

Trabajo Fin de Máster

Máster en Ingeniería Industrial

Estudio de la distribución espacial de la temperatura del interior de una sala utilizando una red de sensores inalámbricos

Autor: Daniel Paredes García

Tutora: Amparo Núñez Reyes

Dep. Ingeniería de Sistemas y Automática
Escuela Técnica Superior de Ingeniería
Universidad de Sevilla

Sevilla, 2021



Trabajo Fin de Máster
Máster en Ingeniería Industrial

Estudio de la distribución espacial de la temperatura del interior de una sala utilizando una red de sensores inalámbricos

Autor:

Daniel Paredes García

Tutora:

Amparo Núñez Reyes

Profesora titular

Dep. Ingeniería de Sistemas y Automática

Escuela Técnica Superior de Ingeniería

Universidad de Sevilla

Sevilla, 2021

Trabajo Fin de Máster: Estudio de la distribución espacial de la temperatura del interior de una sala utilizando una red de sensores inalámbricos

Autor: Daniel Paredes García

Tutora: Amparo Núñez Reyes

El tribunal nombrado para juzgar el Proyecto arriba indicado, compuesto por los siguientes miembros:

Presidente:

Vocales:

Secretario:

Acuerdan otorgarle la calificación de:

Sevilla, 2021

El Secretario del Tribunal

Agradecimientos

Llegados a este punto donde culmina mi etapa universitaria en lo referente a la Ingeniería Industrial, me gustaría dedicar unas palabras de agradecimientos a aquellas personas que han formado parte de mi vida durante este camino.

A mi familia, en especial a mis padres, que siempre han estado ahí cada vez que los necesitaba. Nunca podré devolverles todo lo que han hecho por mí y todo lo que siguen haciendo.

A mis amigos, a aquellos que me llevo de esta bonita etapa y a los de toda la vida, por hacerme más amenos estos años y por sacarme siempre una sonrisa.

A *Dable Red S.L.* y las personas que lo conforman, por ofrecerme la oportunidad de entrar en el mundo laboral, otorgándome toda la flexibilidad necesaria para poder culminar mis estudios y permitiéndome formarme tanto profesional como personalmente.

Y, por último, a mi “pequeñín”. Mi amigo y compañero de cuatro patas que ha estado conmigo durante más de trece años y al que recurría siempre cada vez que tenía un mal día.

Daniel Paredes García

Sevilla, 2021

Resumen

El confort ambiental es un concepto que se relaciona con el bienestar físico y social de los individuos cuando experimentan una sensación de satisfacción respecto a las condiciones ambientales existentes. Los avances tecnológicos y la necesidad de las personas de conseguir dicho confort, lo convierte en un tema cuyo objetivo es intentar satisfacer estas necesidades.

Es por ello que este trabajo se enmarca en el contexto de eficiencia energética de la Unión Europea, tratando sobre el control de calidad ambiental de interiores y la gestión óptima de energía.

Se trata de realizar un análisis de la distribución espacial, de las medidas de calidad ambiental del interior de una habitación, mediante el uso de una red de sensores inalámbricos (WSN) y la aplicación de diferentes técnicas de interpolación, tanto deterministas como geoestadísticas, con el fin de obtener una conciencia espacial total de la temperatura. Las lecturas de temperatura muestreadas por los nodos-sensores tienen la capacidad de informar al sistema sobre las temperaturas en ubicaciones específicas, pero no describen completamente toda el área en cuestión, de ahí la realización del estudio mediante estimaciones de las diferentes zonas no cubiertas, con el fin de conocer la efectividad de estos métodos que permitan posteriormente una optimización para el control de las temperaturas, asegurando así un confort ambiental idóneo.

Abstract

Environmental comfort is a concept that relates to the physical and social well-being of individuals when they experience a sense of satisfaction with existing environmental conditions. Technological advances and people's need to achieve this comfort make it a subject whose objective is to try to satisfy these needs.

