Thesis for the degree of Doctor of Philosophy

Distributed Spatial Prediction for Radio Environment Maps Reconstruction in Heterogeneous Wireless Networks

Vinay Prasad Chowdappa

VNIVERSITAT DE VALÈNCIA

Tecnologías de la Información, Comunicaciones y Matemática Computacional

Valencia, May 2017

Advisor:
Dr. Carmen Botella Mascarell

To my parents Dr. Chowdappa and Sujatha

Resumen

Motivación y objetivos

En los últimos años, el crecimiento en el número de dispositivos inalámbricos y en el número de aplicaciones que se pueden ejecutar sobre ellos ha provocado un aumento rapidísimo del tráfico de datos móviles, creando la necesidad de mejorar la calidad de servicio a los usuarios. El último informe de la empresa Cisco, publicado en febrero de 2017, predice que el tráfico de datos móviles se multiplicará por siete en el periodo entre 2016 y 2021, creciendo con una tasa agregada anual del 47 %.

Para dar servicio a esta demanda, tanto la industria como la academia se están centrando en las redes de quinta generación o 5G. Las redes 5G se espera que constituyan un entorno complejo e interconectado, que además proporcione múltiples servicios y aplicaciones a un número masivo de usuarios y máquinas. En este concepto se incluye la necesidad de dar soporte o de crear servicios para el paradigma conocido como el internet de las cosas o IoT, donde la visión es la de crear un entorno de todo conectado con todo en todo momento. Un factor muy importante será el de la conectividad, que no puede experimentar discontinuidades de ningún tipo. Por este motivo, es necesario que nuevas características y funcionalidades se puedan desarrollar y añadir a los sistemas inalámbricos actuales.

Una de las consecuencias de este cambio en la visión de las comunicaciones es que se tiende hacia un entorno puramente heterogéneo, tanto desde el punto de vista de los dispositivos a los que se va a dar servicio (desde simples sensores a máquinas complejas), como desde el punto de vista del tipo de redes (más específicamente, puntos de transmisión) que van a dar servicio a estos dispositivos. Dentro de las redes heterogéneas y del paradigma de la IoT, es necesario que la conectividad en sus múltiples acepciones no experimente interrupciones, por lo que las técnicas basadas en coordinación y cooperación van a ser fundamentales. En principio, las aplicaciones basadas en la IoT se pueden clasificar en dos tipos: comunicaciones masivas de tipo máquina y aplicaciones de misión crítica. Para dar servicio a aplicaciones específicas en estos dominios es necesario disponer de información de tipo contextual, así como de conocimiento de la ubicación propia del dispositivo. Una de las tecnologías que es una fuente de este tipo de información contextual es la de los mapas de entorno radio. En esta Tesis, un mapa de entorno radio es la herramienta que va a recopilar información limitada de tipo contextual e información de localización mediante el uso de procesado de tipo distribuido, con el fin

último de complementar a un amplio abanico de aplicaciones de 5G.

La 5G se espera que pueda proporcionar información de localización muy precisa, del orden de un metro o incluso menor. Esta precisión es mucho mejor que la que proporcionan hoy en día los sistemas de localización. Por ejemplo, el estándar LTE (*Long Term Evolution*) obtiene una precisión de decenas de metros, los sistemas de navegación por satélite proporcionan precisiones de cinco metros, mientras que aplicaciones sobre las redes de área local pueden llegar a los tres o cuatro metros. Algunos estudios señalan que la información de localización también se podría obtener en redes móviles de tipo *ad hoc*.

Disponer de información de localización abre la posibilidad de diseñar y proporcionar una gran variedad de nuevos servicios sobre las redes 5G. Además, las redes 5G necesitarán disponer de información contextual; de hecho, esta información deberá obtenerse en tiempo real. La posibilidad de que distintos elementos de la red puedan extraer, procesar y almacenar diferentes tipos de información contextual se ha identificado como uno de los mecanismos fundamentales para mejorar de forma global el rendimiento de todo el sistema. Por ejemplo, en redes celulares, la información contextual se consigue mediante los procesos asociados a la minimización de las pruebas de excitación (*minimization of drive tests*). Esta característica se introdujo en la *Release* 10 del 3GPP (*3rd Generation Partnership Project*) para evaluar el comportamiento de la red y reducir el esfuerzo y coste de las pruebas tradicionales. Es una característica que se ha estandarizado tanto para UMTS (*universal mobile telephony system*) como para LTE, y permite a los operadores utilizar equipos de usuario para recoger medidas radio y la información de localización asociada.

Un mapa de entorno radio es una herramienta que recopila información contextual (básicamente, información del enlace o variaciones del canal inalámbrico) y de localización, dando servicio tanto a las tecnologías tradicionales como a las novedosas y disruptivas que se espera que puedan solucionar los retos asociados a la 5G. Los mapas de entorno radio tienen múltiples aplicaciones dentro de los actuales sistemas de comunicaciones inalámbricas, como pueden ser, entre otras:

- Sensado del espectro y enrutado en radios cognitivas: la radio cognitiva es una red inteligente que mejora la eficiencia del uso del espectro radio mediante el conocimiento en el estado del uso y congestión de los distintos canales disponibles. En este caso, los mapas de entorno radio facilitan la operación de la radio cognitiva, puesto que ayuda a los usuarios secundarios a identificar oportunidades de transmisión en redes con gestión dinámica del espectro.

- Análisis de cobertura: los operadores móviles necesitan detectar fallos de cobertura para mejorar la calidad de servicio percibida por los usuarios. Realizar una detección manual del estado de la cobertura es costoso y no es práctico. Los mapas de entorno radio pueden en este caso proporcionar soluciones bastante realistas mediante el uso de herramienta matemáticas apropiadas.

- Gestión de interferencias en redes heterogéneas: las redes heterogéneas van a ser fundamentales en la 5G para hacer frente no sólo al incremento de tráfico generado por los usuarios tradicionales, sino también para dar servicio a nuevos paradigmas

como el internet táctil o la IoT. La arquitectura de una red heterogénea incluye un grupo de celdas pequeñas que utilizan distintas frecuencias de portadora de forma inteligente, incluyendo celdas de tipo macro en bandas licenciadas, como LTE, y celdas pequeñas trabajando en bandas de uso sin licencia o con licencia, como la WiFi. La densificación de pequeñas celdas aumenta el nivel de interferencia cuando la misma frecuencia se reutiliza en distintas celdas. En este escenario, los mapas de entorno radio proporcionan conocimiento del entorno para posibilitar una gestión más eficiente de las interferencias.

- Asignación de recursos proactiva en redes proacticas: los usuarios móviles se desplazan entre localizaciones, y existe una correlación entre la localización geográfica y la calidad del enlace inalámbrico. La localización de un usuario puede predecirse de distintas formas, tales como a partir del histórico de la movilidad, de un modelo de movilidad, de un patrón de movilidad, etc. Si se conoce la localización de un usuario, la calidad del enlace inalámbrico en cualquier otra localización se puede predecir a partir de la base de datos del mapa de entorno radio. La ventaja de este método es que la estación base puede conocer si algún usuario se dirige hacia una área con baja tasa de datos o fallos de cobertura. En ese caso, puede preasignarle los recursos que sean necesarios para que se mantenga su calidad de servicio.

- Comunicaciones dispositivo a dispositivo. Este tipo de comunicaciones es una de la direcciones tecnológicas disruptivas que se ha identificado para el éxito de la 5G. Las comunicaciones dispositivo a dispositivo permitirán crear infraestructuras de bajo coste que manejarán comunicaciones de tipo local sin la necesidad de contar con una estación base. La disponibilidad de un mapa de entorno radio permitiría en este caso optimizar estas comunicaciones locales, especialmente en el proceso de toma de decisiones.

- Sistemas con múltiples agentes o MAS (*multiple agent system*). Este tipo de sistemas utilizan el canal inalámbrico para que los agentes se coordinen entre ellos y se comuniquen con un centro de mando. Por su naturaleza, el potencial de estos sistemas se centraría en situaciones como catástrofes naturales, vigilancia, cirugía remota, etc., donde una conectividad inalámbrica estable con el centro de mando es vital. En este caso, la disponibilidad de un mapa de entorno radio permitiría a los agentes mantenerse en las regiones con un nivel de potencia recibida adecuado, asegurando la conectividad.

Aunque los mapas de entorno radio, como se ha mencionado anteriormente, tienen aplicación en todos los casos anteriores, el trabajo de la Tesis se centra en el caso concreto de las redes heterogéneas. Concretamente, se van a contemplar dos escenarios sobre una misma celda: un primer escenario donde habrá una red de sensores inalámbricos o WSN (*wireless sensor networks*) y un segundo escenario donde lo que se tendrá será un MAS, donde cada agente construirá su base de datos. En la Tesis, la información contextual se refiere a la información de nivel de enlace, concretamente a la evolución del canal físico inalámbrico, es decir, a la calidad del canal y a los parámetros que caracterizan el canal. Las variaciones de un canal inalámbrico se pueden modelar de forma estadística

como un sistema dinámico de escala múltiple, incluyendo los efectos de la propagación a gran escala y los efectos a pequeña escala como el desvanecimiento. Puesto que el canal inalámbrico depende de la localización, es decir, de la ubicación de los transmisores, receptores y el entorno, una de las opciones para predecir el canal en los mapas de entorno radio es el uso de herramientas estándares de regresión, suponiendo que haya medidas de canal e información de localización disponibles. A lo largo de la Tesis, se va a utilizar el término base de datos para referirse al conjunto de medidas de canal e información de localización. El contenido de esta base de datos puede ser global, si incluye todas las medidas disponibles, o bien local, si incluye un subconjunto de éstas.

Dentro de las técnicas de regresión espacial, una técnica proveniente de la geoestadística llamada Kriging y la regresión de procesos Gaussianos o GPR (*Gaussian process regression*), son probablemente las técnicas más conocidas y aplicadas para la construcción de mapas de entorno radio. Tanto Kriging como GPR tienen la capacidad de manejar la componente determinista del canal (las pérdidas por propagación) y las componentes aleatorias que además están correladas espacialmente como las pérdidas por sombreado. Una de las principales desventajas que exhiben tanto Kriging como GPR es su complejidad computacional, que es cúbica con respecto al número de medidas disponibles. Además, tradicionalmente, la predicción de canal es un proceso centralizado, significando en este caso que se necesitan todas las medidas disponibles para predecir el canal en una posición de la que no se tienen medidas disponibles. Para compensar estos problemas, uno de los objetivos de esta Tesis es la reconstrucción de mapas de entorno radio mediante el desarrollo de algoritmos distribuidos basados en Kriging y GPR, que además disminuyan la complejidad computacional asociada y la utilización de los recursos en redes inalámbricas. Concretamente, dentro del primer escenario que contempla el uso de una WSN para obtener una serie de muestras de la potencia recibida desde la estación base a lo largo del área de cobertura, se va a proponer un algoritmo distribuido con formación de agrupaciones de sensores incremental o DICA (*distributed incremental clustering algorithm*) basado en dos versiones de la técnica de Kriging. Dentro del segundo escenario contemplado, donde un MAS va a ser el encargado de recoger las muestras en las bases de datos de los distintos agentes distribuidos por el área de cobertura, se va a proponer una versión distribuida de GPR, incluyendo tanto la fase de aprendizaje como la de predicción.

El desarrollo de esta Tesis y sus resultados forman parte de un proyecto de investigación financiado en la convocatoria del año 2013 de los proyectos I+D+i -programa estatal de investigación, desarrollo e innovación orientada a los retos de la sociedad. Se trata del proyecto 'Radio Access Technologies for Heterogeneous Wireless Networks', RACHEL TEC2013-47141-C4-4-R.

Metodología

En esta Tesis se combinan los métodos de análisis teórico, el diseño de algoritmos considerando todas las restricciones asociadas, y la simulación mediante ordenador. Partiendo del objetivo fundamental, que es la reconstrucción de mapas de entorno radio que puedan proporcionar información contextual dentro del paradigma de las redes he-

terogéneas, se ha realizado un estudio exhaustivo del estado del arte para identificar cuáles son las técnicas más utilizadas con este fin. Tras esta revisión, se han identificado a Kriging y GPR como las dos técnicas más eficaces para conseguir reconstruir el mapa de entorno radio a partir de un conjunto de medidas o base de datos. También se han identificado sus dos principales inconvenientes que constituyen los retos a superar en el desarrollo de esta Tesis y que son, la complejidad computacional de tipo cúbico de la solución y el hecho de que tradicionalmente las soluciones propuestas hayan sido de tipo centralizado, soluciones que no pueden dar respuesta a los escenarios fundamentalmente formados por comunicaciones de tipo local o distribuido en el paradigma de la 5G.

En el primer caso bajo estudio, el escenario donde a lo largo del área de cobertura de la estación base se distribuye una WSN que es la encargada de obtener las muestras de la potencia recibida, se ha desarrollado un algoritmo distribuido basado en la técnica de Kriging, denominado de forma general DICA. Una primera versión del algoritmo utiliza como base la versión del Kriging denominada Kriging ordinario, con lo cual el algoritmo se ha denominado como DICA-OK. El objetivo de esta primera versión del algoritmo era estudiar las capacidades de reconstrucción del mapa de entorno radio y compararlo con una solución centralizada, donde todas las muestras obtenidas por la WSN se envían a un nodo central (la estación base de la celda), que es la que realiza la interpolación del campo y proporciona el mapa de entorno radio final. La evaluación de los resultados se realiza mediante simulación por ordenador mediante el programa MATLAB. Tras este primer estudio, donde la restricción más importante a incorporar era conseguir una versión distribuida que disminuyera la complejidad computacional asociada al Kriging, se propone un segundo algoritmo basado en una versión del Kriging denominada Kriging regresivo, de forma que el algoritmo se denomina como DICA-RK. La parte más importante de esta solución es que se estiman de forma separada las pérdidas de propagación a gran escala deterministas y las aleatorias o asociadas a las pérdidas por sombreado. Para obtener una solución totalmente distribuida, se proponen los distintos algoritmos y las distintas fases que debe seguir el DICA-RK. Las prestaciones de la solución DICA-RK se comparan con la solución previa, el DICA-OK, y con otras soluciones propuestas en la literatura como los *splines* o el *natural neighbor*. Dado que los sensores no siempre consiguen muestras fiables, se incluye también un modelo de incertidumbre sobre la medida obtenida y se evalúa el algoritmo para ver cómo se degrada frente a esta incertidumbre. La evaluación de los resultados se realiza mediante simulación por ordenador mediante el programa MATLAB.

En el segundo caso bajo estudio, es un MAS el que se distribuye a la largo del área de cobertura de la estación base. Los agentes son los encargados de construir las bases de datos. Para el desarrollo de esta segunda parte de la Tesis, se contó con la colaboración de uno de los grupos más punteros en este campo, el del profesor Henk Wymeersch. Durante las dos estancias de investigación realizadas en su grupo de investigación, se desarrolló una versión distribuida de GPR, incluyendo aspectos tales como distintos tamaños de bases de datos o la incertidumbre en las medidas obtenidas por los agentes. En los dos escenarios bajo estudio, se ha evaluado la complejidad computacional de las soluciones propuestas para comprobar que efectivamente disminuye la de la solución tradicional centralizada. Es importante destacar que los escenarios simulados han tenido en cuenta parámetros y modelos de simulación estándar, propuestos por organismos de

estandarización o por la comunidad científica, con el objetivo de facilitar la reproducción de los resultados obtenidos en esta Tesis. Aunque las simulaciones se han realizado con el programa MATLAB, sería posible utilizar otras plataformas de simulación y lenguajes de programación.

Conclusiones y contribuciones

Dentro del paradigma de las redes heterogéneas, se han contemplado dos escenarios de aplicación, un primer escenario donde el área de cobertura de la estación base está cubierta por una WSN y un segundo escenario donde lo que se superpone es un MAS. En el escenario de la WSN, se ha desarrollado un algoritmo basado en la técnica de Kriging, mientras que en el escenario del MAS se ha desarrollado una solución basada en GPR. Los objetivos principales del trabajo han sido:

- Realizar una predicción de canal distribuida en un entorno sin infraestructura de comunicaciones pre-existente.
- Reducir la complejidad computacional de las soluciones basadas en Kriging y GPR.

Para conseguir estos objetivos se ha trabajado sobre dos ideas fundamentales:

- La predicción de canal se debería realizar utilizando el entorno local del punto donde se quiere predecir el canal, puesto que sólo las medidas más cercanas van a influir en el resultado.
- Es importante distribuir los cálculos o computaciones necesarias entre varias entidades (sensores o agentes según el escenario), para posteriormente, volver a combinarlas y obtener una estimación conjunta.

A continuación se resumen las contribuciones de la Tesis en cada uno de los escenarios considerados.

Distributed incremental clustering algorithm

El DICA es un algoritmo desarrollado con el objetivo de reconstruir un mapa de entorno radio a partir de un conjunto de medidas de potencia recibida obtenidas por una WSN. Las contribuciones en este primer escenario han sido las siguientes:

- Se ha formulado el problema de reconstrucción distribuida de mapas de entorno radio y se ha analizado mediante el diseño de algoritmos adecuados y de la evaluación de su complejidad.
- Se ha propuesto una versión del DICA basada en Kriging de tipo ordinario, DICA-OK, donde el paso fundamental es la formación de agrupaciones de sensores de forma adaptativa para estimar el canal en localizaciones en las que no se disponen de medidas. El parámetro que controla la formación de estas agrupaciones es la varianza de Kriging, que está directamente relacionada con el error de estimación

con respecto al valor real del campo a estimar. Se ha demostrado mediante simulación cómo la evolución de este parámetro permite determinar en qué momento dejar de incorporar nuevos sensores a cada agrupación.

- Se ha realizado un estudio de cuál debería ser el tamaño inicial de las agrupaciones de sensores, puesto que se ha demostrado que este parámetro influye en los resultados obtenidos. A partir del desarrollo de las ecuaciones correspondientes, se obtiene que se necesitan tres sensores para poder formar una agrupación inicial que pueda ejecutar el algoritmo DICA y posteriormente actualizarse incluyendo más nodos según la varianza de Kriging.

- El algoritmo DICA-OK se ha comparado con una versión totalmente centralizada y con una versión que divide el área a estimar en secciones independientes. Los resultados concluyen que el algoritmo DICA-OK consigue resultados en función del error cuadrático medio muy cercanos a la solución centralizado con un mucho menor coste computacional.

- La segunda versión del algoritmo DICA propuesta utiliza la versión de Kriging regresivo, de forma que se denomina DICA-RK. Este algoritmo hace uso de varios algoritmos o etapas, como la estimación D-OLS (*distributed ordinary least square*) y la predicción DC-OK (*distributed cluster based ordinary Kriging*). La estimación D-OLS se encarga de estimar las pérdidas por propagación deterministas, mientras que la predicción DC-OK se encarga de estimar las pérdidas por sombreado. En el caso del DC-OK, se construyen agrupaciones iniciales formadas por tres sensores, que realizan la estimación del semivariograma y la predicción de Kriging de forma local y distribuida. Posteriormente, se añadirán nuevos sensores dependiendo del valor de la varianza de Kriging. La estimación final del campo se obtiene sumando los resultados de la estimación D-OLS y la predicción DC-OK.

- Dentro de la predicción DC-OK, se ha detallado cómo conseguir un modelo matemático que actualice en un solo paso los pesos asociados a la predicción de Kriging tras actualizar las agrupaciones de sensores con nuevos nodos.

- Los resultados de simulación obtenidos, especialmente los mapas de interpolación y los resultados en función del error cuadrático medio, demuestran que la capacidad de reconstrucción del DICA-RK es muy alta. El DICA-RK se ha comparado con una solución puramente centralizada y con soluciones clásicas como los *splines* y *natural neighbor*, obteniendo resultados cercanos a la solución centralizada con mucho menor coste computacional.

- La robustez de la solución DICA-RK frente a incertidumbre en las medidas obtenidas por la WSN se ha modelado y evaluado en función del error cuadrático medio.

Distributed gaussian process regression

GPR es un algoritmo desarrollado con el objetivo de reconstruir un mapa de entorno radio a partir de un conjunto de medidas de potencia recibida obtenidas por un MAS. Las contribuciones en este segundo escenario han sido las siguientes:

- Se ha propuesto un GPR distribuido que es capaz de realizar las fases de aprendizaje y predicción de forma distribuida. Este algoritmo puede manejar bases de datos de gran tamaño mediante la distribución de las computaciones necesarias entre agentes móviles independientes. En concreto, cada agente realiza el aprendizaje y la predicción de forma independiente sobre su propia base de datos y posteriormente los parámetros se combinan mediante un algoritmo de consenso distribuido para alcanzar la solución estimada global. De esta forma, la complejidad del cálculo asociado a cada agente disminuye.

- Se ha desarrollado un método de aprendizaje distribuido basado en el método de muestreo por importancia, con el objetivo de aprender los parámetros del canal inalámbrico a partir de las medidas disponibles de forma distribuida. Este método es efectivo para el aprendizaje tanto para el caso de Kriging (Kriging bayesiano) como para GPR.

