control*, vol. 137, no. 3, 2015.

[19] D. Gu and H. Hu, "Spatial Gaussian Process Regression With Mobile Sensor Networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 8, pp. 1279–1290, Aug 2012.

[20] K. Tiwari, V. Honor, S. Jeong, N. Y. Chong, and M. P. Deisenroth, "Resource-Constrained Decentralized Active Sensing for Multi-Robot Systems using Distributed Gaussian Processes," in *2016 16th International Conference on Control, Automation and Systems (ICCAS)*, Oct 2016, pp. 13–18.

[21] 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.

[22] 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.

[23] 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, no. 1, 2016.

[24] R. Olfati-Saber, J. A. Fax, and R. M. Murray, "Consensus and Cooperation in Networked Multi-Agent Systems," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 215–233, Jan 2007.

[25] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.