Therefore, this work is framed in the context of energy efficiency in the European Union, dealing with indoor environmental quality control and optimal energy management.

The aim is to perform an analysis of the spatial distribution of indoor environmental quality measurements in a room, using a wireless sensor network (WSN) and the application of different interpolation techniques, both deterministic and geostatistical, to obtain a total spatial awareness of temperature. The temperature readings sampled by the sensor-nodes could inform the system about temperatures in specific locations, but do not fully describe the entire area in question, hence the study was carried out by estimating the different areas not covered, to know the effectiveness of these methods that subsequently allow an optimisation for the control of temperatures, thus ensuring an ideal environmental comfort.

Índice

Agradecimientos	vii
Resumen	ix
Abstract	xi
Índice	xiii
Índice de Tablas	xv
Índice de Figuras	xvii
1 Introducción	1
1.1 Estructura de la Memoria	1
2 Redes de Sensores Inalámbricas	3
2.1 Redes de sensores inalámbricas	3
2.2 Arquitectura Hardware de un nodo sensor	4
2.3 Características de una WSN	5
2.4 Topologías de red	6
2.4.1 Topología en Estrella	6
2.4.2 Topología en Árbol	7
2.4.3 Topología en Malla	7
2.4.4 Topología Híbrida Estrella-Malla	8
2.5 Pila de protocolos de WSN	8
2.5.1 Capa física	9
2.5.2 Capa de enlace de datos	9
2.5.3 Capa de red	10
2.5.4 Capa de transporte	10
2.5.5 Capa de aplicación	13
2.6 Estándares de comunicación	14
2.6.1 Bluetooth	14
2.6.2 Wi-Fi	14
2.6.3 UWB	15
2.6.4 WirelessHart	15
2.6.5 Zigbee	15
2.6.6 6LoWPAN	16
2.7 Campos de aplicación	16
3 Métodos de interpolación espacial	19
3.1 Métodos de Interpolación Deterministas	19
3.1.1 Polígono de Voronoi	20
3.1.2 Vecino Natural	20
3.1.3 Spline	21
3.1.4 IDW	22
3.2 Métodos de Interpolación Geoestadísticos	22
3.2.1 Estudio exploratorio de los datos	22
3.2.2 Análisis variográfico	23
3.2.3 Kriging o Krigeaje.	26

3.2.4	Validación Cruzada	30
4	Recreación de la red de sensores	31
4.1	<i>Origen de los datos</i>	31
4.2	<i>Funciones la recreación de la WSN</i>	33
4.2.1	ON	34
4.2.2	OFF	36
4.2.3	TOMA DE DATOS GENERAL	36
4.2.4	TOMA DE DATOS ESPECÍFICA	37
4.2.5	ANULAR NODO	38
5	Estudio de la distribución espacial	39
5.1	<i>Estudio exploratorio de los datos</i>	40
5.2	<i>Estimación del variograma</i>	42
5.3	<i>Estimación por Kriging Ordinario</i>	43
5.4	<i>Validación cruzada</i>	44
5.5	<i>Estimación de la temperatura durante un día concreto.</i>	47
5.6	<i>Comprobación de la fiabilidad del modelo</i>	50
6	Conclusiones	53
6.1	<i>Líneas de mejora e investigación futura.</i>	53
	Referencias	55
	Glosario	57

ÍNDICE DE TABLAS

Tablas 5-1. Estimaciones obtenidas por validación cruzada para la media global.	45
Tablas 5-2. Estimaciones obtenidas por validación cruzada.	46
Tablas 5-3. Temperaturas recogidas por cada nodo el día 4 de marzo a las 13:16.	47
Tablas 5-4. Resultado de los medidores de error durante el día 6 de marzo	49
Tablas 5-5. Resultado de las medias de los medidores de error durante el día 6 de marzo.	50
Tablas 5-6. Resultado de las medias de los errores obtenidas de los nodos no utilizados	51
Tablas 5-7. MAE obtenidos de la validación cruzada de cada nodo en °C	51