- Este método no tiene restricciones asociadas con la función de covarianza, o con la topología del MAS. Además, funciona sobre la base de datos completa, en contraste con el caso del GPR disperso.

- Se ha analizado la influencia del tamaño de la base de datos en el aprendizaje, así como la distribución de la base de datos por agente y el impacto de la incertidumbre asociada a las medidas.

El trabajo desarrollado durante la Tesis ha dado lugar a la publicación de tres artículos en congresos internacionales indexados como CORE B, y al envío de dos artículos a revistas indexadas en el primer cuartil del JCR. Se ha publicado un artículo adicional en un congreso internacional que no se incluye en la Tesis, donde el objetivo era modelar el coste de la comunicación entre sensores asociada a la solución DICA-OK. Durante la realización de esta Tesis, el doctorando ha colaborado con dos estudiantes de Trabajo Fin de Máster cuyos objetivos eran caracterizar la solución DICA-OK y compararla con su versión centralizada en cuanto al volumen de mensajes intercambiados entre los sensores y en cuanto al coste de la comunicación.

Las líneas de trabajo futuro comprenden como paso fundamental el incluir las variaciones, no sólo espaciales, sino también temporales de los mapas de entorno radio, para poder utilizar esta información de forma más efectiva y en tiempo real. Aunque en el trabajo realizado ya se ha incluido de forma básica el hecho de que los sensores o agentes dispongan de medidas con incertidumbre, sería conveniente analizar la robustez de la solución frente a modelos más agresivos. De hecho, sería muy importante probar los algoritmos con medidas de campo reales. Por último, todos los algoritmos y esquemas presentados en esta Tesis se han analizado mediante simulaciones. Su implementación en un testbed mediante el uso de dispositivos como los USRPs sería de gran utilidad de cara a su posible aplicación práctica.

Acknowledgements

This Thesis would not have been possible without the guidance and support of my supervisor Assistant Prof. Carmen Botella. I thank her for accepting me as her PhD student in 2014, when my first supervisor Prof. Baltasar Beferull-Lozano left the university. I would also like to thank him for providing me the PhD opportunity. Thanks to Prof. Rafael J Martínez for the help with funding from March 2016 to February 2017.

Many thanks to Prof. Henk Wymeersch, for hosting me in Chalmers and providing me an opportunity to work on GP. Thanks for the weekly meetings and helping me to understand prediction. It was a pleasure working with you. Thanks to Markus for collaboration and discussions.

Finally and most importantly, I would like to convey my deepest gratitude to my parents and brother.

Vinay Prasad Chowdappa
Valencia, May 2017

This work was funded by the regional government of Comunidad Valenciana through the GVA fellowship S. Grisolía GRISOLIA/2012/028 and the government of Spain under Spanish MINECO Grant "Radio Access Technologies for Heterogeneous Wireless Networks" (RACHEL TEC2013-47141-C4-4-R).

List of Publications

Included Papers

This Thesis is based on the following papers.

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

[B] V. P. Chowdappa and C. Botella, "Distributed Radio Map Reconstruction for 5G Automotive," to be submitted to *IEEE Transactions on Intelligent Transportation Systems*, July 2017.

[C] V. P. Chowdappa, M. Fröhle, H. Wymeersch and C. Botella, "Distributed Channel Prediction for Multi-Agent Systems," in *IEEE International Conference on Communications (ICC)*, Paris, May 2017.

[D] V. P. Chowdappa, C. Botella, Sara Santos Sáez, J. Javier Samper and Rafael J. Martínez, "Low Complexity Distributed Cluster based Algorithm for Spatial Prediction", in *International Wireless Communications and Mobile Computing Conference (IWCMC)*, Valencia, June 2017.

[E] V. P. Chowdappa, M. Fröhle, H. Wymeersch and C. Botella, "Distributed Gaussian Processes and Kriging for Channel Learning and Prediction in Multi-Agent Systems," to be submitted to *IEEE Transactions on Wireless Communications*, August 2017.

Other Work

The following paper is an extension of Paper [A] listed above and is not appended in this Thesis.

[a] F. Frantzis, V. P. Chowdappa, C. Botella, J. Javier Samper and Rafael J. Martínez, "Radio Environment Map Estimation based on Communication Cost Modeling for Heterogeneous Networks," in *IEEE 85th Vehicular Technology Conference (VTC Spring)*, Sydney, June 2017.

Abstract

Mobile data traffic is expected to grow sevenfold at a compound annual growth rate of 47 percent from 2016 to 2021. To meet these demands, wireless communication researchers and designers are turning their attention towards fifth generation (5G) networks. 5G will be a key enabler for the Internet of Things (IoT), whose vision is to create an environment of everything connected everywhere and provide a platform to massive machine-type communications and mission-critical applications. Heterogeneous networks will offer ubiquitous connectivity for IoT applications through a variety of coordination and cooperation techniques. Provisioning services and supporting diverse applications requires the network to be context-aware, utilizing contextual information in real-time. 5G will have the ability to extract and process various contextual information coupled with location information to improve the overall system performance.

Radio environment map (REM) is a powerful tool that leverages link contextual information and location information, to support both the traditional and disruptive technologies in addressing the challenges of 5G. Link context refers in this Thesis to the evolution of the wireless propagation channel, which can be probabilistically modeled as a multi-scale dynamical system consisting of path-loss, shadowing and small-scale fading. Since the wireless channel is location-dependent, standard regression tools can be used for channel prediction in REMs. Kriging and Gaussian process regression (GPR) are popular spatial regression tools from Geo-statistics and machine learning, respectively. Drawbacks of Kriging and GPR are a traditional centralized prediction and their computing complexity. To address these limitations, in this Thesis, REMs are developed using a distributed incremental clustering algorithm (DICA) and distributed GPR to minimize the computational complexity of kriging and GPR, respectively. DICA is a kriging based interpolation method that employs the least number of closest measurements to leverage short range variations in the local neighborhood of the unmeasured location. Distributed GPR distributes the overall computations among the independent mobile agents. Learning and prediction phases of GPR are achieved by first performing local prediction and then combining the local information using a consensus scheme to obtain a global estimate. In addition, a novel distributed learning method based on importance sampling suitable for kriging and GPR is presented. The complexity of the proposed methods is analyzed and simulation results are presented to showcase the algorithm efficacy to REM reconstruction.

Keywords: REM; Distributed channel prediction; Kriging; Gaussian process regression.

Contents

List of Figures

List of Tables

Acronyms

3GPP	3rd generation partnership project
5G	fifth generation
ARM	available resource map
BCM	bayesian committee machine
BS	base station
CAGR	compound annual growth rate
CC	command center
D2D	device to device
DAC	distributed adaptive clustering
DICA	distributed incremental clustering algorithm
DC-OK	distributed cluster based ordinary kriging
DICA-RK	DICA based on ordinary kriging
DICA-RK	DICA based on regression kriging
DKPA	distributed kriging prediction algorithm
D-OLS	distributed ordinary least squares
DSA	distributed semivariogram algorithm
EV	experimental semivariogram
GNSSs	global navigation satellite systems
GPR	Gaussian processes regression
GPS	global positioning system
HetNets	heterogeneous networks

IoT	internet of things
IS	importance sampling
LTE	long term evolution
M2M	machine-to-machine
MANETs	mobile adhoc networks
MAS	multi-agent system
MDT	minimization of drive tests
MSE	mean squared error
OK	ordinary kriging
pdf	probability density function
QoS	quality of service
REM	radio environment map
RK	regression kriging
RX	receiver
SV	semivariogram
TX	transmitter
UMTS	universal mobile telephony systems
WiMax	Worldwide Interoperability for Microwave Access
WLAN	wireless local area network
WLSE	weighted least squares estimation
WSN	wireless sensor network

Chapter 1

Introduction

The proliferation of wireless devices and applications on them has led to the rapid growth of mobile data traffic, and the need for a better quality of service (QoS) experience. According to Cisco's forecast [1], as illustrated in Figure 1.1, mobile data traffic is expected to grow sevenfold at a compound annual growth rate (CAGR) of 47 percent from 2016 to 2021. In a similar trend, by 2021, there will be 8.3 billion handheld or personal mobile-ready devices and 3.3 billion machine-to-machine (M2M) connections.

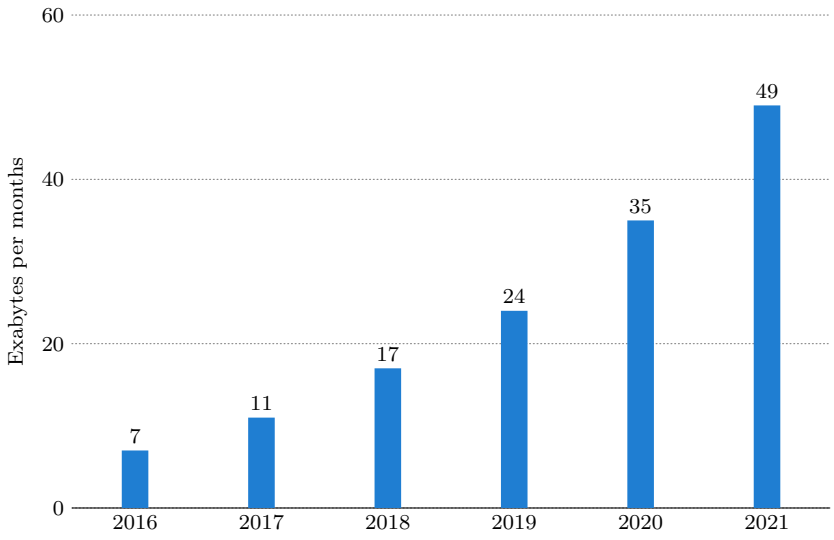


Figure 1.1: Cisco forecasts 49 exabytes per month of mobile data traffic by 2021 [1]

To meet these demands, wireless communication researchers and designers are turning their attention towards fifth generation (5G) networks [2–4]. 5G is expected to evolve

into a complex, interconnected environment with multiple services built on a system that supports a myriad of applications and provides high-speed access to a massive number of subscribers and machines [4]. This includes a desire to support or create services for the internet of things (IoT), whose vision is to create an environment of everything connected everywhere. To enable ubiquitous connectivity for IoT applications, many more features and functionalities will need to be added to the current systems. This inherently leads to a strong heterogeneous network (HetNet) paradigm with a diverse set of devices ranging from simple sensors to complex machines. HetNets will offer the required seamless connectivity for the emerging IoT through a variety of coordination and cooperation techniques [5]. Two classes of IoT-based applications can be identified: massive machine-type communications and mission-critical applications [3]. Provisioning service in those kind of domain specific applications requires the knowledge of context coupled with location awareness [6]. Radio environment map (REM) is one such technology that is used as source of contextual information. In this Thesis, REMs are developed by leveraging limited contextual information and location information through distributed processing to complement a wide range of 5G applications. This Thesis is included in the Spanish MINECO project "Radio Access Technologies for Heterogeneous Wireless Networks" (RACHEL TEC2013-47141-C4-4-R).

1.1 Motivation

5G is expected to provide accurate location information in the order of one meter or below [2, 7–9]. This is significantly better than the accuracy offered by the existing localization systems. For example, a couple of tens of meters in long term evolution (LTE) [10], 5 m in global navigation satellite systems (GNSSs) [11] and 3 m - 4 m in wireless local area network (WLAN) fingerprinting [12]. Studies in [13] show that position information can also be obtained in mobile ad hoc networks (MANETs). Leveraging location information paves the way for wide selection of entirely new features and services in 5G networks [9]. Additionally, 5G systems will need to be context-aware, utilizing contextual information in real-time [14]. The ability to extract, process and store various contextual information by network entities is identified as a vital mechanism to improve the overall system performance [15–17]. For instance, in cellular networks, the contextual information is available through minimization of drive tests (MDT). It is a feature introduced first in 3rd generation partnership project (3GPP) Release 10 [18] to asses network performance and reduce the effort and expense of traditional drive tests [19]. It has been standardized for both universal mobile telephony system (UMTS) and LTE. MDT enables operators to utilize users' equipment to collect radio measurements and associated location information.

A REM is a tool that leverages contextual information and location information, to support both the traditional and disruptive technologies in addressing the challenges of 5G [8]. Contextual information in this Thesis corresponds to the link context [20], which refers to the evolution of the physical wireless channel, i.e., the channel quality and the channel parameters that characterize it. The variation of a wireless channel can be probabilistically modeled as a multi-scale dynamical system consisting of large-

scale and small-scale fading [21]. Since the wireless channel is location-dependent (i.e., dependent on the location of transmitters (TXs), receivers (RXs), and environment), standard regression tools can be used for channel prediction in REMs, provided that channel measurements and location information are available. Thus, the use of the term *database* in this Thesis refers to the collection of channel measurements and location information. The content of the database can be global (all available measurements) or local (subset of global database), depending on the scenario under consideration. In this Thesis, REMs are developed in the framework of HetNets considering two different types of networks: wireless sensor networks (WSNs) and multi-agent systems (MAS). Figure 1.2 shows an example of a REM in terms of power radiated by the TX located in the origin.

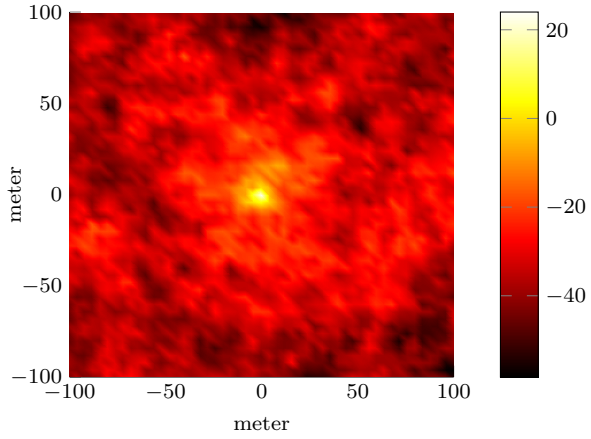


Figure 1.2: Exemplary realization of the received power (in dBm) with respect to a TX located in the origin.

A geo-statistical technique named Kriging [22] and Gaussian process regression (GPR) from machine learning [23] are probably the most well-known spatial regression tools to construct REMs. They can harness both the deterministic components of the channel (path-loss), as well as the spatially correlated random components (shadowing) [24], using well-established correlation models [25]. A drawback of Kriging and GPR is the computing complexity, that is cubic with respect to the number of measurements. Furthermore, traditional channel prediction is centralized, meaning that all available measurements are required to perform a channel prediction at an unmeasured location. To address these deficiencies, this Thesis aims to reconstruct REMs by developing distributed algorithms based on Kriging and GPR to minimize the related computational complexity and resource utilization in wireless networks. More precisely, a distributed incremental clustering algorithm (DICA) based on Kriging and a distributed GPR are presented. Distributed implementation and complexity reduction are the main objectives of this Thesis.

1.2 Contributions

The main publications derived from this Thesis are [26–30]. In [26–30], two distributed methods based on Kriging and GPR for REM reconstruction are proposed, with the objective of: (i) Performing distributed channel prediction when there is no pre-existing communication infrastructure and (ii) Reducing the computational complexity of Kriging and GPR. These objectives are achieved through two key ideas:

- By performing prediction using the local neighborhood where only the closest observations are used for each prediction.
- By distributing computations to independent computational units and, subsequently, recombine them to form an overall estimate.

Based on these ideas, the DICA is introduced in chapter 5, while the distributed GPR is introduced in chapter 6. In addition, the performance comparison of Kriging and GPR methods in a centralized setup is also provided in chapter 4. In the following, the contributions of this Thesis are summarized.

1.2.1 Distributed incremental clustering algorithm

The DICA is a REM reconstruction algorithm developed by adapting the Kriging interpolation method [22], in the framework of WSNs consisiting of a TX and sensor nodes. The TX gives coverage to a single cell and the role of the WSN is to reconstruct the REM of the received power in the coverage area by using the available measurements. The main contributions are the following:

- The problem of distributed spatial field reconstruction of REMs in WSNs is formulated and analyzed in terms of algorithm design and complexity analysis.
- A novel DICA based on regression Kriging (DICA-RK) is presented, which consists of distributed ordinary least square estimation (D-OLS) and distributed cluster based ordinary Kriging 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.
- The cluster formation technique in the DC-OK for shadowing estimation is detailed and an update model to instantly calculate the Kriging weights and Kriging estimates is presented. The proposed 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. The impact of location uncertainty on the performance of the proposed algorithm is also investigated.

1.2.2 Distributed Gaussian process regression

Distributed GPR is built on the work of [31, 32] to develop a distributed method for REM reconstruction, suitable for wireless communication between training databases, as is the case in MAS. Contributions are summarized as follows:

- A novel distributed GPR method capable of performing learning and prediction in a distributed way is proposed. This method can handle large training databases, by distributing the computations among the independent mobile agents. In particular, each agent performs learning/ predictions independently on its own local database. The individual learned parameters/ predictions are then combined using distributed consensus to obtain a global parameter estimate/ prediction. In doing so, the computation complexity at every agent is kept low.
- A novel distributed learning method is presented by adapting importance sampling (IS) method [33], to learn wireless channel parameters from the measurements in a distributed way. The proposed learning method is suitable for learning parameters in both Kriging (bayesian Kriging) and GPR.
- 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 [31].
- As part of the analysis, the influence of database size on learning, the impact of how the training database is distributed and how location uncertainty impacts predictions are investigated.
- The proposed prediction method is also extended to perform Kriging prediction, so that a fair comparison can be made.

1.3 Thesis organization

This Thesis is organized as follows. Chapter 2 introduces the background of REMs, summarizes the related works in Kriging and GPR, and presents the challenges faced in the Thesis. Chapter 3 describes the basics of the wireless channel followed by the channel model, and the RX models for both Kriging and GPR, and how the databases are built for each case. Chapter 4 provides detailed information about parameter learning and prediction in Kriging and GPR, respectively, in a centralized way. Chapter 5 presents the DICA algorithm in the framework of WSNs as one of the solutions to address the challenges of this Thesis, whereas chapter 6 presents the second solution under a MAS setup. Chapter 7 concludes the study in this Thesis and points out several future directions in the research on REMs.

Chapter 2

State of the Art

In this chapter, concepts of REMs, related work and the challenges faced in this Thesis are presented. In Section 2.1, the background of REMs including the definition and architecture, details on spatial interpolation and REMs applications are given. The related work in Section 2.2 is organized around the two main spatial interpolation methods considered in this Thesis: (i) Kriging; (ii) GPR. In Section 2.3, challenges of Kriging and GPR are mentioned.

2.1 Radio environment map

2.1.1 Definition and architecture

A REM is an advanced framework that enhances the contextual awareness of the radio environment in the spectral, temporal and spatial domains, by processing the geo-location aware measurements. Thus, REM creation is a multidimensional problem. REMs gained enormous attention in the research community, due to spectrum sensing in cognitive radio [34], where REMs were proposed as a key tool for the evaluation and optimal reusage of vacant spectrum bands. In cognitive radio paradigm, REMs were thought as a dynamic database containing information about radio frequency signal environment, location of TXs and RXs, radio regulations and relevant historical experiences [35–37]. REMs were studied in FP7 FARAMIR project ("Flexible and spectrum-aware radio access through measurements and modeling in cognitive radio systems") and a functional REM architecture was developed and implemented as one of the core results [38, 39]. Architecture of FARAMIR's REM is illustrated in Figure 2.1, which consists of four main functional blocks: (1) *Measurement capable devices* are responsible for gathering measurement information from the environment. REM architecture should be able to integrate different types of devices such as sensor nodes, unmanned aerial vehicles, etc.; (2) *REM data storage and acquisition* has the dual function of communicating with measurement capable devices and storing raw data or processed data in the form of maps; (3) *REM manager* is the brain of the architecture and it is responsible for evaluating the measurements and generating the REM; and (4) *REM user* sends requests for REM

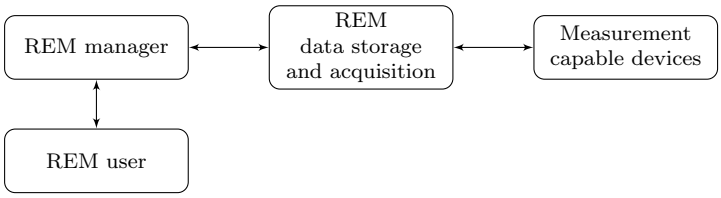


Figure 2.1: Architecture of a REM

information. This information can be used by variety of users such as spectrum regulators, networks operators, etc. On the contrary, a REM based architecture and protocols for MANETs are proposed in [40].

The basic notion of REM reconstruction is to collect geo-located measurements from various network entities and combine them using adequate signal processing techniques [36]. By definition, REMs 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, mathematical models from geo-statistics and machine learning can be used for spatial modeling. In other words, spatial interpolation tools can be used to interpolate a REM, by predicting the field at unmeasured spatial locations based on a set of available measurements.