ÍNDICE DE FIGURAS

Figura 2-1. Esquema de una red de sensores inalámbricos [2].	3
Figura 2-2. Estructura de un nodo de una WSN.	4
Figura 2-3. Esquema topología en estrella.	6
Figura 2-4. Esquema topología en árbol.	7
Figura 2-5. Esquema topología en malla.	8
Figura 2-6. Esquema topología híbrida estrella-malla [8].	8
Figura 2-7. Pila de protocolo de una WSN.	9
Figura 2-8. Sistema de envío de paquetes Tiny TCP/IP.	12
Figura 3-1. Polígono de Voronoi. [22]	20
Figura 3-2. Método de interpolación Vecino Natural.	21
Figura 3-3. Cálculo del variograma según distancia y ángulo.	24
Figura 3-4. Variograma Experimental.	25
Figura 3-5. Modelos de Variograma Teóricos [28]	26
Figura 4-1. Esquema de la sala de laboratorio de Intel Berkeley.	31
Figura 4-2. Comparativa del laboratorio y la recreación de la red de sensores.	32
Figura 4-3. Interfaz de la demostración de la WSN en Matlab.	33
Figura 4-4. Envío de mensajes desde los Nodos al Nodo Sumidero.	34
Figura 4-5. Envío de mensajes desde el Nodo Sumidero al Servidor.	35
Figura 4-6. Interfaz tras la primera toma de datos automática. Nodo 5 desactivado.	35
Figura 4-7. Envío de mensajes Servidor-Nodo Sumidero (a) y Nodo Sumidero-Nodos (b)	36
Figura 4-8. Ventana emergente Toma de Datos Específica	37
Figura 4-9. Envío de mensajes Servidor-Nodo Sumidero y Nodo Sumidero-5 Nodos más cercanos	37
Figura 4-10. Envío de mensajes 5 Nodos más cercanos-Nodo Sumidero y Nodo Sumidero-Servidor.	37
Figura 4-11. Ventana emergente Anular Nodo.	38
Figura 4-12. Nodo nº14 eliminado.	38
Figura 5-1 . Distribución de nodos utilizados y reservados.	39
Figura 5-2. Histograma de la media global de los nodos.	40
Figura 5-3. Gráfica q-q de la media global de los nodos.	41
Figura 5-4. Histograma 8 de marzo sin normalizar (izquierda) y normalizado (derecha).	41
Figura 5-5. Variogramas direccionales de la media global	42
Figura 5-6. Variograma experimental omnidireccional de la media global	42
Figura 5-7 Variograma experimental y esférico de la media global	43
Figura 5-8. Estimaciones y varianza Kriging de la media global con modelo exponencial y esférico	44

Figura 5-9. Validación cruzada de los modelos exponencial y esférico	45
Figura 5-10. Variograma teórico esférico puntual (13:16 del día 4 de marzo).	46
Figura 5-11. Variogramas teóricos esféricos con las medias de cada hora.	48

1 INTRODUCCIÓN

En los tiempos en los que vivimos y en especial en las últimas décadas, el crecimiento de la tecnología en todos los ámbitos de la vida parece incluso ir por delante de lo que la mente humana puede llegar a imaginar. Esto también se aplica al campo de las redes de sensores. En los últimos años, las redes de sensores tradicionales, con gran cantidad de cableados y de corto alcance, han dado paso al uso de las redes de sensores inalámbricas, permitiendo llegar a zonas y ámbitos donde antes no podía ser aplicados. Tanto es así, que en campos como la medicina o en aplicaciones militares están utilizándose con gran éxito, permitiendo mejorar la calidad de vida de las personas o aportando información crucial en las defensas territoriales. Además de esto, estos avances tecnológicos permiten una mayor duración y una reducción de tamaños y costes en las redes de sensores.