2.1.2 Spatial interpolation

Spatial interpolation is a statistical method that combines the available measurements with geo-location information to construct a complete map. The principle underlying spatial interpolation is the first law of geography established by Waldo Tobler, which says "everything is related to everything else, but near things are more related than distant things" [41]. Spatial interpolation methods employ statistical tools such as semi-variance and covariance, to capture the spatial correlation and make inference about it. Note that these concepts will be in introduced in chapter 4. One of the key challenging tasks in REM reconstruction is choosing an appropriate interpolation method offering a good quality and complexity trade-off. In [42], about 72 methods/ sub-methods of spatial interpolation in environmental sciences are reviewed. However, when the methods applied to REM reconstruction are considered, Kriging and GPR are the most frequently used. The choice of method depends on the data, the level of accuracy required, complexity and the availability of computation resources [35].

2.1.3 Application of radio environment maps

Although REMs started with cognitive radio, they can be found in several other areas of current wireless systems. Some of them are listed below.

- Spectrum sensing and routing in cognitive radios: Cognitive radio is an intelligent network to enhance the utilization efficiency of radio spectrum resources. REMs

are a tool to support the operation of cognitive radio networks, where their availability could benefit secondary users to identify spectrum opportunities in dynamic spectrum access network [35, 43, 44]. In [17], a routing algorithm was designed for cognitive radio ad-hoc networks by leveraging REMs.

- Coverage analysis: Mobile operators must detect the coverage holes in order to offer satisfactory QoS to the end user. Manual coverage detection and prediction is expensive and practically inefficient. REMs provide a realistic solution by making use of mathematical tools [45–47].

- Interference management in HetNets: HetNets are expected to be fundamental in 5G [48] to deal with the enormous traffic generated not only by traditional mobile users, but also by new paradigms such as the tactile internet [49] and the IoT. HetNets architecture comprises a group of small cells using different carrier frequencies in a smart way, including macro-cells in licensed band (e.g., LTE) and small cells in licensed or unlicensed bands (e.g., Wi-Fi). Densification of small cells in HetNets generates interference when the same frequency is shared by different cells. In this situation, REMs can provide knowledge about the environment to support interference management in HetNets [50].

- Proactive resource allocation in anticipatory networks: Mobile users move from one location to another and there is a correlation between the geographical location and wireless link quality. User location can be predicted from various ways such as user mobility history, mobility model, mobility pattern etc. With the user location, the wireless link quality at any location can be predicted from a REM database. The advantage of this method is that the BS can know if users are heading to low data rate areas or coverage holes. Therefore, it can pre-allocate the required resources to maintain the QoS [20, 51, 52].

- Device to device communication (D2D): In [14, 53], D2D has been identified as one of the key disruptive technology directions for 5G. D2D communication will form a low cost infrastructure that handles local communication without the use of BSs. For instance, to share content or interact (e.g., social networking or peer-to-peer content sharing) through a wireless channel among several co-located devices. In this case, D2D devices can employ a REM to optimize local communications [54].

- MAS: 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 CC is crucial for communication and control. Hence, the availability of REMs can improve the system performance. Typical application scenarios in MAS include formation control [55] and connectivity maintenance [56].

As mentioned before, the REMs reconstructed in this Thesis could be applied in any of the above areas. However, this Thesis focus on a HetNet framework consisting of WSNs and MAS. More precisely, a DICA based on Kriging is developed for a WSN

framework, while a distributed GPR is developed for the MAS scenario. In the following, related work for both Kriging and GPR is presented.

2.2 Related work

In REM literature, there are several papers that have attempted to develop spatial interpolation methods suitable for wireless networks. The first paper to introduce the concept of knowledge database combining environment with location was the available resource maps (ARM) by [57], specifically for cognitive radio applications. Later, it was extended to REMs by [58]. Thereafter, many authors studied the problem with different methods from geo-statistics, machine learning, etc. Some of the notable methods that have been specifically proposed to develop REMs are: nearest neighbor [59], thin plate splines [37, 60], natural neighbor [37, 61], inverse distance weighting [62–65], Kriging [36] and GPR [66]. Kriging is the most frequently cited spatial interpolation method in the REM literature [26, 35–37, 43, 44, 46, 47, 59, 62, 63, 67–73]. Kriging and GPR are stochastic methods capable of modeling deterministic variations (large-scale), spatially autocorrelated variations (small-scale) and uncorrelated noise. However, the other methods are deterministic. The advantage of Kriging and GPR is that they produce estimation error maps i.e., uncertainties associated with predictions. Hence, this Thesis considers Kriging and GPR for building REMs. Kriging and GPR are two staged processes consisting both of parameter learning and prediction. In the case of Kriging, several variants can be found, among which the most commonly used are simple Kriging, ordinary Kriging (OK) and universal Kriging. More details on Kriging and GPR are provided in chapter 4.

2.2.1 Kriging

In [36], the interference cartography concept in secondary spectrum usage is presented and OK is applied to the interference data obtained from a radio network simulator. In [37], Kriging was compared with thin-plate spline and natural neighbor with the aim of evaluating the performance, in the context of cognitive radio. There are several publications by Riijijärvi *at al.* that debate spatial statistics from geo-statistics literature to model the spatial spectrum [68–71, 74–76]. In [43], the first real world application of geo-statistical modeling and interpolation to the problem of REMs is provided. Here, the OK is used to map the coverage of worldwide interoperability for microwave access (WiMax) operating at 2.5 GHz. Further, in [60], the authors experimentally study the problem of spatial interpolation with crowd-sourced measurements, where the MDT measurements in an urban environment are used to evaluate the accuracy of different interpolation techniques, including OK. In [47], fixed rank Kriging is applied to cellular coverage analysis with the aim of lowering the complexity of the spatial interpolation. In [46, 77], bayesian Kriging is implemented to construct the REM for the purpose of coverage hole detection in cellular networks. In [63], universal Kriging is utilized for estimating the REM and the proposed method is compared with OK and inverse distance weighting. In [59], a simple Kriging approach is considered for modeling of interference maps, where a performance

comparison between inverse distance weighting, nearest neighbor and Kriging by means of detection theory is presented. Another comparison study on interpolation method for generating REMs in the context of cognitive radio is conducted in [67]. This study involves the comparison of three interpolation methods, i.e., Kriging, inverse distance weighting and gradient plus inverse distance squared. In [78], the sensor selection problem for an OK in WSNs via alternating direction of multipliers was studied. In [79], a distributed Kriging algorithm based on OK is proposed in 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. In [80], a kriged kalman filtering is adopted and implemented in a distributed fashion to track the time varying shadowing field. In [81], a distributed method is developed by adopting consensus-based Kalman filter to estimate and track the temporal dynamic REM variation. Note, this Thesis deals with purely spatial models and temporal variations are not considered.

REMs reconstruction is a well studied topic in the field of cognitive radio. However, recent studies show its applications in 5G networks [8, 20]. From the above literature review, it can be noticed that the number of distributed methods proposed for REM reconstruction using Kriging are limited. Hence, this Thesis aims at presenting a distributed method for Kriging interpolation, with the objective of employing the least number of measurements in order to reduce the computational complexity.

2.2.2 Gaussian process regression

In [66], a GPR framework is proposed to learn and predict the wireless channel in a centralized setup. In addition, the authors demonstrate that not considering location uncertainty in GPR leads to poor learning and prediction. GPR has two limitations: (1) it is computationally complex and (2) the prediction is centralized. To address the first drawback, in [82], sparse GPR has been proposed to limit the computing complexity with an increasing size of the training dataset. To address the second issue, a different approach was taken in [31, 32], where prediction calculation is distributed among many computation units, all part of a (wired) computation network. With the help of a specialized covariance function, [83] and [84] 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 [85], a distributed GPR is proposed for the problem of area coverage, where each robot's predictions are centrally fused to obtain a global prediction.

In this Thesis a distributed GPR method for REMs reconstruction is proposed to reduce the computational complexity. In contrast to the usual parameter learning by maximum-likelihood (ML) estimation method [31, 66, 84] in GPR, a distributed learning method based on IS is also presented in this Thesis.

2.3 Challenges

In practice, REM reconstruction using Kriging and GPR methods are faced with two main challenges:

1. Centralized reconstruction: To perform prediction at an unmeasured location, Kriging and GPR require all the available measurements. However, the influence of distant measurements on the estimate is minimal. Hence, it is practically inefficient to consider all measurements for reconstruction.

2. Complexity: The estimation quality and accuracy of both Kriging and GPR methods increase with the number of measurements and their correlation with the test location. Due to the inversion of a $N \times N$ semivariance matrix $\mathbf{A}$ in Kriging and a $N \times N$ covariance matrix $\mathbf{K}$ in GPR, 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 Kriging and GPR. Note that matrices $\mathbf{A}$ and $\mathbf{K}$ will be introduced in chapter 4.

In order to tackle these challenges, this Thesis focuses on two main objectives: (1) to propose distributed algorithms for REMs reconstruction; (2) to reduce the computational complexity of Kriging and GPR based solutions.

Chapter 3

System Model

The wireless propagation channel can be probabilistically modeled as a multi-scale dynamical system consisting of path-loss, shadowing and small-scale fading [21]. Small-scale fading characterize the variation of the received signal strength over very short travel distances (which will be considered averaged out in this Thesis), whereas the variation due to path-loss and shadowing occur over relatively large distances. Shadowing de-correlates over 50-100 m [21] for outdoors and 1-5 m for indoors [86], and can be statistically modeled using well established correlation models [25]. In this chapter, details about the channel model, the shadowing model and network models for WSN and MAS are given.

3.1 Channel model

Consider a geographical area $\mathcal{A} \subset \mathbb{R}^2$ with a single transmitter TX and N mobile RXs located at $\mathbf{x}_t$ and $\mathbf{x}_i$, respectively, where $i = 1, 2, \ldots, N$. Note that the RXs will be the sensors in the case of WSN application scenario, and the agents in the case of the MAS. Let P_{TX} denote the power transmitted through a wireless propagation channel. Due to the imperfect RX characteristics, each RX is assumed to obtain a noisy version of the received power,

$$y(\mathbf{x}_i) = P_{\mathrm{RX}}(\mathbf{x}_i) + n_i, \qquad i = 1, 2, \ldots, N, \tag{3.1}$$

where n_i accounts for zero mean additive white Gaussian noise with variance σ_n^2 .

The received signal power in a wireless propagation channel is mainly affected by path-loss, shadowing and small-scale fading [21]. Assuming that the small-scale fading is averaged out by the RX, the received power in dB scale can be expressed as the sum of path-loss and shadowing components:

$$P_{\mathrm{RX}}(\mathbf{x}_i) = G_0 - 10\eta \log_{10} ||\mathbf{x}_t - \mathbf{x}_i|| + s(\mathbf{x}_i). \tag{3.2}$$

Constant $G_0 = P_{\mathrm{TX}} + K_{dB} + 10\eta \log_{10} d_0$, where K_{dB} is the constant path-loss factor, η is the path-loss exponent and d_0 is the reference distance. The term $||\mathbf{x}_t - \mathbf{x}_i||$ is the distance between RX at location $\mathbf{x}_i$ and TX at location $\mathbf{x}_t$ and $s(\mathbf{x}_i)$ is the location

dependent shadowing. We assume that the shadowing follows a log-normal distribution i.e., $s(\mathbf{x}_i) \sim \mathcal{N}(0, \sigma_\Psi^2)$, where σ_Ψ^2 is the shadowing variance. The channel model (3.2) has been empirically confirmed by [24, 87–89], allowing to model the variations of the received signal power in a wireless channel. The path-loss component is clear from equation (3.2) whereas for modeling the spatial correlation of shadowing, we employ the Gudmundson model [25]:

$$C(\mathbf{x}_i, \mathbf{x}_j) = \mathbb{E}[\Psi(\mathbf{x}_t, \mathbf{x}_i)\Psi(\mathbf{x}_t, \mathbf{x}_j)], \tag{3.3}$$

$$C(\mathbf{x}_i, \mathbf{x}_j) = \begin{cases} \sigma_\Psi^2 \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|}{d_c}\right), & \|\mathbf{x}_i - \mathbf{x}_j\| > 0, \\ \sigma_0^2 + \sigma_\Psi^2, & \mathbf{x}_i = \mathbf{x}_j, \end{cases} \tag{3.4}$$

where d_c is the correlation distance of the shadowing. An example of a realization of the received power with respect to a TX located at the origin has been shown in Fig. 1.2.

3.2 Network model

This section introduces the network model for the HetNets application scenarios under consideration, that is, WSN and MAS. In the case of WSNs, sensors are the RXs and measure the spatial field, that is, the received power from the TX. In the MAS scenario, a system of multiple agents with local databases is envisioned.

3.2.1 WSN model

The WSN is modeled as a connectivity graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$, composed by a set of N sensor nodes $\mathcal{V} = \{1, 2, ..., N\}$ and a set of links $\mathcal{E}$. The sensor nodes are deployed over a square area $\mathbf{x} \in \mathbb{R}^2$ to monitor a 2-D spatial field denoted by $y(\mathbf{x}_i)$, where $i = 1, 2, ..., N$. In matrix notation, $\mathbf{y} = [y(\mathbf{x}_1), y(\mathbf{x}_2), ..., y(\mathbf{x}_N)]$ and $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N]$. The spatial location where the field needs to be estimated is denoted as $\mathbf{x}_*$. Due to WSNs 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 $d(\mathbf{x}_i, \mathbf{x}_j) = ||\mathbf{x}_i - \mathbf{x}_j||$ is less than R. We assume that sensor nodes are equipped with global positioning system (GPS) capability, allowing them to estimate the inter-node distance $d(\mathbf{x}_i, \mathbf{x}_j)$ with one hop neighbors and also, calculate the distance $d(\mathbf{x}_i, \mathbf{x}_*)$ with spatial location $\mathbf{x}_*$. Otherwise, it is assumed that sensor nodes can perform estimation of inter-node distances. One of the features of the proposed DICA is that clusters of sensor nodes are formed. 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 Thesis is a random distribution of sensor nodes as shown in Figure 3.1.

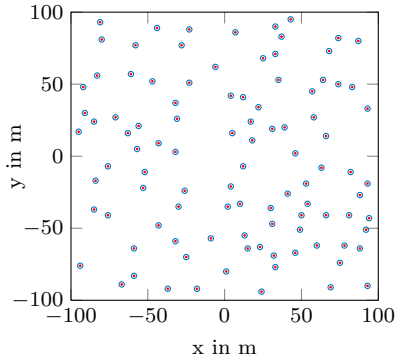


Figure 3.1: Random deployment of sensor nodes

3.2.2 MAS model

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 $\mathbf{A}$ 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 $[\mathbf{A}]_{ij} = 1$, where $[\cdot]_{ij}$ stands for the element in the i-th row and j-th column of the matrix. 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 Figure 3.2. The number of visited locations will be denoted by $|\mathcal{D}_i|$, where $|.|$ is the cardinality of the set. 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. No overlap is assumed between the local databases, so that $|\mathcal{D}| = \sum_{i=1}^{N} |\mathcal{D}_i|$.

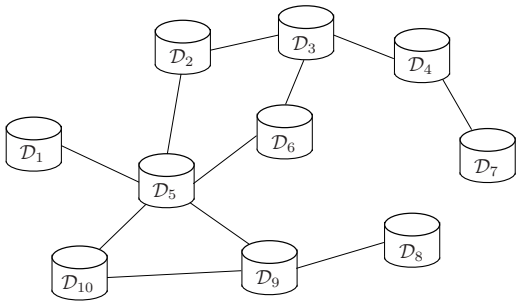


Figure 3.2: Each agent i has a local database $\mathcal{D}_i$, containing $|\mathcal{D}_i|$ entries.

3.2.3 Comparison of two receiver models

In this Thesis, with the objective of considering two different applications of REMs, distinct HetNet scenarios were considered for DICA and distributed GPR algorithms. The main difference among the RX models is related to the database sizes. In case of a WSN, a sensor is assumed to contain one measurements whereas in a MAS setting, an agent is assumed to contain multiple measurements. However, if an agent contains one measurement, a DICA based on kriging can be developed for MAS, and if a sensor contains multiple measurements, a distributed GPR can be developed for the WSN scenario.

Chapter 4

Spatial Regression

Spatial regression methods combine the RX measurements with Geo-location information to construct accurate and reliable REMs. Kriging and GPR are well-known spatial regression tools developed in the context of Geo-statistics and machine learning, respectively. These methods rely on the spatial correlation between the available measurements to construct a complete REM over the geographical area of interest. Kriging and GPR can harness both the deterministic components of the channel (path-loss), as well as the spatially correlated random components (shadowing) [24], using well-established correlation models. In this chapter, the standard Kriging and GPR approaches are modified to be suitable for wireless channel prediction assuming first a centralized setting, which means that all the measurements are available at a central entity. Note that the measurements are gathered by the sensors in the WSN setting (Kriging) and by the agents in the MAS case (GPR).

4.1 Kriging

Kriging is a very popular Geo-statistical interpolation method that depends on spatial and statistical relationships to calculate the surface [22, 90]. Kriging is a two-stage process: semivariogram modeling and Kriging prediction. The semivariogram model characterizes the spatial correlation in the data and the Kriging method employs this model to generate the best linear unbiased estimates. Kriging treats the various processes in the environment as the realizations of random processes. Since a two-dimensional space is considered, the measurements are treated as a random field. In this section, the OK method, which is the basis of contributions [26, 28] will be stated in its centralized form. Then, regression Kriging (RK) will be presented, as it is the basis for contribution [29].

4.1.1 Random field

Measurements obtained at various RX locations represent a continuous spatial phenomenon that can be modeled as random field. The random field in two dimensions

is denoted as $\{y(\mathbf{x}) : \mathbf{x} \in D \subset \mathbb{R}^2\}$, where D is the domain of interest [90]. $y(\mathbf{x})$ is decomposed into:

$$y(\mathbf{x}) = \boldsymbol{\mu}(\mathbf{x}) + \mathbf{s}(\mathbf{x}), \tag{4.1}$$

where $\boldsymbol{\mu}(\mathbf{x})$ is the mean function:

$$\begin{aligned} \boldsymbol{\mu}(\mathbf{x}) &= E[y(\mathbf{x})] \\ &= P_{\mathrm{TX}} + G_0 - 10\eta \log_{10} ||\mathbf{x}_i - \mathbf{x}_t|| , \end{aligned} \tag{4.2}$$

and $\mathbf{s}(\mathbf{x})$ is a random quantity with mean of zero and a variogram function $\gamma(\mathbf{h})$:

$$\begin{aligned} \gamma(\mathbf{h}) &= \frac{1}{2}\mathrm{Var}[\mathbf{s}(\mathbf{x}) - \mathbf{s}(\mathbf{x}+\mathbf{h})] \\ &= \frac{1}{2}\mathbb{E}[\big(\mathbf{s}(\mathbf{x}) - \mathbf{s}(\mathbf{x}+\mathbf{h})\big)^2], \end{aligned} \tag{4.3}$$

where $\mathbf{s}(\mathbf{x})$ and $\mathbf{s}(\mathbf{x}+\mathbf{h})$ are the values of shadowing $\mathbf{s}$ at location $\mathbf{x} \in D$ and $\mathbf{x}+\mathbf{h} \in D$, 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 (3.4) is known as exponential model in Geo-statistics literature and it is given by:

$$\gamma(\mathbf{h}) = \begin{cases} \sigma_0^2 + \sigma_\Psi^2 \left(1 - \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|}{d_c}\right)\right), & \|\mathbf{x}_i - \mathbf{x}_j\| > 0, \\ 0, & \mathbf{x}_i = \mathbf{x}_j, \end{cases} \tag{4.4}$$

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 correlation distance of shadowing. The quantity $\sigma_0^2 + \sigma_\Psi^2$ is known as sill and it is the value at which the semivariogram attains the range. Note that the range, sill and nugget are terms used in geo-statistics for semivariogram model parameters. The semivariogram concept will be presented in the next section.

Under the second-order stationarity assumptions on $\mathbf{s}(\mathbf{x})$, the covariance can be derived from the variogram as:

$$C(\mathbf{h}) = \sigma_0^2 + \sigma_\Psi^2 - \gamma(\mathbf{h}). \tag{4.5}$$

The behavior of semivariance and covariance over distance based on models (3.4) and (4.4), respectively, is illustrated in Figure 4.1.

4.1.2 Semivariogram estimation

The semivariogram describes the spatial variability of a random field $\mathbf{y}(\mathbf{x})$ 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 $y(\mathbf{x})$ can be obtained from the set of measurements $y(\mathbf{x}_i)$,

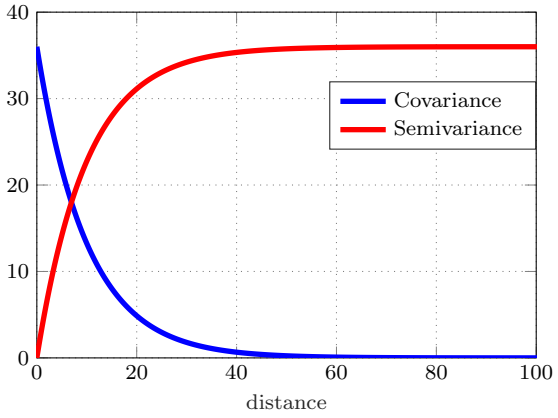


Figure 4.1: Behavior of semivariance and covariance with distance in meters

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 Matheron's method of moments estimator [90]:

$$\hat{\gamma}(\mathbf{h}) \equiv \frac{1}{2|\mathcal{N}(\mathbf{h})|} \sum_{\mathcal{N}(\mathbf{h})} (y(\mathbf{x}_i) - y(\mathbf{x}_j))^2, \tag{4.6}$$

where $y(\mathbf{x}_i)$ and $y(\mathbf{x}_j)$ are field values at locations $\mathbf{x}_i$ and $\mathbf{x}_j$, respectively. $\mathcal{N}(\mathbf{h}) = \{(\mathbf{x}_i, \mathbf{x}_j) : \mathbf{x}_i - \mathbf{x}_j \in \mathbf{h} \text{ for } i, j = 1, \ldots, N\}$ 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 Thesis, we consider the exponential model, denoted as $\overline{\gamma}(\mathbf{h})$, to model the spatial correlation, as it is equivalent to the Gudmundson model (3.4) proposed for cellular networks. Note, $\overline{\gamma}(\mathbf{h})$ is presented earlier in equation (4.4).

After the model selection, the model parameters $\hat{\boldsymbol{\theta}} = [\hat{\sigma}^2_{\Psi}, \hat{d}_c]$ can be determined by fitting the semivariogram model to the EV using the weighted least squares estimation (WLSE). 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

[91]:

$$\sigma^2_{0_{ini}} = max\left[0, \hat{\gamma}(\mathbf{h}_1) - \frac{\mathbf{h}_1}{\mathbf{h}_2 - \mathbf{h}_1}\hat{\gamma}(\mathbf{h}_2) - \hat{\gamma}(\mathbf{h}_1)\right],$$
$$d_{c_{ini}} = \frac{\mathbf{h}_{\mathrm{N}}}{2}, \tag{4.7}$$
$$\sigma^2_{0_{ini}} + \sigma^2_{\psi_{ini}} = \frac{\hat{\gamma}(\mathbf{h}_{\text{N-2}}) + \hat{\gamma}(\mathbf{h}_{\text{N-1}}) + \hat{\gamma}(\mathbf{h}_{\mathrm{N}})}{3},$$

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

4.1.3 Ordinary Kriging

Once the semivariogram estimated, it can be employed in the Kriging prediction. 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 [92]. (2) Along with the estimates, it also quantifies the Kriging variance. Kriging is a statistical interpolation method for obtaining an estimation of the received power $y(\mathbf{x}_*)$ at an unmeasured spatial location $\mathbf{x}_*$, from the weighted linear combinations of available data. This is achieved by allocating weights to each RX in such way that the Kriging variance is minimized. Let $\mathbf{y} = [(y(\mathbf{x}_1), y(\mathbf{x}_2), \ldots, y(\mathbf{x}_N))]$ be the N received power measurements at spatial locations $\mathbf{X} = [(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N)]$. Then, the kriged estimate $y(\mathbf{x}_*)$ at unmeasured spatial location $\mathbf{x}_*$ using N RXs is the weighted average of the data in its neighborhood given by [22]:

$$y(\mathbf{x}_*)|_N = \sum_{i=1}^{N} w_i(\mathbf{x}_*) y(\mathbf{x}_i), \tag{4.8}$$

where N is the number of available measurements, $y(\mathbf{x}_*)|_N$ is the received power estimate using N RXs and $w_i(\mathbf{x}_*)$ is the weight allocated for RX i from an estimation performed using N RXs. These weights fulfill the unbiased conditions of the estimator, that is:

$$\sum_{i=1}^{N} w_i(\mathbf{x}_*) = 1, \tag{4.9}$$

and the expected error between the estimated value and the actual value at location $\mathbf{x}_*$, $\mathbb{E}[y(\mathbf{x}_*) - y(\mathbf{x}_*)]$, is 0. The weights in equation (4.8) 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:

$$\sum_{i=1}^{N} w_i(\mathbf{x}_*)\overline{\Gamma}(\mathbf{x}_i, \mathbf{x}_j) + \sum_{l=1}^{L} \mathcal{L}_l(\mathbf{x}_*) f_l(\mathbf{x}_i) = \sum_{i=1}^{N} \overline{\gamma}(\mathbf{x}_i, \mathbf{x}_*), j = 1, 2, \ldots, \mathrm{N}, \tag{4.10}$$

where $\overline{\Gamma}(\mathbf{x}_i, \mathbf{x}_j)$ is the semivariogram between measurements from RX locations $\mathbf{x}_i$ and $\mathbf{x}_j$, and $\overline{\gamma}(\mathbf{x}_i, \mathbf{x}_*)$ is the semivariogram between samples from RX location $\mathbf{x}_i$ and target location $\mathbf{x}_*$. Note that $\overline{\Gamma}(\mathbf{x}_i - \mathbf{x}_j)$ and $\overline{\gamma}(\mathbf{x}_i - \mathbf{x}_*)$ are obtained from the theoretical exponential model (4.4). $\mathcal{L}_1(\mathbf{x}_*), \ldots, \mathcal{L}_L(\mathbf{x}_*)$ are the Lagrange multipliers introduced to achieve minimization [22, 90] and $f_l(\mathbf{x}_i)$ is a drift function that depends on the type of Kriging:

- For *simple Kriging* $f_l = 0, \ \forall\ l$.
- For *ordinary Kriging* $f_l = 1, \forall\ l$.
- For *universal Kriging* $f_l = f_l(\mathbf{x})$, where $f_l(\mathbf{x})$ are monomials of the spatial coordinates. From (4.2), $f_l(\mathbf{x})$ is the mean function $\boldsymbol{\mu}(\mathbf{x})$.

The Kriging system (4.10) can be rewritten for OK as:

$$\sum_{i=1}^{N} w_i(\mathbf{x}_*)\overline{\Gamma}(\mathbf{x}_i, \mathbf{x}_j) + \mathcal{L}(\mathbf{x}_*) = \sum_{i=1}^{N} \overline{\gamma}(\mathbf{x}_i, \mathbf{x}_*), j = 1, 2, \ldots, N, \tag{4.11}$$

The OK system can be represented in matrix form as:

$$\mathbf{A}\boldsymbol{\lambda} = \mathbf{b}, \tag{4.12}$$

where

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

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

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

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

$$\sigma^2(\mathbf{x}_*)|_N = \sum_{i=1}^{N} w_i(\mathbf{x}_*)\overline{\gamma}(\mathbf{x}_i, \mathbf{x}_*) + \mathcal{L}(\mathbf{x}_*). \tag{4.13}$$

4.1.4 Regression Kriging

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 [93]. The auxiliary information (known as drift in Geo-statistics) is assumed to be available at all the RXs and it is the path-loss. The key assumption of 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 [94]. In RK, the drift is accounted by the regression model, while the model residuals by OK. 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 [95–97].

RK provides an estimate of channel parameters $\hat{\boldsymbol{\theta}} = [\hat{\eta}, \hat{\sigma}^2_{\psi}, \hat{d}_c]$ and predicts $y(\mathbf{x}_*)$ at unmeasured spatial locations $\mathbf{x}_*$ using N RXs, by summing the estimated mean and the residuals:

$$y(\mathbf{x}_*) = \mu(\mathbf{x}_*) + s(\mathbf{x}_*), \tag{4.14}$$

where the mean $\mu(\mathbf{x}_*)$ is obtained using linear regression analysis and the residual $s(\mathbf{x}_*)$ is interpolated using OK. Comparing the model (4.14) and the channel model (3.2), we can deduce that the mean is the path-loss and the residual is the shadowing.

In general, there is a correlation between the path-loss component and the random shadowing component. According to [93, 98], the correlation between those components is assumed to be zero in the following. Thus, the ordinary least square (OLS) estimate of the path-loss exponent $\hat{\eta}$ can be computed as:

$$\hat{\eta} = (\mathbf{h}^{\mathrm{T}}\mathbf{h})^{-1}\mathbf{h}^{\mathrm{T}}(\mathbf{y} - \mathbf{1}^{\mathrm{T}}G_0), \tag{4.15}$$

where $\mathbf{h}^{\mathrm{T}} = -10[\log_{10}(||\mathbf{x}_1 - \mathbf{x}_t||), \log_{10}(||\mathbf{x}_2 - \mathbf{x}_t||), ..., \log_{10}(||\mathbf{x}_N - \mathbf{x}_t||)]^{\mathrm{T}}$. Once $\hat{\eta}$ is estimated, the path-loss at $\mathbf{x}_i$,

$$\mu(\mathbf{x}_i) = \mathbf{1}^{\mathrm{T}}G_0 + \eta\mathbf{h}, \tag{4.16}$$

and $\mathbf{x}_*$,

$$\mu(\mathbf{x}_*) = \mathbf{1}^{\mathrm{T}}G_0 + \eta\mathbf{h}_*, \tag{4.17}$$

where $\mathbf{h}_*^{\mathrm{T}} = -10[\log_{10}(||\mathbf{x}_1 - \mathbf{x}_*||), \log_{10}(||\mathbf{x}_2 - \mathbf{x}_*||), ..., \log_{10}(||\mathbf{x}_N - \mathbf{x}_*||)]^{\mathrm{T}}$ can be calculated. The random residual for shadowing estimation is obtained by de-trending the measurements $y(\mathbf{x}_i)$:

$$s(\mathbf{x}_i) = y(\mathbf{x}_i) - \mu(\mathbf{x}_i), \qquad i = 1, 2, \ldots, N, \tag{4.18}$$

where $s(\mathbf{x}_i)$ becomes a zero-mean random variable and then, semivariogram estimation is performed to obtain the parameters $\hat{\boldsymbol{\theta}} = [\hat{\sigma}^2_{\Psi}, \hat{d}_c]$. Once the $\hat{\boldsymbol{\theta}}$ is obtained, OK prediction can be performed to obtain the estimate $s(\mathbf{x}_*)$ and prediction variance $\sigma^2(\mathbf{x}_*)$.

Semivariogram estimation using RK has advantages, specifically when fitting a SV model and when there is trend in the data. Figure 4.2 shows EV and SV for a given set of measurements $y(\mathbf{x}_i)$, $i = 1, 2, \ldots, N$. 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. As a result, by modeling $\mathbf{y}(\mathbf{x})$ as a combination of a deterministic mean function (known as drift in Geo-statistics) and a random component, semivariogram parameters can be estimated more efficiently. Also, it can be noticed in Figure 4.2 (EV and SV of $s(\mathbf{x}_i)$) that after de-trending, the model fits the data better.

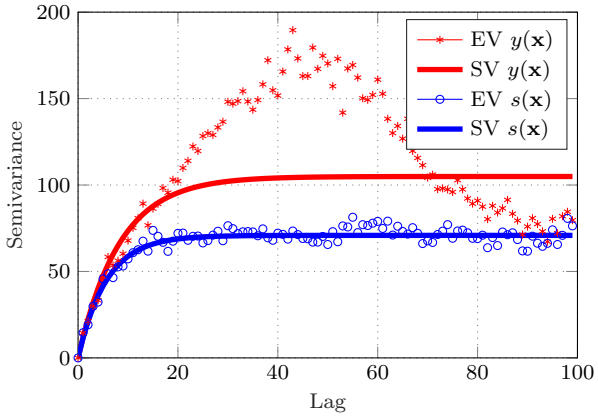


Figure 4.2: Semivariogram parameters and their impact on the behavior of semivariogram models

Finally, the RK prediction estimate and variance are obtained by summing the individual estimates and variances of OLS and OK as follows:

$$y(\mathbf{x}_*)|_N = G_0 + \hat{\eta}\mathbf{h} + \sum_{i=1}^{N} w_i(\mathbf{x}_*)s(\mathbf{x}_i), \tag{4.19}$$

$$\sigma^2(\mathbf{x}_*)|_N = \mathbf{h}_*^{\mathrm{T}}(\mathbf{h}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{h})^{-1}\mathbf{h}_* + \mathbf{b}^{\mathrm{T}}\lambda. \tag{4.20}$$

4.2 Gaussian process regression

In this section, GPR method, which is the basis of contributions [27, 30] will be stated in its centralized form. GPR is formally defined as a collection of random variables, any finite number of which shows a joint Gaussian distribution [23]. A GPR is completely specified by its mean and covariance function. Similar to Kriging, a GPR consists of two phases: training/ learning and testing/ prediction. The purpose of the learning phase is to estimate the underlying parameters of the model that best explain the data, while the prediction phase employs the learnt parameters to perform prediction at the test location. The received power $P_{\mathrm{RX}}(\mathbf{x}_i)$ at location $\mathbf{x}_i$ represents a continuous spatial process and it is modeled as a Gaussian process:

$$P_{\mathrm{RX}}(\mathbf{x}_i) \sim \mathcal{GP}\big(\mu(\mathbf{x}_i), C(\mathbf{x}_i, \mathbf{x}_j)\big), \tag{4.21}$$

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

$$\begin{aligned} \mu(\mathbf{x}_i) &= \mathbb{E}[P_{\mathrm{RX}}(\mathbf{x}_i)] \\ &= P_{\mathrm{TX}} + G_0 - 10\eta \log_{10} \|\mathbf{x}_t - \mathbf{x}_i\|, \end{aligned} \tag{4.22}$$

$$C(\mathbf{x}_i, \mathbf{x}_j) = \sigma_\psi^2 \exp\left(- \frac{\|\mathbf{x}_i - \mathbf{x}_j\|^\alpha}{d_c^\alpha} \right), \tag{4.23}$$

where d_c is the correlation distance of the shadowing. By setting $\alpha = 1$, we will obtain the Gudmundson model (3.4), which will be used in this Thesis to model the correlation properties of shadowing.

4.2.1 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 (4.22) and (4.23), the hyper-parameters of our GPR model depend on:

$$\boldsymbol{\theta} = [\eta, G_0, \sigma_\psi, d_c, \sigma_n]^{\mathrm{T}}. \tag{4.24}$$

The vector $\boldsymbol{\theta}$ can be learned by maximizing the log-likelihood with respect to $\boldsymbol{\theta}$, which is given by

$$\hat{\boldsymbol{\theta}} = \arg\max_{\theta} \log p(\mathbf{y} \mid \mathbf{X}, \boldsymbol{\theta}), \tag{4.25}$$

$$\log p(\mathbf{y} \mid \mathbf{X}, \boldsymbol{\theta}) = -\frac{1}{2} \left(\mathbf{y} - \boldsymbol{\mu}(\mathbf{X})\right)^{\mathrm{T}} \mathbf{K}^{-1} \left(\mathbf{y} - \boldsymbol{\mu}(\mathbf{X})\right) - \frac{1}{2} \log |\mathbf{K}|, \tag{4.26}$$

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 $\mathbf{X}$, and $\boldsymbol{\mu}(\mathbf{X})$ is the $N \times 1$ mean vector containing path-loss components at training locations $\mathbf{X}$ where the i-th entry is computed with the help of (4.22), i.e., $\boldsymbol{\mu}(\mathbf{X}) = [\mu(\mathbf{x}_1), \mu(\mathbf{x}_2), \ldots, \mu(\mathbf{x}_N)]^{\mathrm{T}}$. Note that $\boldsymbol{\mu}$ and $\mathbf{K}$ depend on $\boldsymbol{\theta}$.

4.2.2 Prediction

Once the hyper-parameter vector $\boldsymbol{\theta}$ 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 as $p\big(P_{\mathrm{RX}}(\mathbf{x}_*) \mid \mathcal{D}, \boldsymbol{\theta}, \mathbf{x}_*\big)$ and follows a Gaussian distribution with mean $\bar{P}_{\mathrm{RX}}(\mathbf{x}_*)$ and variance $\sigma^2_{\mathrm{RX}}(\mathbf{x}_*)$ given by [23]:

$$\bar{P}_{\mathrm{RX}}(\mathbf{x}_*) = \mu(\mathbf{x}_*) + \mathbf{k}_*^{\mathrm{T}} \mathbf{K}^{-1} \big(\mathbf{y} - \boldsymbol{\mu}(\mathbf{X})\big), \tag{4.27}$$

$$\sigma^2_{\mathrm{RX}}(\mathbf{x}_*) = k_{**} - \mathbf{k}_*^{\mathrm{T}} \mathbf{K}^{-1} \mathbf{k}_*, \tag{4.28}$$

where $\mathbf{k}_* = C(\mathbf{x}_i, \mathbf{x}_*)$ 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, 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 $\mathcal{D}$.

4.3 Simulations

In this section, simulation results are provided to compare the performance over a centralized setting of OK, RK, and GPR. The scenario under consideration is a LTE-sensor network in a sub-urban environment in a square area of 200 m $\times$ 200 m. The LTE pico-cell BS is placed at location $\mathbf{x} = [0, 0]^{\mathrm{T}}$ and the WSN consisting of N sensor nodes with

2 m inter-node spacing are randomly and uniformly deployed. We simulate a realistic radio environment based on the propagation model (3.2) 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 4.1 [99, 100]. The complete REM is obtained by performing predictions in a fine grid of $N_L = 2601$ locations, when considering a 4 m resolution grid on the square area under study.

Table 4.1: 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

The accuracy of prediction is an important criteria to benchmark the performance of spatial interpolation methods. In order to analyze the accuracy of REM reconstruction, the mean squared error (MSE) between the ground truth $y(\mathbf{x}_{*,l})$ and the predicted mean value $\hat{y}(\mathbf{x}_{*,l})$ for each spatial location l, where $l = \{1, 2, ..., N_L\}$ is considered. It is defined as:

$$\text{MSE} = \frac{1}{N_L} \sum_{l=1}^{N_L} \left(y(\mathbf{x}_{*,l}) - P_{\text{RX}}(\mathbf{x}_{*,l}) \right)^2 , \tag{4.29}$$

Figure 4.3 shows the mean MSE for different number of measurements for centralized REM reconstruction, averaged over 100 realization of the channel field. It also illustrates the performance comparison of the spatial regression methods considered in this chapter. In a centralized setup, the CC collects measurements from sensors/ agents and reconstructs the REM centrally. Irrespective of RXs, the central database contains one measurement for each location. The following conclusion can be drawn: (1) Prediction error in-terms of MSE decreases with the increase in number of measurements; (2) when the number of measurements are higher, all the methods perform equally well. However, when $N < 10$, GPR performs better than RK and OK, and when $N > 10$, GPR and

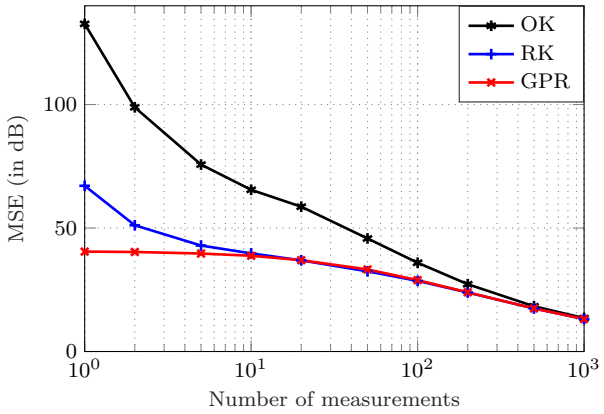


Figure 4.3: MSE versus number of measurements for OK, RK and GPR

RK provide the same performance. In terms of Kriging, RK outperforms OK as it can capture the path-loss components in addition to the shadowing (3) Kriging and GPR yield the same result when the same correlation model for characterizing shadowing is used. Thus, the above comparison is fair.

Chapter 5

Distributed Incremental Clustering Algorithm

In this chapter, one of the two approaches considered in this Thesis to address the challenges of spatial prediction for REM reconstruction is presented. The scenario under consideration is the WSN setting introduced in section 3.2.1, where the measurements obtained by a sensor node will be the basis for reconstructing the REM. 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, a novel DICA[1] based on the RK method is proposed. The objective is REM reconstruction in terms of average received power at locations where no sensor measurements are available. The 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 $\mathcal{O}(n^3)$. 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 OK variance. The complexity of the proposed algorithm is analyzed and simulation results are presented to showcase the algorithm efficacy to field reconstruction.

5.1 Problem statement

The objective is to obtain a high-quality REM for future 5G applications or in other words, to perform:

1. *Distributed semivariogram/ parameter estimation:* Estimate the channel parameters $\boldsymbol{\theta} = [\eta, G_0, \sigma_\Psi^2, d_c, \sigma_n^2]^{\mathrm{T}}$ from measurements $y(\mathbf{x}_i)$, $i = 1, 2, \ldots, N$.
2. *Distributed Prediction:* Obtain an estimate of the field value $y(\mathbf{x}_*)$ at location $\mathbf{x}_*$ where value is not known, using the least number of geo-location aware sensor node measurements $y(\mathbf{x}_i)$.