En lo que respecta a este proyecto, el uso de estas redes tiene como principal objetivo permitir estimar las temperaturas en zonas interiores, con el fin de poder adaptar, en la mejor medida posible, la climatología del interior de las salas a las personas, produciendo una mayor satisfacción en ellas y permitiendo un mayor confort térmico. Así, debido a las imposibilidades de poder realizar una base de datos propia obtenida de una red de sensores inalámbricas personal, se realizará un proceso de estimación de temperatura de una sala de interior gracias a una base de datos facilitada por Intel que, con la distribución de 54 nodos-sensores repartidos por un laboratorio, realizó la toma de medidas de la temperatura durante un periodo de tiempo prolongado de forma que ha hecho posible la realización de dicho proceso de estimación.

Utilizando diferentes técnicas de interpolación, tanto deterministas como geoestadísticas, se realizarán las estimaciones de toda la sala con el fin de conocer la efectividad de los distintos modelos utilizados y poder plantear la posibilidad de estimar la temperatura en cada punto.

Estos estudios se llevarán a cabo con el software informático Matlab, permitiendo comprobar la viabilidad del propio estudio.

1.1 Estructura de la Memoria

Este proyecto está dividido en los siguientes capítulos:

- **Redes de Sensores Inalámbricas.** En este capítulo se realiza la descripción general de las WSN. Este abarca temas como la definición de estas o la explicación de la arquitectura y componentes que la conforman. Además, se describen sus características, se diferencian las diferentes topologías de red que pueden usar, los diferentes protocolos y estándar de comunicación permitidos y los distintos campos de aplicaciones en los que serían de utilidad.
- **Métodos de Interpolación Espacial.** Aquí se hace una explicación de los diferentes métodos de interpolación que podemos encontrar, dividiéndolos en dos grandes grupos, los métodos deterministas y los métodos geoestadísticos. Del primero de ellos, se hará hincapié en el método IDW; mientras en el segundo se centrará en el Kriging, siendo ambos los escogidos para la realización de las estimaciones de la sala.
- **Recreación de una Red de Sensores.** En este capítulo, mediante el software informático Matlab, se hace una simulación de cómo sería el funcionamiento de una WSN y las diferentes funciones que podrían realizarse con ella.
- **Estudio de la Distribución Espacial.** Dentro de este capítulo se realizan los estudios prácticos para conocer la distribución espacial de la temperatura en el interior de una sala. Para ello, mediante el uso de distintos métodos de interpolación, se seguirán los pasos que permitirá conocer las estimaciones

dentro de dicha sala y conocer la viabilidad y eficacia de los distintos métodos implementados.

- **Conclusiones y Líneas de mejora.** En este último capítulo se recogen las diferentes conclusiones alcanzadas tras la realización del proyecto y se exponen distintas medidas que permitirían un desarrollo más avanzado con nuevas ideas de mejora a partir de lo expuesto en este proyecto.

2 REDES DE SENSORES INALÁMBRICAS

Durante los últimos años, el uso de las redes de sensores inalámbricas está ganando gran protagonismo en diversidad de campos de aplicación debido a los avances que día tras día se realizan en el ámbito tecnológico, lo que hace que no solo mejoren en aquellos campos donde ya se utilizan sino que aparezcan nuevos donde aplicarse y desarrollarse.

2.1 Redes de sensores inalámbricas

Una red de sensores inalámbricos (Wireless Sensor Networks, WSN) es una red formada por un grupo de nodos sensores, cuyo número puede ir desde las decenas hasta los miles de ellos, de tamaño reducido que se despliegan a lo largo de nuestro campo de aplicación y que tienen capacidad de comunicación de manera inalámbrica, permitiendo la detección, recopilación y procesamiento de la información obtenida a partir de ellos[1].