[1]The contributions are based on [26, 28, 29]

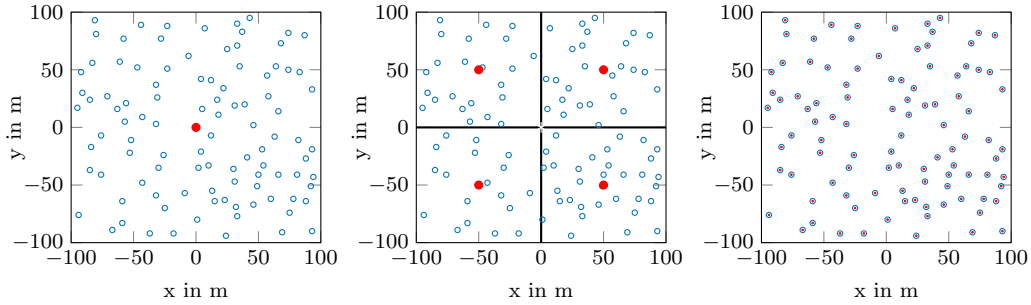


Figure 5.1: Random deployment of 225 sensor nodes. (Left) Centralized estimation, (center) Partitioned estimation, (right) Distributed estimation

5.2 Field reconstruction

In this Thesis, the performance of three approaches ranging from centralized to completely distributed over the same WSN topology (see Figure 5.1) will be compared and their effect on the reconstruction quality will be studied. The estimation quality depends on the sensor node deployment, the quality of measurements and the number of sensor nodes.

5.2.1 Centralized estimation

Sensor nodes send their measurements and positions to a central node for global field reconstruction. The solution to the field reconstruction problem is known only to the central node, which is adequate in many applications. However, for dynamic applications where the solution has to be known locally at the sensor nodes, the central node can communicate the solutions to the sensor nodes for local field reconstruction. This implies that additional communication requirements need to be added to the basic forwarding of measurements via multi-hop to the central node. In the centralized scheme, the complete reconstruction process can be affected due to the malfunctioning of the central node or appearance of faulty sensor measurements.

5.2.2 Partitioned estimation

The 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 estimation problem is known only by the subregion heads. The communication requirements are reduced due to short-range transmissions within the subregion. The malfunctioning of a subregion head or the faulty sensor measurements may influence a subregion without having an effect on the other subregion. However, since the inter-subregion sensor node sharing is not allowed, border effects may appear

in the reconstructed REM. The intensity of the border effect depends on the WSN size.

5.2.3 Distributed estimation

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. In this case, communication requirements are significantly reduced and a small region is affected in case of a sensor node failure or faulty measurement. 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. This is the architecture followed by DICA.

5.3 Algorithm description

In centralized estimation, the Kriging equations (4.19) and (4.20) use all the available sensor measurements to calculate the spatial interpolations. In this Thesis, the objective is to minimize the number of measurements through a DICA, which is based on RK, and hence, it is named as 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 distributed ordinary least squares (D-OLS) and a distributed cluster based OK (DC-OK) algorithms. The complete algorithm is represented in the flowchart (see Figure 5.2). For comparison purpose, a version of DICA based on OK (DICA-OK) has been explored. In order to avoid replication, a flowchart of DICA-OK is presented in Figure 5.3, from which the operation of the algorithm can easily be understood.

DICA-RK consists of the following phases:

5.3.1 Neighbor discovery

A broadcasting protocol is employed in order to find the set of sensor nodes in the neighborhood of $\mathbf{x}_*$, i.e., $\mathcal{N}_{\mathbf{x}_*}$, and the node closest to $\mathbf{x}_*$. Note that the neighborhood of $\mathbf{x}_*$ includes all the nodes that are in-range and at one-hop distance from $\mathbf{x}_*$. Neighbor discovery is performed by each node to gather information about its multi-hop neighborhood. During this stage, each node broadcasts 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.

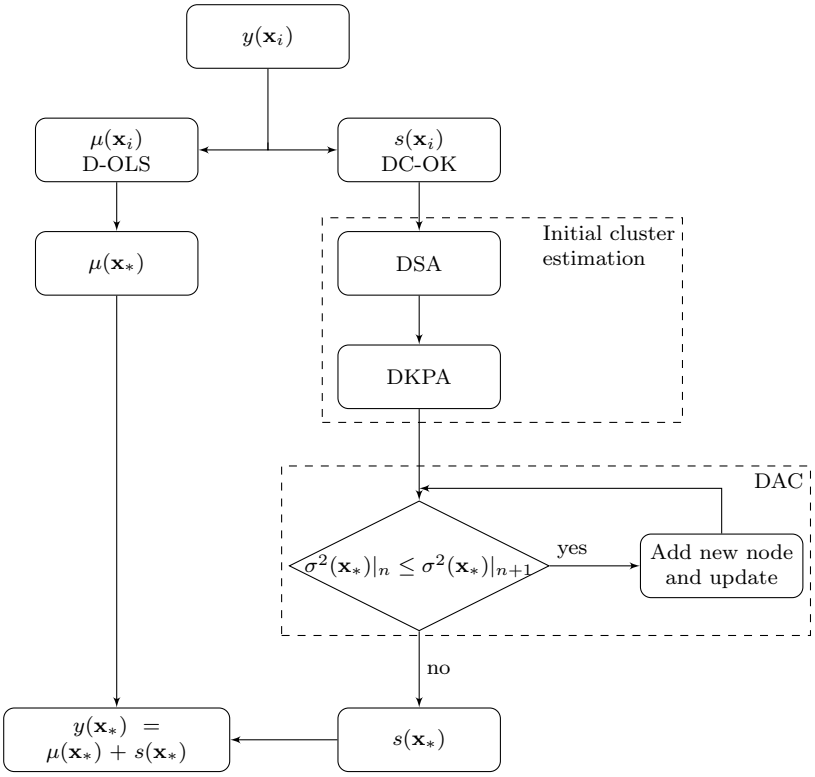


Figure 5.2: DICA-RK algorithm

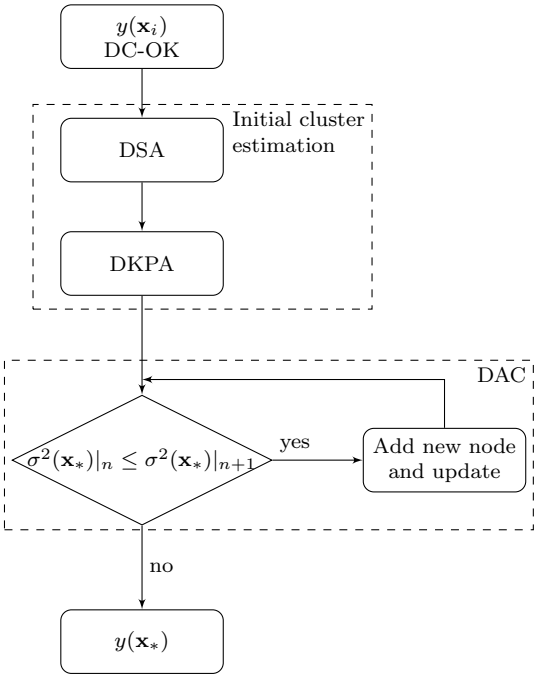


Figure 5.3: DICA-OK algorithm

5.3.2 Distributed ordinary least squares

Once the sensor nodes are aware about the information of their multi-hop neighborhood, they can estimate $\hat{\eta}$ and thereby compute $\mu(\mathbf{x}_*)$. OLS equation (4.15) can be solved using standard distributed methods, such as distributed recursive least-squares algorithm from [101]. Each node can compute the path-loss component and subtract it with its own measurement to obtain the shadowing component $s(\mathbf{x}_i)$ using (4.18).

5.3.3 Distributed cluster based ordinary Kriging

For shadowing estimation, equations (4.8) and (4.11) 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 [102]. Due to this, the assumption of stationarity of the mean of the shadowing $\mathbf{s}(\mathbf{x})$ can be restricted 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 (4.11) 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, a distributed cluster based OK (DC-OK) algorithm is presented to minimize the computation complexity of the shadowing estimation, by 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 and that the clusters can overlap. 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 an initial cluster estimation followed by adaptive cluster estimation.

5.3.3.1 Initial cluster estimation

An initial set of one-hop, in-range, p sensor nodes, that are closest to unmeasured spatial location $\mathbf{x}_*$, forms an initial cluster and begins the initial estimation process. In this Thesis, the WLSE analytical fitting method is employed and the initial values of the semivariogram model parameters is chosen using equation (4.7). To compute the initial values $\sigma^2_{0_{ini}}$, $d_{c_{ini}}$ and $\sigma^2_{0_{ini}} + \sigma^2_{\psi_{ini}}$, sensor nodes require the values of $\hat{\gamma}(\mathbf{h})$ from three node measurements. Therefore, 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:

1. 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.

Algorithm 1 Distributed semivariogram algorithm (DSA)

$\mathcal{N}_{\mathbf{x}_*}$: Set of sensor nodes in $\mathbf{x}_*$ neighborhood // p: Number of nodes in the initial cluster // $\mathbf{h}$: Lag distance // $d(\mathbf{x}_i, \mathbf{x}_j)$: 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 $\mathbf{x}_*$, $i \in \mathcal{N}_{\mathbf{x}_*}$, activates and looks for next closest sensor node j to $\mathbf{x}_*$, $j \in \mathcal{N}_{\mathbf{x}_*}$.
2: Sensor node i sends a packet containing $y(\mathbf{x}_i)$, location $\mathbf{x}_i$ and $\hat{\gamma}(\mathbf{h})$ to sensor node j.
3: Sensor node j receives the packet and calculates $d(\mathbf{x}_i, \mathbf{x}_j)$ and $\hat{\boldsymbol{\gamma}}(\mathbf{h})$ using (4.6).
4: Sensor nodes $i, j \in \mathcal{N}_p$ look for the next closest sensor node k to $\mathbf{x}_*$.
5: **while** $|N_p| < p$ **do**
6: **if** exists a sensor node k closest to $\mathbf{x}_*$, $k \in \mathcal{N}_{\mathbf{x}_*}$ **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.

Due to this iterative process, an initial cluster is built, the field similarity is known and the EV is estimated among all the sensor nodes in the cluster. To compute the EV, a distributed semivariogram algorithm (DSA) (see Algorithm 1) is developed which solves equation (4.6) in a distributed way. Sensor node i closest to $\mathbf{x}_*$, $i \in \mathcal{N}_{\mathbf{x}_*}$, activates and sends a packet containing its field measurement value and geo-location information to the next closest sensor node j to $\mathbf{x}_*$, $j \in \mathcal{N}_{\mathbf{x}_*}$, where $\mathcal{N}_{\mathbf{x}_*}$ is the set of one-hop, in-range sensor nodes that can be found in $\mathbf{x}_*$ 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 $\mathbf{x}_*$, $k \in \mathcal{N}_{\mathbf{x}_*}$, 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.

2. Distributed Kriging prediction:

After the parameter estimation, the sensor nodes in the initial cluster are able to predict $s(\mathbf{x}_*)|_p$ by solving the Kriging estimator (4.8) and Kriging system (4.11) 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 $\overline{\Gamma}(\mathbf{x}_i, \mathbf{x}_j)$ between all the sensor nodes in the cluster and the semivariogram $\overline{\gamma}(\mathbf{x}_i - \mathbf{x}_*)$ to the target location $\mathbf{x}_*$. The Kriging system is solved by the modified Gaussian elimination method as described in Algorithm 2. We adapted the algorithm from [103] to work with the cluster of sensor nodes. The closest sensor node to $\mathbf{x}_*$, $m \in \mathcal{N}_p$, is chosen to begin the iterative process. The sensor node m also constructs

Algorithm 2 Distributed Kriging prediction algorithm (DKPA)

$\mathbf{r}_i$: row of Kriging system (4.11) created by sensor node i // $\overline{\Gamma}(\mathbf{x}_i, \mathbf{x}_j)$: 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 $\overline{\Gamma}(\mathbf{x}_i, \mathbf{x}_j),\ \forall j \in \mathcal{N}_p$ using stored distances and model parameters.
3: Each sensor node $i \in \mathcal{N}_p$ assigns values for all elements in row $\mathbf{r}_i$:
$\mathbf{r}_i(1:p) \leftarrow [\overline{\gamma}(\mathbf{x}_i, \mathbf{x}_1), \overline{\gamma}(\mathbf{x}_i, \mathbf{x}_2), \ldots, \overline{\gamma}(\mathbf{x}_i, \mathbf{x}_p)]$
$\mathbf{r}_i(p+1:p+2) \leftarrow [\overline{\gamma}(\mathbf{x}_i, \mathbf{x}_*), 1]$
4: Closest sensor node, $m \in \mathcal{N}_p$, to $\mathbf{x}_*$ 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 $w_{i|p}(\mathbf{x}_*)$ for each sensor node i and $\mathcal{L}$ are obtained.

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 $w_{i|p}(\mathbf{x}_*)$ 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 $s(\mathbf{x}_*)|_p$ can be obtained from the Kriging estimator (4.8) in a distributed way, i.e., by multiplying the weight of each sensor node with its measurement and summing at each sensor node.

5.3.3.2 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 (DAC) is proposed, which employs the OK variance in combination with the Kriging system to optimize the cluster size. The clustering procedure for shadowing estimation is explained in Algorithm 3.

Subsequent to the initial estimation of $s(\mathbf{x}_*)|_p$, the Kriging variance $\sigma^2(\mathbf{x}_*)|_p$ is computed. Note that the Kriging variance given by equation (4.13) can be computed distributively by summing the local multiplications with the Lagrange multiplier. A new sensor node $q \in \mathcal{N}_{\mathbf{x}_*}$ is added to the initial cluster, and the estimate $s(\mathbf{x}_*)|_{p+1}$ and the Kriging variance $\sigma^2(\mathbf{x}_*)|_{p+1}$ 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 [104]:

$$s(\mathbf{x}_*)|_{p+1} = s(\mathbf{x}_*)|_p - w_{q|p+1}(\mathbf{x}_*)[s(\mathbf{x}_q)|_p - s(\mathbf{x}_q)], \tag{5.1}$$

$$w_{i|p+1}(\mathbf{x}_*) = w_{i|p}(\mathbf{x}_*) - w_{q|p+1}(\mathbf{x}_*)w_{i|p}(\mathbf{x}_q), i = 1, 2, \ldots, p, \tag{5.2}$$

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

The Kriging variance of clusters with p and $p+1$ sensor nodes are compared. If $\sigma^2(\mathbf{x}_*)|_p$ is greater than $\sigma^2(\mathbf{x}_*)|_{p+1}$, 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 $\sigma^2(\mathbf{x}_*)|_{p+t+1}$ is higher or equal to the one with $\sigma^2(\mathbf{x}_*)|_{p+t}$. 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 $\mathbf{x}_*$ 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. The update models (5.1) and (5.2) for adding one sensor node can be extended to the case when a group of t sensor nodes are added:

$$s(\mathbf{x}_*)|_{p+t} = s(\mathbf{x}_*)|_p - \sum_{j=1}^{t} w_{p+j|p+t}(\mathbf{x}_*)[s(\mathbf{x}_{p+j})|_p - s(\mathbf{x}_{p+j})], \tag{5.3}$$

$$w_{i|p+t}(\mathbf{x}_*) = w_{i|p}(\mathbf{x}_*) - \sum_{j=1}^{t} w_{p+j|p+t}(\mathbf{x}_*)w_{i|p}(\mathbf{x}_{p+j}). \tag{5.4}$$

The advantage of these update models is that the Kriging estimate and weights can be updated at once instead of recalculating from the beginning. This model is convenient for estimation using a large number of sensor nodes, as it reduces the communication burden and the computational cost to solve the Kriging system.

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 // $s(\mathbf{x}_*)|_n$: Final field estimate value obtained through clustering // $\mathbf{x}_*$: location where field value is not known // t: Number of sensor nodes added to the initial cluster // d_c: Range parameter of semivariogram model // $d(\mathbf{x}_i, \mathbf{x}_*)$: Euclidean distance between $\mathbf{x}_i$ and $\mathbf{x}_*$.

1: **for** all $\mathbf{x}_*$ **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 $s(\mathbf{x}_*)|_n$ and $s(\mathbf{x}_*)|_{n+1}$.
5: Kriging variances $\sigma^2(\mathbf{x}_*)|_n$ and $\sigma^2(\mathbf{x}_*)|_{n+1}$ are computed by using equation (4.13).
6: **if** $\sigma^2(\mathbf{x}_*)|_n \leq \sigma^2(\mathbf{x}_*)|_{n+1}$ **then**
7: Terminate update process and cluster formation.
8: Field estimate value at location $\mathbf{x}_*$ and Kriging variance are obtained.
 $s(\mathbf{x}_*) = s(\mathbf{x}_*)|_n$
 $\sigma^2(\mathbf{x}_*) = \sigma^2(\mathbf{x}_*)|_n$
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**

5.4 Complexity analysis

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

- 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 $cost_{\mathrm{DSA}} \in \mathcal{O}(n)$.

- 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 (4.8) 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 $cost_{\mathrm{DKPA}} \in \mathcal{O}(n^3)$.

- In DAC, a cluster is formed with $n = p + t$ sensor nodes, where p is the initial

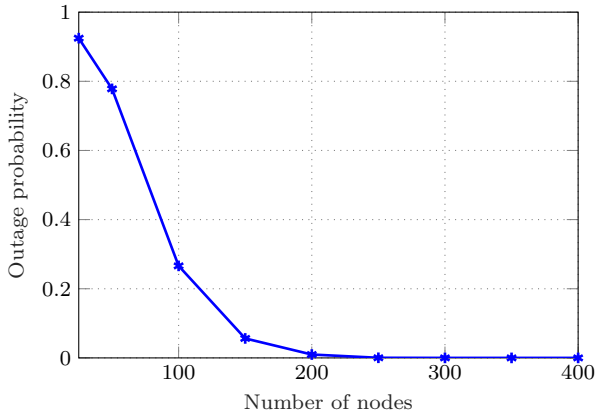


Figure 5.4: Outage probability versus number of sensor nodes

set of sensor nodes to begin the estimation and t is the number of additional sensor nodes incorporated to improve the quality. Spatial estimation $s(\mathbf{x}_*)|_p$ using the initial cluster requires one operation. However, $s(\mathbf{x}_*)|_t$ 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 $|t+1| \times (cost_{DSA} + cost_{DKPA} + cost_{DAC}) \in \mathcal{O}(n^3)$. If the estimates are required at m unmeasured locations, the complexity further increases to $\mathcal{O}(mn^3)$.

5.5 Simulation results

To assess the performance of the proposed DICA-RK, simulation parameter values of Table 4.1 are considered. Sensor node coverage radius of 21 m is obtained by considering the channel model (3.2). 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.

Important factors for implementing the DICA-RK are the network size and the initial cluster size. Based on the model (4.7) presented in section 4.1.2, DICA-RK 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, the outage probability is defined as the probability that each spatial location $\mathbf{x}_*$ is not in the communication range of three sensor nodes for initial cluster formation. Figure 5.4 illustrates the outage probability for different WSNs sizes in the scenario under consid-

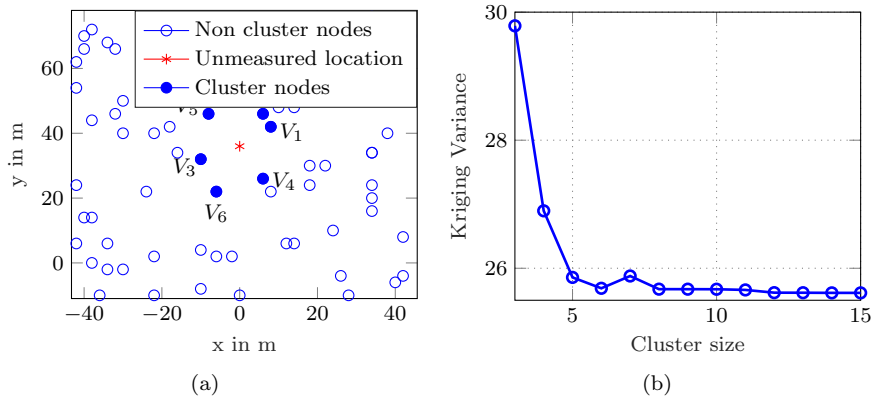


Figure 5.5: (a) Figure 5.1 zoomed to show cluster formation (b) Kriging variance versus cluster size for location $\mathbf{x_0} = [0, 36]^{\mathrm{T}}$

eration.

In order to demonstrate the cluster formation procedure, the unmeasured location $\mathbf{x}_* = [0, 36]^{\mathrm{T}}$ of Figure 5.5(a) is considered 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 $\mathbf{x}_*$. 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 Figure 5.5(b), which illustrates how Kriging variance changes as new sensor nodes are added to the initial cluster. As seen from the Figure 5.5(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.

In Figure 5.6, the DICA-RK is compared with a centralized and a partitioned estimation, which are labeled as Centralized RK and Partitioned RK, respectively. DICA-OK stands for DICA based on OK and is proposed in our previous work [26, 28]. Appendix A includes results for DICA-OK complexity analysis, in terms of average size of clusters required to build REMs and interpolated maps using multiple TXs. In Figure 5.6, the mean MSE is averaged over 50 realizations of the channel field. 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 Figure 5.6, 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

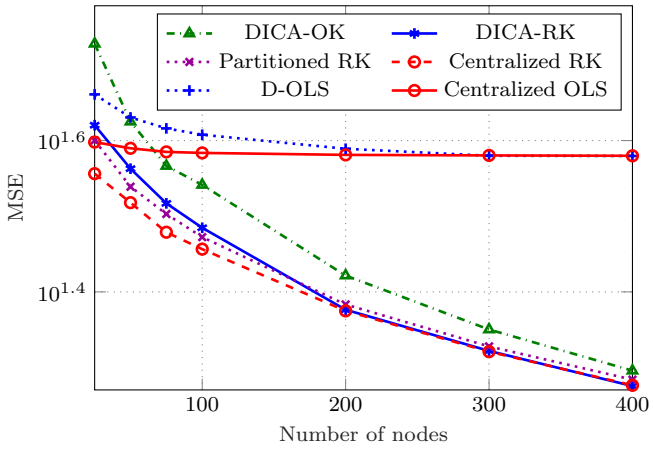


Figure 5.6: Comparison of DICA-RK with centralized and partitioned estimation

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 Figure 5.9(b). Finally, it can be observed 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.

Localization is subject to various error sources. Hence, sensor nodes are now assumed to obtain noisy location measurements $\mathbf{X}_i = \mathbf{X}_i + \mathbf{L}_i$, where $L_i \sim \mathcal{N}(0, \sigma^2)$ is zero mean additive white Gaussian noise with the location error standard deviation σ. Parameter σ is drawn from an exponential distribution, i.e., $\sigma \overset{iid}{\sim} \exp(\lambda)$, where λ is the average error location standard deviation. True location is considered with $\lambda = 6\,\mathrm{m}$ and $\lambda = 0\,\mathrm{m}$ for estimated location N_L. It can be seen in the Table 5.1 how the performance degrades under the impact of location noise.

Table 5.1: 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

The performance of DICA-RK is also compared with two classical interpolation methods such as natural neighbor and cubic spline interpolation. Figure 5.7 demonstrates that the MSE for the classical interpolation methods is significantly higher than for the DICA-RK. In order to analyze the prediction precision, a contour plot (see Figure 5.8) is

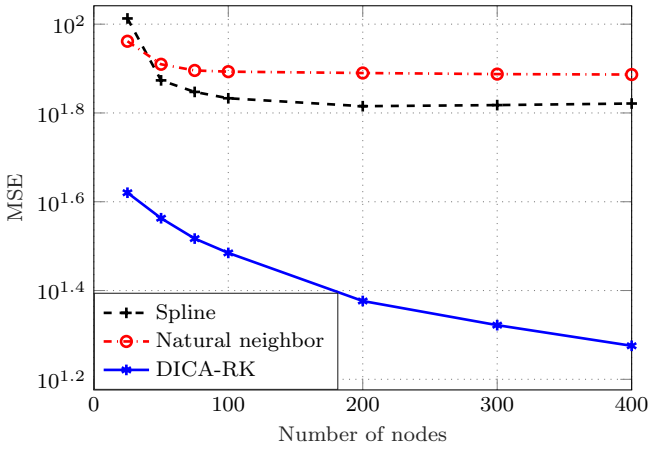


Figure 5.7: Comparison of DICA-RK with classical interpolation methods

obtained where the intensity of the spatial field exceeding -30dB is considered. The level curve for the DICA-RK is very close to the actual area and shows that it is possible to localize the active TX area effectively. To illustrate the excellent reconstruction quality of DICA-RK, interpolated maps are presented in Figure 5.9, where all the interpolation methods 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.

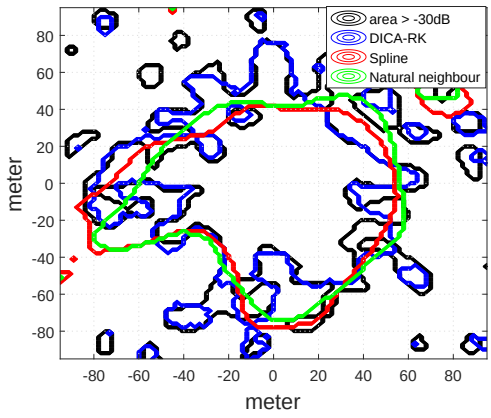
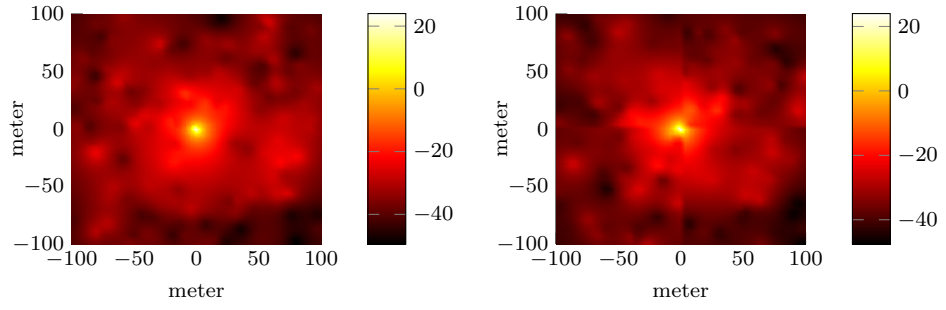
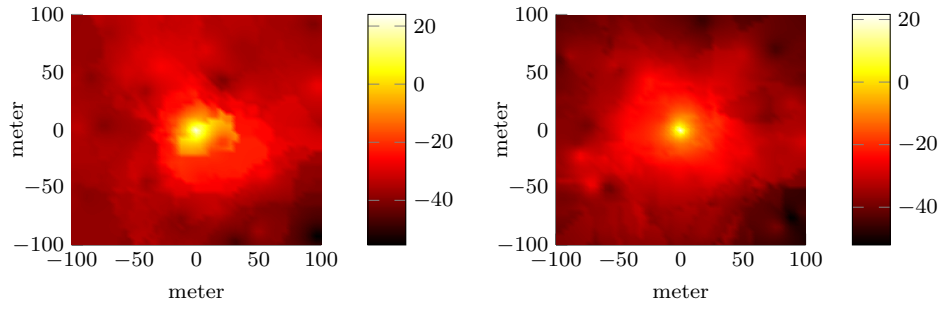


Figure 5.8: Level curves for Z>-30dB with N=300



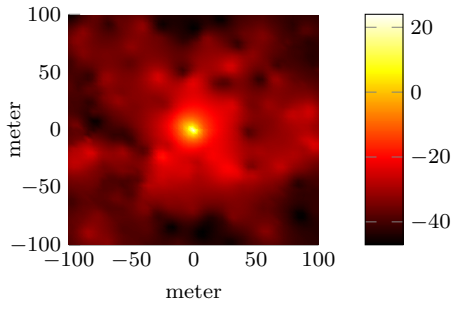
(a) Centralized RK N=200

(b) Partitioned RK N=200



(c) DICA OK N=50

(d) DICA RK N=50



(e) DICA RK N=200

Figure 5.9: Interpolated maps for different estimation frameworks

Chapter 6

Distributed Gaussian Process Regression

In this chapter, a second approach is considered to address the challenges presented in this Thesis. The scenario under consideration is the MAS setting introduced in section 3.2.2, where measurements gathered by agents will be the basis for REM reconstruction. 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. In this chapter, a distributed GPR[1] framework for REM reconstruction in terms of the received power in MAS is presented. The proposed solution combines a Bayesian committee machine (BCM) with an average consensus scheme, thus distributing not only the memory, but also computational and communication loads. Through Monte Carlo simulations, the performance of the proposed GPR is demonstrated.

6.1 Problem statement

The aim is to perform:

1. *Distributed Learning*: Compute one common estimate of $\theta = [\eta, \sigma_\Psi^2, d_c]^{\mathrm{T}}$ for all agents using their respective databases $\mathcal{D}_i = \{\mathbf{X}_i, \mathbf{y}_i\}$, through distributed processing.

2. *Distributed Prediction*: Determine one common 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.

[1] The contributions are based on [27, 30]

6.2 Distributed Gaussian process regression

In centralized GPR, the CC performs the prediction computations (i.e., mean (4.27) and variance (4.28)) by collecting the measurements from all the agents. In this section, a distributed GPR where each agent performs the prediction computations independently in a distributed way is presented. For this, independence among the agent databases $\mathcal{D}_i$ is assumed, similar to [31]. Each agent determines a local pdf conditioned on the local database, and subsequently combines the local predictions using a BCM from [32] to obtain the overall predictive pdf $p\big(P_{\mathrm{RX}}(\mathbf{x}_*) \mid \mathcal{D}, \theta, \mathbf{x}_*\big)$ for a test location $\mathbf{x}_*$. The BCM is used to develop a distributed GPR in MAS by adapting a distributed average consensus algorithm [105]. Distributed GPR consists of two phases: distributed learning and distributed prediction.

6.2.1 Distributed learning

For distributed learning, each mobile agent relies on its own database. Hence, the following GP approximation, similar to [32], can be performed:

$$p\left(\mathcal{D}_i \mid \mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_{i-1}, \mathcal{D}_{i+1}, \ldots, \mathcal{D}_N, \boldsymbol{\theta}\right) \approx p\big(\mathcal{D}_i \mid \boldsymbol{\theta}\big). \tag{6.1}$$

According to [32], (6.1) 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 [32], by partitioning the database $\mathcal{D}$ and assigning the database of each cluster $\mathcal{D}_i$ to a separate agent i. As a result:

$$p(\boldsymbol{\theta}|\mathcal{D}) \propto p(\boldsymbol{\theta}) \prod_{i=1}^{N} p(\mathcal{D}_i|\boldsymbol{\theta}), \tag{6.2}$$

where $p(\mathcal{D}_i|\boldsymbol{\theta})$ is the likelihood (4.26), and $p(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\mu}_p, \boldsymbol{\Sigma}_p)$ is a Gaussian prior with mean $\boldsymbol{\mu}_p = [\mu_\eta, \mu_{\sigma_\psi^2}, \mu_{d_c}]^{\mathrm{T}}$ and a diagonal covariance matrix $\boldsymbol{\Sigma}_p = diag([\sigma_\eta^2, \sigma_{\sigma_\psi^2}^2, \sigma_{d_c}^2]^{\mathrm{T}})$. Applying Bayes' rule:

$$p(\boldsymbol{\theta}|\mathcal{D}) \propto \frac{\prod_{i=1}^{N} p(\boldsymbol{\theta}|\mathcal{D}_i)}{\left(p(\boldsymbol{\theta})\right)^{N-1}}. \tag{6.3}$$

6.2.1.1 Agent learning

The goal is to determine the posterior distribution $p(\boldsymbol{\theta}|\mathcal{D}_i)$ of agent i by using its own database. Let $p(\boldsymbol{\theta})$ denote the joint probability distribution of $\boldsymbol{\theta} = [\eta, \sigma_\psi^2, d_c]$. The posterior distribution of $\boldsymbol{\theta}$ can be determined through application of Baye's theorem:

$$p(\boldsymbol{\theta}|\mathcal{D}_i) \propto p(\mathcal{D}_i|\boldsymbol{\theta}) p(\boldsymbol{\theta}). \tag{6.4}$$

To evaluate the moments of the posterior distribution $p(\boldsymbol{\theta}|\mathcal{D}_i)$, a simple approach would be to uniformly discretize the domain of $\boldsymbol{\theta}$. As this operation scales exponentially with the dimensionality of $\boldsymbol{\theta}$, an attractive approach is to represent the distribution by a finite

list of L independent particles. The IS method is adapted to generate such a particle system $\{(w^{(l)}, \boldsymbol{\theta}^{(l)})\}_{l=1}^{L}$, where $w^{(l)}$ is the weight of the particle $\boldsymbol{\theta}^{(l)}$. IS technique consists of choosing an importance sampling distribution $q(\boldsymbol{\theta})$ from which particles are easy to sample. The generic IS algorithm can be summarized as follows:

1. Sample L particles according to importance distribution, $\boldsymbol{\theta}^{(l)} \sim q(\boldsymbol{\theta}), l = \{1, 2, \ldots, L\}$.

2. For each particle, compute the unnormalized importance weights:

$$\tilde{w}^{(l)} \propto p(\mathcal{D}_i|\boldsymbol{\theta}^{(l)}) \frac{p(\boldsymbol{\theta}^{(l)})}{q(\boldsymbol{\theta}^{(l)})}. \tag{6.5}$$

3. Normalize weights so that they add up to unity:

$$w^{(l)} = \frac{\tilde{w}(\boldsymbol{\theta}^{(l)})}{\sum_{l=1}^{L} \tilde{w}(\boldsymbol{\theta}^{(l)})}. \tag{6.6}$$

After the normalization of weights, the set of particles and weights of the particle system $\{(w^{(l)}, \boldsymbol{\theta}^{(l)})\}_{l=1}^{L}$ is approximately distributed as the posterior distribution $p(\boldsymbol{\theta}|\mathcal{D}_i)$. As a result, agent i consists of $p(\boldsymbol{\theta}|\mathcal{D}_i)$ described by its mean $\boldsymbol{\mu}_i$ and covariance matrix $\boldsymbol{\Sigma}_i$:

$$p(\boldsymbol{\theta}|\mathcal{D}_i) = \mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i). \tag{6.7}$$

$\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ are used to generate a new $q(\boldsymbol{\theta}^{(l)})$ and the IS algorithm is iteratively repeated until a reliable estimate of $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ are obtained.

6.2.1.2 Global learning

After each agent i computes its pdf $p(\boldsymbol{\theta}|\mathcal{D}_i)$ by conditioning $\boldsymbol{\theta}$ on its respective database $\mathcal{D}_i = \{\mathbf{X}_i, \mathbf{y}_i\}$, under BCM approximation, the global pdf $p(\boldsymbol{\theta}|\mathcal{D})$ is obtained by communicating with the neighbors. Since the product and ratio of Gaussian distributions are Gaussian, (6.3) can be rewritten as:

$$p(\boldsymbol{\theta}|\mathcal{D}) \propto \frac{\mathcal{N}(\boldsymbol{\mu}_{\mathrm{n}}, \boldsymbol{\Sigma}_{\mathrm{n}})}{\mathcal{N}(\boldsymbol{\mu}_{\mathrm{d}}, \boldsymbol{\Sigma}_{\mathrm{d}})}, \tag{6.8}$$

with

$$\boldsymbol{\mu}_{\mathrm{n}} = \frac{\sum_{i=1}^{N} \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\mu}_i}{\sum_{i=1}^{N} \boldsymbol{\Sigma}_i^{-1}}, \tag{6.9}$$

$$\boldsymbol{\Sigma}_{\mathrm{n}} = \left[\sum_{i=1}^{N} \boldsymbol{\Sigma}_i^{-1} \right]^{-1}, \tag{6.10}$$

$$\boldsymbol{\mu}_{\mathrm{d}} = \boldsymbol{\mu}_{\mathrm{p}}, \tag{6.11}$$

$$\boldsymbol{\Sigma}_{\mathrm{d}} = \frac{\boldsymbol{\Sigma}_{\mathrm{p}}}{N-1}, \tag{6.12}$$

such that

$$p(\boldsymbol{\theta}|\mathcal{D}_i) \propto \mathcal{N}\left(\boldsymbol{\mu}_{\mathrm{g}}, \boldsymbol{\Sigma}_{\mathrm{g}}\right), \tag{6.13}$$

with

$$\boldsymbol{\Sigma}_{\mathrm{g}} = \left(\boldsymbol{\Sigma}_{\mathrm{n}}^{-1} - \boldsymbol{\Sigma}_{\mathrm{d}}^{-1}\right)^{-1}, \tag{6.14}$$

$$\boldsymbol{\mu}_{\mathrm{g}} = \boldsymbol{\Sigma}_{\mathrm{g}}(\boldsymbol{\mu}_{\mathrm{n}}\boldsymbol{\Sigma}_{\mathrm{n}}^{-1} - \boldsymbol{\mu}_{\mathrm{d}}\boldsymbol{\Sigma}_{\mathrm{d}}^{-1}). \tag{6.15}$$

A distributed learning is achieved through the following steps. First, agents compute the terms $\boldsymbol{\mu}_{\mathrm{d}}$ and $\boldsymbol{\Sigma}_{\mathrm{d}}$ by using their prior knowledge. Later, they compute $\boldsymbol{\mu}_{\mathrm{g}}$ and $\boldsymbol{\Sigma}_{\mathrm{g}}$ through distributed average consensus [105].

6.2.2 Distributed prediction

Once an estimate of $\boldsymbol{\theta}$ is obtained, the next goal is to determine $p\big(P_{\mathrm{RX}}(\mathbf{x}_*) \mid \mathcal{D}, \boldsymbol{\theta}, \mathbf{x}_*\big)$ at an unvisited test location $\mathbf{x}_*$. Based on GP approximation (6.1), the posterior pdf at $\mathbf{x}_*$ is given by:

$$p\big(P_{\mathrm{RX}}(\mathbf{x}_*) \mid \mathcal{D}, \boldsymbol{\theta}, \mathbf{x}_*\big) \propto p\big(P_{\mathrm{RX}}(\mathbf{x}_*)\boldsymbol{\theta}, \mathbf{x}_*\big) \times \prod_{i=1}^{N} p\big(\mathcal{D}_i \mid P_{\mathrm{RX}}(\mathbf{x}_*), \boldsymbol{\theta}, \mathbf{x}_*\big). \tag{6.16}$$

Applying Bayes' rule:

$$p\big(P_{\mathrm{RX}}(\mathbf{x}_*) \mid \mathcal{D}, \boldsymbol{\theta}, \mathbf{x}_*\big) \propto \frac{\prod_{i=1}^{N} p\big(P_{\mathrm{RX}}(\mathbf{x}_*) \mid \mathcal{D}_i, \boldsymbol{\theta}, \mathbf{x}_*\big)}{\big(p(P_{\mathrm{RX}}(\mathbf{x}_*)|\boldsymbol{\theta}, \mathbf{x}_*)\big)^{N-1}}, \tag{6.17}$$

where $p(P_{\mathrm{RX}}(\mathbf{x}_*)|\boldsymbol{\theta}, \mathbf{x}_*)$ is the a priori predictive pdf at test location $\mathbf{x}_*$ with mean $\mu(\mathbf{x}_*)$ and variance k_{**}. Since all distributions involved in (6.17) are Gaussian, and the product of Gaussians is also Gaussian, (6.17) can be written as:

$$p\big(P_{\mathrm{RX}}(\mathbf{x}_*) \mid \mathcal{D}, \boldsymbol{\theta}, \mathbf{x}_*\big) \propto \frac{\mathcal{N}_{P_{\mathrm{RX}}(\mathbf{x}_*)}\left(\mu_{\mathrm{c}}(\mathbf{x}_*), \sigma_{\mathrm{c}}^2(\mathbf{x}_*)\right)}{\mathcal{N}_{P_{\mathrm{RX}}(\mathbf{x}_*)}\left(\mu_{\mathrm{p}}(\mathbf{x}_*), \sigma_{\mathrm{p}}^2\right)}, \tag{6.18}$$

with

$$\mu_{\mathrm{c}}(\mathbf{x}_*) = \frac{\sum_{i=1}^{N} \sigma_{\mathrm{RX},i}^{-2}(\mathbf{x}_*)\ \bar{P}_{\mathrm{RX},i}(\mathbf{x}_*)}{\sum_{i=1}^{N} \sigma_{\mathrm{RX},i}^{-2}(\mathbf{x}_*)}, \tag{6.19}$$

$$\sigma_{\mathrm{c}}^2(\mathbf{x}_*) = \left[\sum_{i=1}^{N} \sigma_{\mathrm{RX},i}^{-2}(\mathbf{x}_*)\right]^{-1}, \tag{6.20}$$

and

$$\mu_{\mathrm{p}}(\mathbf{x}_*) = \mu(\mathbf{x}_*), \tag{6.21}$$

$$\sigma_{\mathrm{p}}^2 = \frac{k_{**}}{N-1}, \tag{6.22}$$

where $\bar{P}_{\mathrm{RX},i}(\mathbf{x}_*)$ and $\sigma_{\mathrm{RX},i}^{-2}(\mathbf{x}_*)$ are the predictions of agent i, based on $\mathcal{D}_i$, computed similarly to (4.27)–(4.28), 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 $p\big(P_{\mathrm{RX}}(\mathbf{x}_*) \mid \mathcal{D}, \boldsymbol{\theta}, \mathbf{x}_*\big)$ is proportional to a Gaussian distribution with mean:

$$\mu_{\mathrm{post}}(\mathbf{x}_*) = \sigma_{\mathrm{post}}^2(\mathbf{x}_*)\left(\mu_{\mathrm{c}}(\mathbf{x}_*)/\sigma_{\mathrm{c}}^2(\mathbf{x}_*) - \mu_{\mathrm{p}}(\mathbf{x}_*)/\sigma_{\mathrm{p}}^2\right), \tag{6.23}$$

and variance:

$$\sigma_{\mathrm{post}}^2(\mathbf{x}_*) = \frac{1}{1/\sigma_{\mathrm{c}}^2(\mathbf{x}_*) - 1/\sigma_{\mathrm{p}}^2(\mathbf{x}_*)}. \tag{6.24}$$