Esta información recolectada por cada nodo sensor es enviada, también mediante una comunicación inalámbrica, hacia un nodo central de coordinación, al que se denomina *Estación base* o *Sink*, que es el encargado de recibir los datos del resto de nodos sensores. Este nodo *Sink*, el cual puede incluir también sensores si así se estima oportuno, es el encargado de realizar la función de enlace entre los nodos sensores de la red y el servidor, conectándose inalámbricamente con los nodos sensores y mediante una conexión TCP/IP con el servidor.

Por último, con la información recibida por los nodos sensores en el servidor, el usuario tendrá la capacidad de analizarla y procesarla.

En la Figura 2-1 se puede ver un esquema de una estructura habitual de una red inalámbrica de sensores.

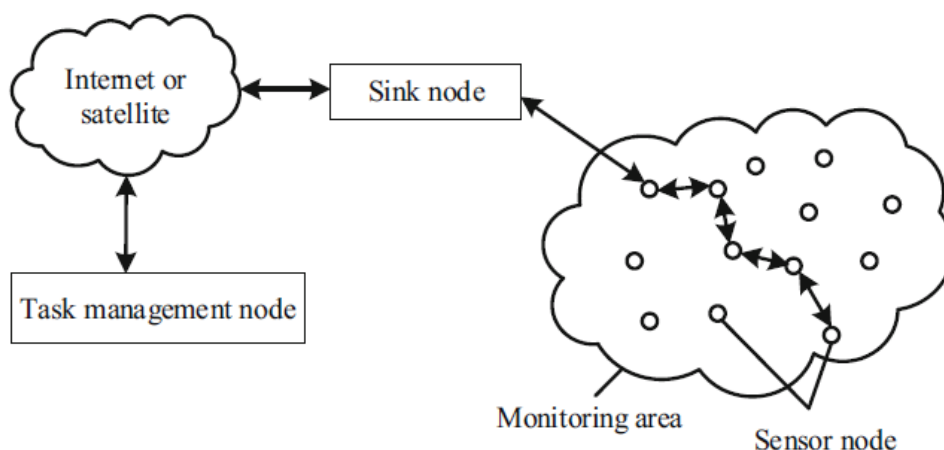


Figura 2-1. Esquema de una red de sensores inalámbricos [2].

2.2 Arquitectura Hardware de un nodo sensor

Los nodos sensores, también conocidos como motas, son elementos compuestos principalmente por un módulo sensor, un módulo de procesamiento, un módulo de comunicación y un módulo de energía.

En el módulo sensor encontramos dos elementos diferentes: uno o varios sensores y un convertidor AC/DC. Los sensores son dispositivos que permiten medir diferentes condiciones físicas. Entre ellos podemos encontrar sensores térmicos, sísmicos, acústicos, mecánicos, lumínicos, biológicos, ópticos, etc. Estos sensores envían la información mediante señales analógicas, de ahí que sea necesario la introducción de un convertidor para convertir dicha señal analógica en una señal digital para que pueda ser transmitida al módulo de procesamiento.

El módulo de procesamiento está formado por un procesador y una memoria. Es el encargado de coordinar todo el nodo, procesar la información recogida por los sensores y de gestionar el almacenamiento de esta información en la memoria. Entre los procesadores más destacados se encuentran las FPGAs, los microprocesadores y los microcontroladores. En la actualidad, debido al alto consumo de energía de las FPGAs, lo cual es un factor importante a tener en cuenta para la vida de los nodos, ya que, a diferencia de los microprocesadores, los microcontroladores incluyen tanto el procesador como la memoria, son estos últimos los más utilizados. Igualmente, aunque los microcontroladores contengan una memoria incorporada, es habitual la implementación de una memoria flash que libere esta memoria del controlador, permita una mayor recogida de datos y a una mayor velocidad, y que estos datos almacenados se mantengan incluso cuando el nodo deja de funcionar [3].

El módulo de comunicación está constituido por un transceptor, el cual es el encargado de permitir la transferencia de datos entre los diferentes nodos mediante una comunicación inalámbrica. La comunicación puede realizarse por diferentes medios como son la comunicación infrarroja, la óptica o por radio frecuencia, siendo esta última la más utilizada. Las bandas de frecuencia más utilizadas son la de 400 MHz, la de 800-900 MHz y la de 2.4 GHz o más conocida como la banda ISM (Industrial, Scientific and Medical). Debido a que el medio de radio es de uso libre y compartido, es necesaria la aplicación del protocolo de acceso o MAC, el cual se implemente entre el transceptor y el procesador. Al igual que los microcontroladores, los transceptores pueden funcionar en modos de transmisión, recepción, inactivo y suspensión. Debido a la importancia del ahorro energético y a que el modo inactivo consume prácticamente la misma energía que en modo de transmisión, es aconsejable utilizar el modo suspensión cuando no esté en funcionamiento.

El módulo de energía está compuesto por una fuente de alimentación, la cual es una batería. Debido a que en muchas ocasiones las WSN se ubican en lugares de difícil acceso, el aprovechamiento de estas baterías adquiere una importante relevancia para la durabilidad y el buen funcionamiento del nodo. Debido a que el principal consumo de batería se produce en los procesos de comunicación, se suele optar por un modelo donde los nodos sensores se encuentren en estado de suspensión y simplemente se activen cuando sea requerido, impidiendo así un gasto innecesario de energía. Con la ayuda de los avances en energías renovables, se están implementando fuentes de energía suplementarias que permitan una mayor durabilidad en el funcionamiento de los nodos. Estas fuentes de energías proceden, entre otros, del sol o los movimientos sísmicos, según el tipo de ambiente donde se desarrolle la aplicación de la WSN [4].

En la imagen de la Figura 2-2 se puede observar un esquema de la estructura de un nodo.

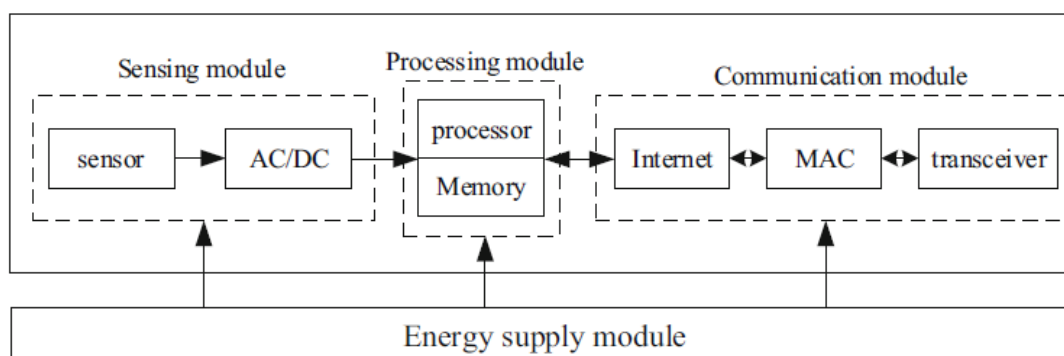


Figura 2-2. Estructura de un nodo de una WSN.

2.3 Características de una WSN

El hecho de que las WSN estén formadas por una gran cantidad de nodos sensores o tengan una comunicación inalámbrica, entre otras muchas de sus características, hace que estas redes tengan que enfrentarse a una serie de factores y problemas de diseño que las haga viables para su utilización. A continuación, se analizan estos factores cuyos resultados infieren directamente en el diseño de una WSN.