From the above derivations, it is clear that the agents should agree on $\mu_{\mathrm{c}}(\mathbf{x}_*)$ and $\sigma_{\mathrm{c}}^2(\mathbf{x}_*)$ in order to determine $\mu_{\mathrm{post}}(\mathbf{x}_*)$ and $\sigma_{\mathrm{post}}^2(\mathbf{x}_*)$. Both $\mu_{\mathrm{c}}(\mathbf{x}_*)$ and $\sigma_{\mathrm{c}}^2(\mathbf{x}_*)$ 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. Initialize variables:

$$\xi_i^{(0)} = N\sigma_{\mathrm{RX},i}^{-2}(\mathbf{x}_*)\ \bar{P}_{\mathrm{RX},i}(\mathbf{x}_*), \tag{6.25}$$

and

$$\lambda_i^{(0)} = N\sigma_{\mathrm{RX},i}^{-2}(\mathbf{x}_*). \tag{6.26}$$

Letting $\boldsymbol{\zeta}_i^{(l)} = [\xi_i^{(l)}, \lambda_i^{(l)}]^{\mathrm{T}}$, agents then apply a consensus update rule [105]. For instance:

$$\boldsymbol{\zeta}_i^{(l+1)} = \boldsymbol{\zeta}_i^{(l)} + \eta \sum_{j \in \mathcal{N}_i} \big(\boldsymbol{\zeta}_j^{(l)} - \boldsymbol{\zeta}_i^{(l)}\big), \tag{6.27}$$

where η is a small constant[2], and $\mathcal{N}_i$ is the neighborhood set of agent i, as determined by the adjacency matrix $\mathbf{A}$. As $l \to +\infty$, it can be verified that:

$$\xi_i^{(l)} \to \sum_{i=1}^{N} \sigma_{\mathrm{RX},i}^{-2}(\mathbf{x}_*)\ \bar{P}_{\mathrm{RX},i}(\mathbf{x}_*), \tag{6.28}$$

and

$$\lambda_i^{(l)} \to \sum_{i=1}^{N} \sigma_{\mathrm{RX},i}^{-2}(\mathbf{x}_*) = 1/\sigma_{\mathrm{c}}^2(\mathbf{x}_*), \tag{6.29}$$

from which $\mu_{\mathrm{c}}(\mathbf{x}_*)$ and $\sigma_{\mathrm{c}}^2(\mathbf{x}_*)$ are easily retrieved. By substituting $\mu_{\mathrm{c}}(\mathbf{x}_*)$ and $\sigma_{\mathrm{c}}^2(\mathbf{x}_*)$ in (6.23) and (6.24), the global predictive pdf (6.18) has mean $\mu_{\mathrm{post}}(\mathbf{x}_*)$ and variance $\sigma_{\mathrm{post}}^2(\mathbf{x}_*)$, 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.

[2] A sufficient condition to reach consensus is $\eta < 1/\Delta$, where Δ is the maximum node degree of graph $\mathcal{G}$ [105].

6.2.3 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 (4.26) and (4.27)) 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.

6.3 Numerical results

In this section, numerical results to illustrate the performance of predicting the received power P_{RX} of centralized GPR and distributed GPR are presented.

6.3.1 Simulation setup

A MAS scenario in a square area $\mathcal{A}$ of $200\,\mathrm{m} \times 200\,\mathrm{m}$ is considered. The transmitting agent (TX) is placed at location $\mathbf{x}_t = [0, 0]^{\mathrm{T}}$ and transmits with a power of 33 dBm. The received power is generated according to a Gaussian field (4.21), with $\alpha = 1$ in (4.23), path-loss exponent 3.5, 8 dB of shadowing standard deviation, a shadowing correlation distance of 50 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 $p = 0.1$. The prediction is performed in a fine grid of $N_p = 2061$ locations (corresponding to a resolution of 4 meters). For training measurements, the total size of database is fixed to $|\mathcal{D}| = 1000$, and let $|\mathcal{D}_i| = |\mathcal{D}|/N$, whereby N is varied. Hence, $N = 1$ corresponds to a centralized GPR, while $N = 1000$ corresponds to one measurement per agent.

Two distinct types of agent databases are considered:

- **Asymmetric databases:** each agents contains different number of measurements. Agents collect measurements randomly in $\mathcal{A}$.
- **Equal databases:** all agent contains the same number of measurement. In this, two ways of assignment are considered: (1) *Random assignment:* measurements are randomly distributed in $\mathcal{A}$ and each measurement is assigned randomly to one of the N agents. (2) *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.

6.3.2 Learning

To test the distributed learning, an asymmetric database is considered with the following simulation setup: $|\mathcal{D}_i| = [1, 19, 80, 200, 700]$ such that $|\mathcal{D}| = 1000$, number of samples $L = 1000$, number of iterations $I = 20$ and $\boldsymbol{\mu}_p \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}_p)$, where $\boldsymbol{\mu} = [3, 36, 10]$, $\boldsymbol{\mu_p} = [3.5, 42, 13]$ and $\boldsymbol{\Sigma}_p = diag([4, 36, 25])$. Note, for learning, simulation parameters are from Table 4.1.

Figure 6.1 show agent's posterior distribution $p(\boldsymbol{\theta}|\mathcal{D}_i)$, global posterior distribution $p(\boldsymbol{\theta}|\mathcal{D})$ and prior distribution for channel parameters η, σ_Ψ^2 and d_c. The following conclusions can be drawn from the plots: (1) agents with bigger $\mathcal{D}_i$ show narrow distributions while the one with smaller $\mathcal{D}_i$ presents wider distributions. This is because the agents with more entries in their $\mathcal{D}_i$ have lower variance (uncertainty) and are more certain about their estimated values. This observation can be seen in Figure 6.2, where η, σ_Ψ^2 and d_c are plot for agents with different number of measurements. Agents with small $\mathcal{D}_i$ have higher estimation variance when compared with agents with bigger $\mathcal{D}_i$. The clear advantage with the proposed method is that, after reaching consensus, each agent achieves a global posterior distribution. Even though it had smaller $\mathcal{D}_i$ and performed poor learning, with distributed learning it obtains global parameter values. It can improve its prediction by employing global $\boldsymbol{\theta}$ knowledge.

6.3.3 Prediction

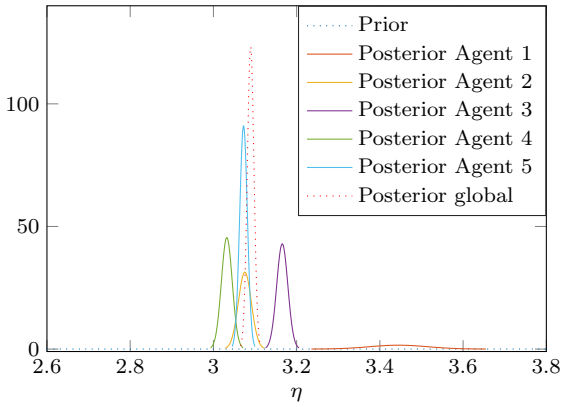
The prediction is performed in a fine grid of $N_p = 2061$ locations (corresponding to a resolution of 4 meters). To assess the prediction quality using the centralized GPR and the distributed GPR, the MSE between the ground truth and the predicted mean is considered. It is defined as:

$$\text{MSE} = \frac{1}{N_p} \sum_{i=1}^{N_p} \left(\mu_{\text{post}}^{(l)}(\mathbf{x}_{*,i}) - P_{\text{RX}}(\mathbf{x}_{*,i}) \right)^2, \tag{6.30}$$

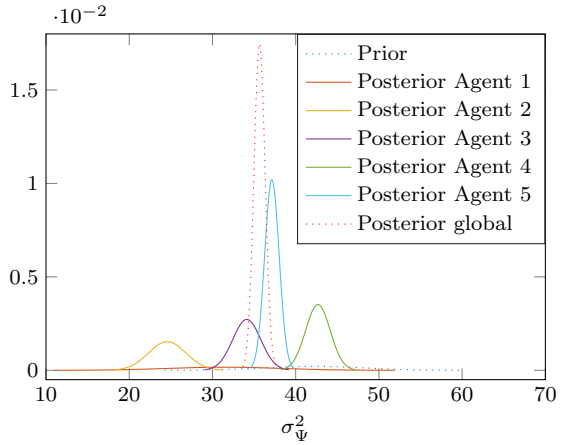
where $\mu_{\text{post}}^{(l)}(\cdot)$ is defined in (6.23) evaluated at consensus iteration l.

Qualitative comparison

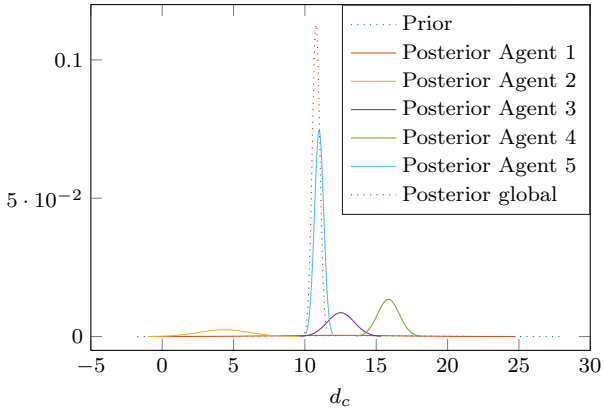
To illustrate the reconstruction quality, the predicted mean of the received power $\bar{P}_{\text{RX}}$ for $N = 1$ (i.e., fully centralized) and $N = 1000$ (i.e., fully distributed) is presented in Figure 6.3. If the mean predicted received power of fully distributed is compared with the fully centralized, it can be noticed that the quality of Figure 6.3(b) is comparatively lower than Figure 6.3(a). 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.



(a) Path loss exponent η

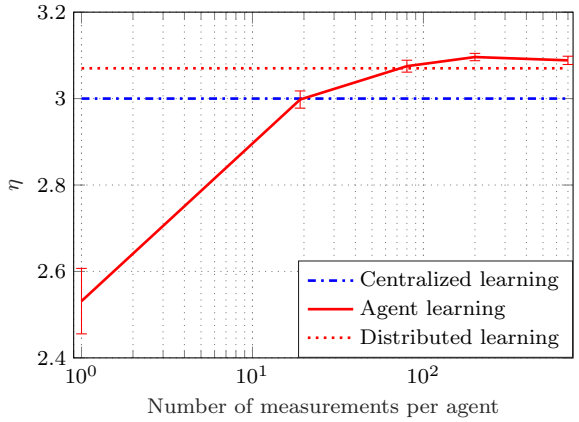


(b) Shadowing variance σ^2_Ψ

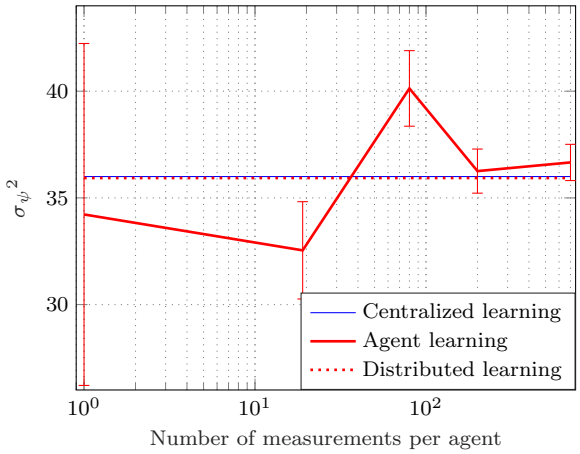


(c) Correlation distance d_c

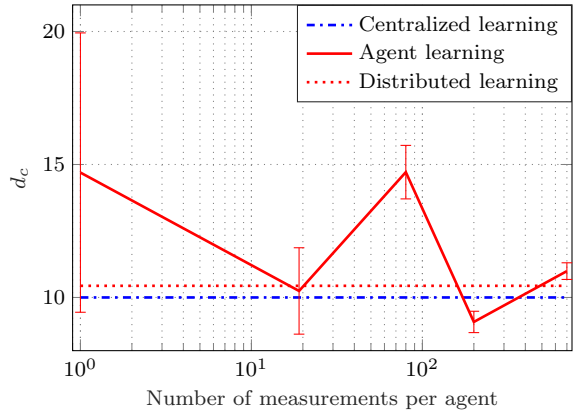
Figure 6.1: Posterior distributions of estimated parameters $\boldsymbol{\theta} = [\eta, \sigma^2_\Psi, d_c]^{\mathrm{T}}$.



(a) Path-loss exponent η



(b) Shadowing variance σ_Ψ^2



(c) Correlation distance d_c

Figure 6.2: Estimated parameters $\boldsymbol{\theta} = [\eta, \sigma_\Psi^2, d_c]^{\mathrm{T}}$ versus size of the database per agent. Error bar indicates one standard deviation.

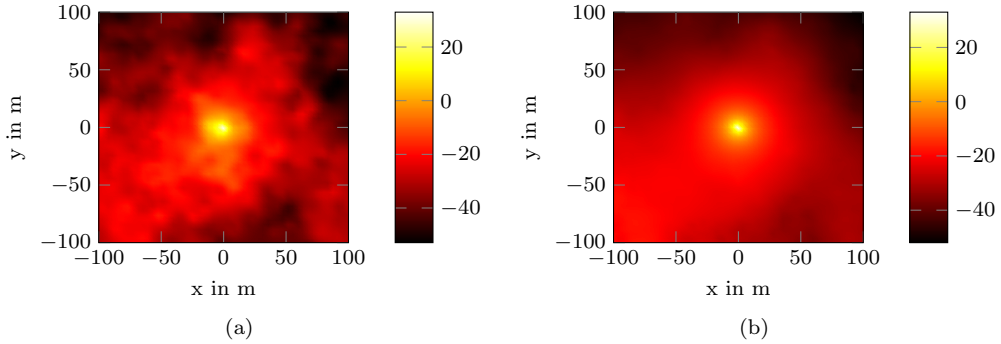


Figure 6.3: Mean of GPR prediction (in dBm) for $N = 1$ (Centralized) in (a) and $N = 1000$ (Distributed) in (b).

Convergence behavior

Now the impact of the number of consensus iterations l on the MSE for the case of $N = 20$ and $|\mathcal{D}_i| = 50$ is analyzed. In Figure 6.4, the mean MSE over the N agents together with the minimum and maximum MSE at each iteration l is plot for random and clustered assignments. It can be observed 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.

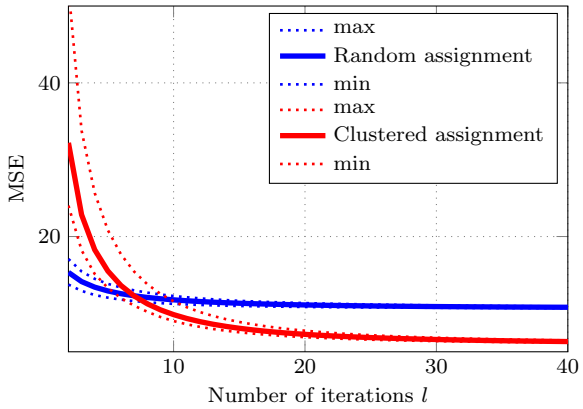


Figure 6.4: Mean MSE for networks with different connectivity versus number of consensus iterations l for 20 agents with each having 50 measurements. The dashed lines indicate the minimum and maximum MSE, respectively.

This leads to a larger initial disagreement for the clustered case, and thus a slower convergence for the clustered assignment.

Quantitative comparison

The prediction quality with respect to the number of measurements per agent $|\mathcal{D}_i|$ is assessed. In Figure 6.5, 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 transition from the centralized prediction to the completely distributed prediction can be observed. 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 a fully distributed approach, both assignments lead to a 400 % increase in MSE. This highlights the fact that with a distributed approach comes a performance penalty. Finally, for a simplified path-loss only model, a mean MSE of 58.34 was obtained, which is significantly higher than the MSE of distributed GPR with any database size using model (3.2), highlighting the fact that considering shadowing correlation is important when predicting wireless channels.

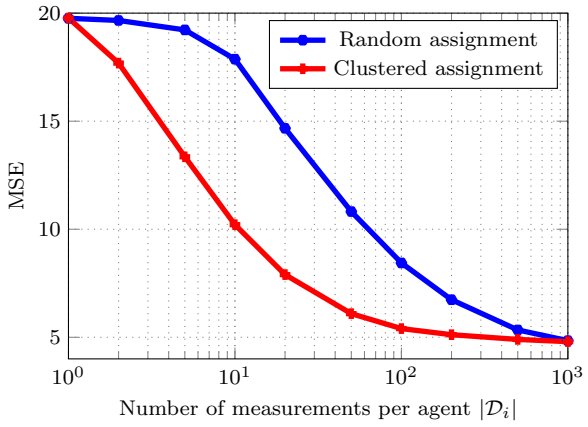


Figure 6.5: MSE versus average size of the database per agent, for a total database size of $|\mathcal{D}| = 1000$.

Comparison with kriging

In order to perform a fair comparison between distributed GPR and kriging, the proposed prediction method in Section 6.2.2 is adapted to be suitable for RK. $\bar{P}_{\mathrm{RX},i}(\mathbf{x}_*)$ and $\sigma^{-2}_{\mathrm{RX},i}(\mathbf{x}_*)$ in (6.19) and (6.20) are the prediction of agents, based on $\mathcal{D}_i$. Earlier, $\bar{P}_{\mathrm{RX},i}(\mathbf{x}_*)$ and $\sigma^{-2}_{\mathrm{RX},i}(\mathbf{x}_*)$ were computed using GPR. However, if each agent performs RK prediction by computing its predictions using (4.19) and (4.20), and substituting in (6.19) and (6.20) provides a distributed kriging version of the proposed method.

Figure 6.6 shows the MSE versus the size of the database per agent for distributed kriging and distributed GPR. An assymmetric $\mathcal{D}_i$ is considered, where $|\mathcal{D}_i| = [1, 19, 80, 200, 700]$ such that $|\mathcal{D}| = 1000$. Firstly, it can be seen that kriging and GP provide the same prediction with larger $|\mathcal{D}_i|$. In Figure 6.6, for $|\mathcal{D}_i| < 20$, GPR performs better than kriging. Note that centralized GPR/ kriging refers to the agent with $|\mathcal{D}_i| = 1000$.

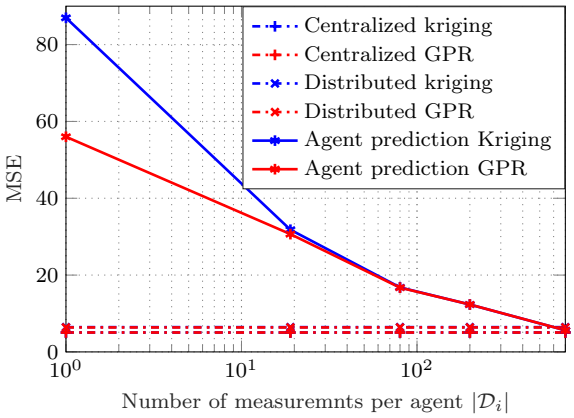


Figure 6.6: MSE versus size of the database per agent

Prediction under location uncertainty

In practical MAS applications, the agents' may not have knowledge of their true locations. Consider a 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 $1/\lambda$. Uncertain training with $\lambda = 6\,\mathrm{m}$ for a fraction of the agents and $\lambda = 0\,\mathrm{m}$ for the remaining agents are considered. Test locations have no location uncertainty. For $N = 5$, i.e., $|\mathcal{D}_i| = 200$ of Figure 6.5, the impact of localization errors on prediction quality is presented in Table 6.1. 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 6.1: 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

Chapter 7

Conclusions and Future Work

7.1 Conclusions

In this Thesis REMs were envisioned as a powerful tool to provide contextual information to support the operation of both traditional and disruptive technologies of 5G. REMs were constructed by leveraging link context and location information in particular to enhance the contextual awareness in the spatial domain. For REM reconstruction, Kriging from geo-statistics and GPR from machine learning were chosen for spatial interpolation. In general, Kriging and GPR are faced with issues of centralized implementation and computational complexity. To overcome these drawbacks, this Thesis presented distributed methods to the problem of REM reconstruction based on two unique approaches, offering solutions depending on the application of interest. First, the DICA was proposed to minimize the complexity by employing the closest observations from the local neighborhood. Second, the distributed GPR method was proposed to reduce the complexity by distributing global computations among independent computation units. In this chapter, conclusions are drawn and future work is presented.