Tolerancia a fallos

Debido a que los sensores normalmente se utilizan en zonas donde las condiciones físicas no son las óptimas, los sensores pueden sufrir problemas de hardware o daños físicos. Este mismo hecho hace que sea de esperar que las redes inalámbricas, a diferencia de las redes cableadas, tengan fallos en la comunicación entre sus nodos. Es por ello, por lo que las WSN deben ser lo suficientemente robustas y poder reestructurarse de manera que, ante la falta de funcionalidad de algunos de sus nodos, la red mantenga su operabilidad, pudiendo ser capaz de crear rutas alternativas para poder enviar la información correctamente sin la necesidad de estos nodos perdidos. Igualmente, siempre que sea posible, es aconsejable la introducción de mecanismos en los nodos de manera que advierta de que estos no están funcionando correctamente, antes de perder por completo su funcionalidad [5].

Escalabilidad

Como ya se ha comentado anteriormente, las redes de sensores inalámbricas están formadas por una gran cantidad de nodos, que deben ser capaces de adaptarse a tener grandes cantidades de trabajo e incluir más nodos de los que originalmente se diseñaron. A esto nos referimos con escalabilidad. Una red es escalable cuando es capaz de mantener su funcionamiento correctamente incluso con la adición de nuevos nodos sensores y, con ello, una mayor cantidad de datos. Por tanto, los protocolos de comunicación de los nodos deben poder adaptarse a estas circunstancias para así mantener un rendimiento óptimo de la red [1].

Reducción de costes

Una de las principales ventajas de las redes de sensores inalámbricas es la reducción de costes respecto a las redes de sensores tradicionales con cable. Gracias al reducido tamaño de los nodos sensores y a evitar ese cableado, las infraestructuras de las redes se hacen más dinámicas, permitiendo la modificación de sus topologías cuando sea necesario.

Al tratarse, por norma general, a los nodos como unidades desechables, en caso de sufrir defectos el coste de producción e instalación de estos debe ser mínimo. Este es otro punto a favor de las redes inalámbricas respecto a las tradicionales ya que sus costes son inferiores, lo que hace que se puedan instalar un mayor número de nodos sensores a menor precio y así, poder adaptar la cantidad de nodos utilizados a nuestras necesidades.

Consumo de energía

En los nodos, el consumo de energía se produce en la transmisión, recepción de datos y en los estados de inactividad. Además de esto, la utilización de redes con un gran número de nodos hace que el consumo sea mayor al relacionarse un mayor número de nodos entre ellos y, por tanto, consumir más energía. En entornos donde el uso de redes no tiene complicaciones de instalación, como por ejemplo en hogares u hospitales, el reemplazo de las baterías de los nodos sensores no suponen grandes dificultades. En cambio, en entornos más hostiles, como en aplicaciones militares, reemplazar una batería por otra se convierte en una tarea verdaderamente complicada. Por todo esto, es necesario adaptar nuestras redes a las distintas necesidades y reducir al máximo el consumo de energía para poder preservar el mayor tiempo posible la viabilidad de nuestra red de sensores [6].

Comunicación

Al tratarse de redes inalámbricas que suelen operar mediante una comunicación por radiofrecuencia en la banda ISM y que se utilizan en numerosas ocasiones en ambientes poco óptimos, es posible que la transmisión de datos no sea siempre posible y se produzcan fallos en dicha transmisión. Es por ello que se hace necesario el uso de sistemas de enrutamientos que optimicen al máximo el rendimiento de la red, por lo que se busca que nuestras comunicaciones tengan la menor latencia y la mayor exactitud posible para que, junto con el resto de las características, la instalación sea lo más viable posible.

Seguridad

Las redes de sensores transmiten una gran cantidad de información mediante una comunicación inalámbrica accesible a todo el mundo, lo que las hace vulnerables frente a posibles ataques cibernéticos que expongan a los datos tomados por estas redes. Este es un aspecto importante en todos los ámbitos a los que se pueden aplicar las WSN, desde la exposición a terceros de nuestros datos en el ámbito doméstico como a los ataques recibidos por el enemigo en el ámbito militar. Por tanto, los nodos deben ser capaces de velar por la seguridad de la red, cumpliendo una serie de requisitos tales como un control que impida el acceso al medio por parte de intrusos; la encriptación de mensajes para que se obtenga una confidencialidad de estos; o impedir el acceso a los mensajes por parte de los atacantes de manera que se mantenga la integridad de estos mensajes.