7.1.1 Distributed incremental clustering algorithm

In chapter 5, a novel DICA for spatial prediction in wireless sensor networks based on the RK interpolation method was presented. The algorithm minimizes the total number of sensor measurements required for REM 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, RK leads to a superior performance than the plain regression and standard OK.

7.1.2 Distributed Gaussian process regression

In chapter 6, a distributed GPR framework for channel learning and prediction in MAS is presented. The proposed method reduces computation and memory requirements compared to a centralized GPR, thus allowing to increase the size of the training dataset. For learning/ 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 parameter/ 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 MSE 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.

7.2 Future work

In the following, some future directions of work are listed.

- Testing with empirical measurements: In this Thesis, measurements were generated from simulations using path-loss and shadowing models that characterize the wireless channel. However, it is interesting to employ empirical measurements and consider practical implementation issues.
- Modeling temporal variations: REM reconstruction is a multidimensional problem. In this Thesis, pure spatial models were considered. However, modeling correlation of shadowing in the temporal dimension provides more flexibility and paves way for more applications. For example, resource allocation in anticipatory networks.
- Prediction under location uncertainty: Algorithms were designed in this Thesis by assuming perfect knowledge of location information. However, accounting location uncertainty in learning and prediction will improve the practicality of the proposed method.
- Quantifying loss of information: With the aim of offering a good trade-off between complexity and prediction complexity, distributed methods proposed in this Thesis were based on using the least number of measurements in DICA and independence assumption between databases in distributed GPR. However, quantifying loss of information will give an insight about the performance of the proposed methods.

Appendix: A

DICA-OK

In the following, preliminary results obtained by the DICA-OK are presented for completeness. The main objective of this study was to prove the reconstruction capacity of the DICA-OK solution, and to explore the adaptive clustering formation algorithm. Simulations were performed considering the scenario depicted in Figure 7.1, where $n = 100$ sensor nodes are randomly deployed in a square area of 570 m $\times$ 570 m. In order to test the algorithm performance, a field was created based on channel model (3.2) with path-loss exponent γ=6, reference distance d_0=10 m, number of transmitters N_{TX}=3, sampling spacing ls=5, loss at 10m d_0=20 dB, correlation length d_c=60 m and location variablity σ^2_{Ψ}=10 dB.

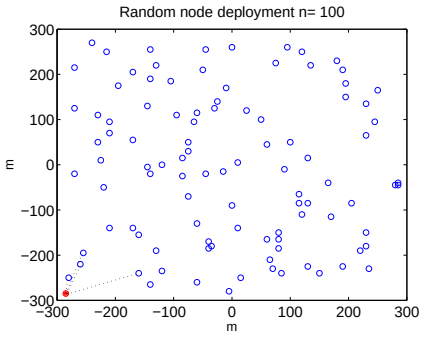
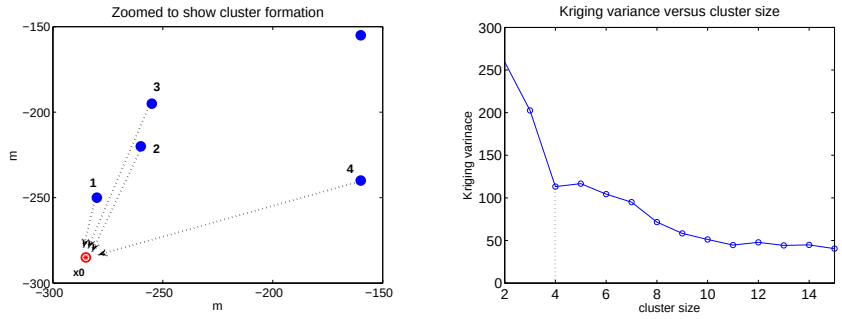


Figure 7.1: Random node placement of size N=100

In the DICA-OK algorithm, initial clusters close to the location target are selected to begin the estimation. Nodes in the cluster distributively compute the semivariogram and solve the Kriging equations by using only the readings from the nodes. As a result, local semivariogram and weights are obtained for each node in the cluster. The advantage of this approach is that nodes with high spatial correlation to the target location are given more weights while the non-cluster nodes are neglected. Because of the adaptive nature

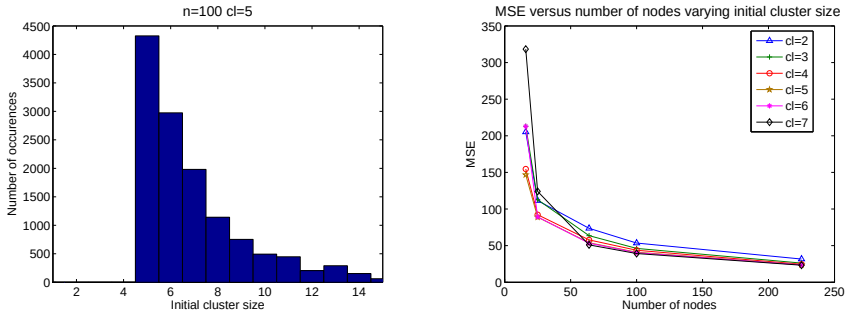


(a) (b)

Figure 7.2: (a) Figure 7.1 zoomed to show cluster formation (b) Kriging variance versus cluster size for location $\mathbf{x_0}$=[-285,-285]

of the algorithm, the initial cluster size can be increased by adding new nodes. However, increasing the size of the cluster does not guarantee a better estimation quality. Hence, our clustering algorithm employs the Kriging variance to choose an optimal cluster size.

The cluster formation procedure was demonstrated with an example. The size of the cluster at each unknown location $\mathbf{x}_0$ changes adaptively depending on the Kriging variance. In Figure 7.2(a), an unknown location $\mathbf{x}_0 = [-285, -285]$ is considered, where an initial set $t = 3$ composed by $\{1, 2, 3\}$ performs Kriging. Figure 7.2(b) shows the Kriging variance as new nodes are added in the cluster. In this case, the addition of a new node $\{4\}$ improves the estimation and the cluster will be formed by the set of 4 nodes. From Figure 7.2(b), one can see that including additional nodes in the cluster may further decrease the Kriging variance. However, the computational complexity increases significantly, without having a clear impact on the estimation quality.



(a) (b)

Figure 7.3: (a) Histogram (b) MSE as a function of initial cluster size

An important factor that must be taken into account is the initial cluster size. The only parameter influencing the formation of initial clusters is the parameter R and battery restrictions or communication constraints which were out of the scope of this analysis. Therefore, we present simulations obtained with several initial cluster sizes. The histogram plot 7.3(a) illustrates the number of occurrences of initial cluster sizes for reconstructing the entire area. By observing the MSE as a function of the initial cluster size in Figure 7.3(b) and comparing with the histogram plot, a realistic initial cluster size (cl) of 5 nodes was selected for reconstructing the REM.

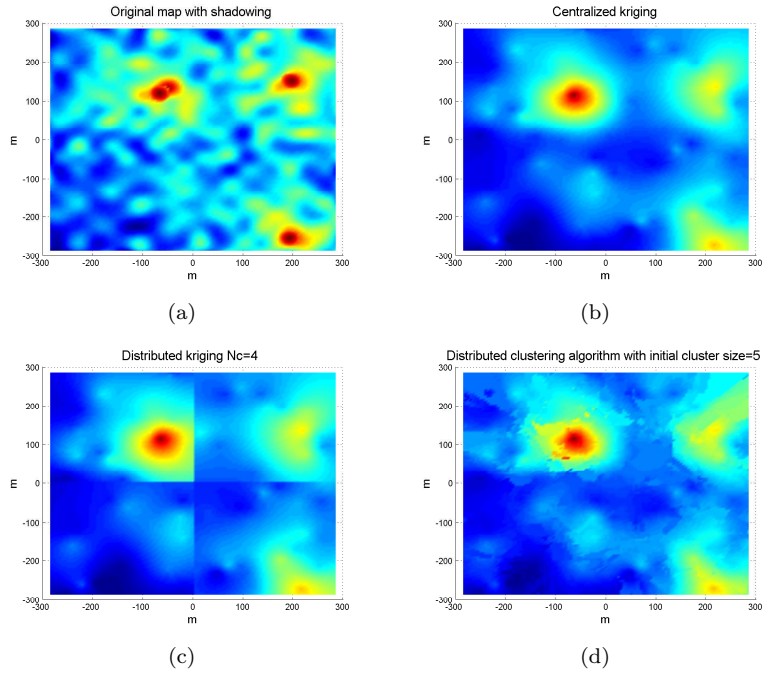


Figure 7.4: Interpolated maps with n=100 (a) Original map (b) Centralized Kriging (c) Distributed Kriging with Nc=4 (d) DICA-OK with cl=5

Figure 7.4 shows the original REM with 3 TXs, the reconstructed REM obtained using a centralized Kriging solution, the result of applying DICA-OK after dividing the area in 4 independent subregions, and the reconstructed REM when DICA-OK is applied with an initial cluster size of 5 nodes, following the results from Figure 7.3.

DICA-OK was extended by beginning the clustering procedure with 3 nodes in [26] by employing the model (4.7) presented in section 4.1.2. The average size of clusters required to build REMs for various WSNs sizes were computed. Results in Table 7.1 showed that DICA-OK required an average of 5 nodes, which is prominently lower than the centralized case.

Table 7.1: 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

References

[1] "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016–2021," Cisco Systems, Tech. Rep. C11-738429-00, Feb 2017.

[2] "5G Vision: The Next Generation of Communication Networks and Services," 5G Infrastructure Association, Tech. Rep., Feb 2015.

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

[4] N. Al-Falahy and O. Y. Alani, "Technologies for 5G Networks: Challenges and Opportunities," *IT Professional*, vol. 19, no. 1, pp. 12–20, Jan 2015.

[5] M. R. Palattella, M. Dohler, A. Grieco, G. Rizzo, J. Torsner, T. Engel, and L. Ladid, "Internet of Things in the 5G Era: Enablers, Architecture, and Business Models," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 3, pp. 510–527, Mar 2015.

[6] S. Nunna, A. Kousaridas, M. Ibrahim, M. Dillinger, C. Thuemmler, H. Feussner, and A. Schneider, "Enabling Real-Time Context-Aware Collaboration through 5G and Mobile Edge Computing," in *2015 12th International Conference on Information Technology - New Generations*, Apr 2015, pp. 601–605.

[7] "NGMN 5G white paper," NGMN Alliance, Tech. Rep., Feb 2015.

[8] 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.

[9] M. Koivisto, A. Hakkarainen, M. Costa, P. Kela, K. Leppänen, and M. Valkama, "High-Efficiency Device Positioning and Location-Aware Communications in Dense 5G Networks," *CoRR*, vol. abs/1608.03775, Sept 2016.

[10] J. Medbo, I. Siomina, A. Kangas, and J. Furuskog, "Propagation Channel Impact on LTE Positioning Accuracy: A Study based on Real Measurements of Observed Time difference of arrival," in *2009 IEEE 20th International Symposium on Personal, Indoor and Mobile Radio Communications*, Sept 2009, pp. 2213–2217.

[11] D. Dardari, P. Closas, and P. M. Djurić, "Indoor Tracking: Theory, Methods, and Technologies," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 4, pp. 1263–1278, Apr 2015.

[12] H. Liu, Y. Gan, J. Yang, S. Sidhom, Y. Wang, Y. Chen, and F. Ye, "Push the Limit of WiFi Based Localization for Smartphones," in *Proceedings of the 18th Annual International Conference on Mobile Computing and Networking*, ser. Mobicom '12. New York, NY, USA: ACM, 2012, pp. 305–316.

[13] A. Dammann, G. Agapiou, J. Bastos, L. Brunelk, M. Garcia, J. Guillet, Y. Ma, J. Ma, J. J. Nielsen, L. Ping, R. Raulefs, J. Rodriguez, D. Slock, D. Yang, and N. Yi, "WHERE2 Location Aided Communications," in *European Wireless 2013; 19th European Wireless Conference*, Apr 2013, pp. 1–8.

[14] B. Bangerter, S. Talwar, R. Arefi, and K. Stewart, "Networks and Devices for the 5G Era," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 90–96, Feb 2014.

[15] P. Makris, D. N. Skoutas, and C. Skianis, "A Survey on Context-Aware Mobile and Wireless Networking: On Networking and Computing Environments' Integration," *IEEE Communications Surveys Tutorials*, vol. 15, no. 1, pp. 362–386, Apr 2012.

[16] V. Pejovic and M. Musolesi, "Anticipatory Mobile Computing: A Survey of the State of the Art and Research Challenges," *ACM Comput. Surv.*, vol. 47, no. 3, 2012.

[17] 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.

[18] "Universal Mobile Telecommunications System (UMTS); LTE; Universal Terrestrial Radio Access (UTRA) and Evolved Universal Terrestrial Radio Access (E-UTRA); Radio measurement collection for Minimization of Drive Tests (MDT); Overall description; Stage 2," *3GPP TS 37.320 version 10.1.0 Release 10*, Apr 2011.

[19] J. Johansson, W. A. Hapsari, S. Kelley, and G. Bodog, "Minimization of Drive Tests in 3GPP Release 11," *IEEE Communications Magazine*, vol. 50, no. 11, pp. 36–43, Nov 2012.

[20] 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

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

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

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

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

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

[26] V. P. Chowdappa and 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.

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

[28] 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.

[29] V. P. Chowdappa, C. Botella, and B. Beferull-Lozano, "Distributed Clustering Algorithm for Spatial Field Reconstruction in Wireless Sensor Networks," *Sensors*, June 2017, to be submitted.

[30] V. P. Chowdappa, M. Fröhle, H. Wymeersch, and C. Botella, "Distributed Gaussian Processes and Kriging for Channel Learning and Prediction in Multi-Agent Systems," *IEEE Transactions on Wireless Communications*, Aug 2017, to be submitted.

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

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

[33] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174–188, Feb 2002.

[34] J. Mitola, "Cognitive Radio: An Integrated Agent Architecture for Software Defined Radio," Ph.D. dissertation, KTH, 2000.

[35] 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.

[36] 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.

[37] 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.

[38] T. Cai, J. van de Beek, B. Sayrac, S. Grimoud, J. Nasreddine, J. Riihijärvi, and P. Mähönen, "Design of Layered Radio Environment Maps for RAN Optimization in Heterogeneous LTE Systems," in *2011 IEEE 22nd International Symposium on Personal, Indoor and Mobile Radio Communications*, Sept 2011, pp. 172–176.

[39] J. V. D. Beek, T. Cai, S. Grimoud, B. Sayrac, P. Mahonen, J. Nasreddine, and J. Riihijarvi, "How a Layered REM Architecture Brings Cognition to Today's Mobile Networks," *IEEE Wireless Communications*, vol. 19, no. 4, pp. 17–24, Aug 2012.

[40] L. Iacobelli, P. Fouillot, and C. J. L. Martret, "Radio Environment Map based Architecture and Protocols for Mobile Ad hoc Networks," in *2012 The 11th Annual Mediterranean Ad Hoc Networking Workshop (Med-Hoc-Net)*, June 2012, pp. 32–38.

[41] W. R. Tobler, "A Computer Movie Simulating Urban Growth in the Detroit Region," *Economic Geography*, vol. 46, no. sup1, pp. 234–240, 1970.

[42] Jin Li and Andrew D. Heap, "A Review of Comparative Studies of Spatial Interpolation Methods in Environmental Sciences: Performance and Impact factors ," *Ecological Informatics*, vol. 6, no. 3–4, pp. 228 – 241, 2011.

[43] 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, pp. 422–433.

[44] H. B. Yilmaz, T. Tugcu, F. Alagöz, 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.

[45] 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.

[46] 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.

[47] 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.

[48] S. Sun, K. Adachi, P. H. Tan, Y. Zhou, J. Joung, and C. K. Ho, "Heterogeneous Network: An Evolutionary Path to 5G," in *2015 21st Asia-Pacific Conference on Communications (APCC)*, Oct 2015, pp. 174–178.

[49] G. P. Fettweis, "The Tactile Internet: Applications and Challenges," *IEEE Vehicular Technology Magazine*, vol. 9, no. 1, pp. 64–70, Mar 2014.

[50] 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.

[51] 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.

[52] 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.

[53] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five Disruptive Technology Directions for 5G," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 74–80, Feb 2014.

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

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

[56] 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.

[57] W. Krenik and A. Batra, "Cognitive Radio Techniques for Wide Area Networks," in *Proceedings. 42nd Design Automation Conference, 2005.*, June 2005, pp. 409–412.

[58] Q. Zhao and B. M. Sadler, "A Survey of Dynamic Spectrum Access," *IEEE Signal Processing Magazine*, vol. 24, no. 3, pp. 79–89, May 2007.

[59] 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.

[60] 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.

[61] S. Üreten, A. Yongaçoğlu, and E. Petriu, "Interference Map Generation based on Delaunay Triangulation in Cognitive Radio Networks," in *2012 IEEE 13th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, June 2012, pp. 134–138.

[62] 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.

[63] 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.

[64] 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, Aprl 2005, pp. 248–253 Vol. 2.

[65] D. Denkovski and V. Atanasovski and 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.

[66] 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.

[67] 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.

[68] J. Riihijarvi, P. Mahonen, M. Wellens, and M. Gordziel, "Characterization and Modelling of Spectrum for Dynamic Spectrum Access with Spatial Statistics and Random fields," in *2008 IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications*, Sept 2008, pp. 1–6.

[69] M. Wellens, J. Riihijarvi, M. Gordziel, and P. Mahonen, "Spatial Statistics of Spectrum Usage: From Measurements to Spectrum Models," in *2009 IEEE International Conference on Communications*, June 2009, pp. 1–6.

[70] 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.

[71] "Spatial Statistics and Models of Spectrum Use," *Computer Communications*, vol. 32, no. 18, pp. 1998 – 2011, 2009, cognitive Radio and Dynamic Spectrum Sharing Systems.

[72] 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.

[73] 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, pp. 724–729.

[74] J. Riihijarvi, P. Mahonen, M. Petrova, and V. Kolar, "Enhancing Cognitive Radios with Spatial Statistics: From Radio Environment Maps to Topology Engine," in *2009 4th International Conference on Cognitive Radio Oriented Wireless Networks and Communications*, June 2009, pp. 1–6.

[75] J. Riihijärvi, P. Mähönen, and S. Sajjad, "Influence of Transmitter Configurations on Spatial Statistics of Sadio Environment Maps," in *2009 IEEE 20th International Symposium on Personal, Indoor and Mobile Radio Communications*, Sept 2009, pp. 853–857.

[76] C. Jardak, J. Riihijärvi, F. Oldewurtel, and P. Mähönen, "Parallel Processing of Data from Very Large-scale Wireless Sensor Networks," in *Proceedings of the 19th ACM International Symposium on High Performance Distributed Computing*, ser. HPDC '10. New York, NY, USA: ACM, 2010, pp. 787–794.

[77] 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.

[78] S. Liu, E. Masazade, M. Fardad, and P. K. Varshney, "Sparsity-Aware Field Estimation via Ordinary Kriging," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2012, pp. 3948–3952.

[79] 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.

[80] 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.

[81] R. Kong, W. Zhang, and Y. Guo, "Distributed Estimation and Tracking for Radio Environment Mapping," in *2014 American Control Conference*, June 2014, pp. 464–470.

[82] 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.

[83] 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.

[84] 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.

[85] K. Tiwari, V. Honoré, 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.

[86] N. Jalden, P. Zetterberg, B. Ottersten, A. Hong, and R. Thoma, "Correlation Properties of Large Scale Fading Based on Indoor Measurements," in *Wireless Communications and Networking Conference, 2007.WCNC 2007. IEEE*, Mar 2007, pp. 1894–1899.

[87] 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.

[88] 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.

[89] 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.

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

[91] 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.

[92] 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.

[93] 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. 1–2, pp. 75 – 93, 2004.

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

[95] I.O.A. Odeha and A.B. McBratney and D.J. Chittleborough, "Spatial prediction of soil properties from landform attributes derived from a digital elevation model," *Geoderma*, vol. 63, no. 3, pp. 197 – 214, 1994.

[96] T.F.A Bishop and A.B McBratney, "A Comparison of Prediction Methods for the Creation of Field-Extent Soil Property Maps," *Geoderma* , vol. 103, no. 1–2, pp. 149 – 160, 2001, estimating uncertainty in soil models.

[97] P. Goovaerts, "Using Elevation to Aid the Geostatistical Mapping of Rainfall Erosivity ," *CATENA*, vol. 34, no. 3–4, pp. 227 – 242, 1999.

[98] 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.

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

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

[101] F. S. Cattivelli and 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.

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

[103] 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, pp. 724–729.

[104] 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.

[105] 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.