2.4 Topologías de red

Como bien es sabido, los nodos sensores se encargan de recoger una serie de datos del entorno y transmitir esta información al nodo *Sink* o Estación base. Es posible que la información sea transmitida directamente del nodo sensor a la estación base o que se requiera de nodos intermedios antes de llegar a ella. Según la estructura utilizada para la comunicación de la información en la red, podemos encontrar diferentes topologías para las redes de sensores inalámbricos.

2.4.1 Topología en Estrella

La topología en estrella es la más simple que se puede encontrar. En ella, cada uno de los nodos-sensores se encuentran conectados únicamente con la estación base, siendo ella la única que puede enviar y/o recibir información de los nodos que conforman la red inalámbrica, es decir, los nodos no se pueden transmitir información entre ellos.

Las principales ventajas que presenta esta topología son la simplicidad en el enrutamiento, ya que solo existe comunicación entre el nodo origen y el nodo destino; la fiabilidad y seguridad que otorga el hecho de solo existir un salto en la transmisión de la información, haciendo que se dificulte la posibilidad de pérdida de información en el camino, o el reducido consumo energético producido en los nodos sensores, ya que solo se comunican con la estación base cuando se requiere información directa de ellos.

Por su parte, esta topología también presenta una serie de desventajas como son el excesivo consumo producido en la estación base ya que, al ser el único en contacto con los nodos sensores, es el que mayor consumo energético tiene. También presenta como inconveniente que es un sistema que solamente es utilizable cuando el campo de amplitud de la red de sensores no es muy extenso, debido a que las comunicaciones entre los nodos y la estación base tendrán un rango máximo de transmisión que en caso de excederse impediría la comunicación entre ellos. Otra desventaja que presenta esta topología es el hecho de que, si un nodo pierde la comunicación con la estación base debido, por ejemplo, a una congestión producida en la esta última por el hecho de estar recibiendo mucha información de otros nodos, toda la información enviada por el nodo se perdería, al estar conectado única y exclusivamente a la estación base y no tener una ruta alternativa de transmisión [7].

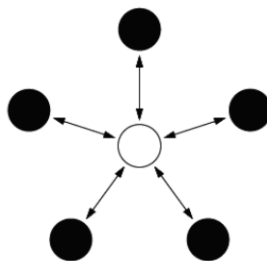


Figura 2-3. Esquema topología en estrella.

2.4.2 Topología en Árbol

A esta topología también se le conoce como topología jerárquica, en la cual la estación base es la encargada de recolectar los datos que llegan desde las distintas ramificaciones que parten de ella. Se puede decir que esta estructura está formada por una colección de redes en estrellas. De cada nodo sensor parte uno o varios nodos de los que a su vez pueden conectarse otros nodos en un nivel inferior. Así la transmisión de la información, a diferencia de las redes en estrella, se produce mediante un enrutamiento de múltiples saltos donde un nodo sensor adquiere los datos y este los transmite a su nodo padre, es decir, al nodo conectado a nivel superior. Este nodo padre puede agregar a la información recibida los datos obtenidos por sus propios sensores y enviar la información acumulada a su propio nodo padre, repitiéndose así el proceso hasta llegar la información de toda la ramificación a la estación base.

La ventaja que presenta esta topología es que el consumo de energía producido en el nodo *Sink* se reduce considerablemente al no tener conexión directamente con todos los nodos que conforman la red. Igualmente, también se produce una descongestión en las transmisiones realizadas en la estación base. Además, también se produce la ventaja de que es posible ampliar la red de sensores tanto en número como en rango de amplitud al no recaer todas las conexiones directamente en la estación base.

Por su parte, la principal desventaja producida en este tipo de topología es que, al igual que en las redes en estrella, cuando un nodo pierde un enlace de comunicación, se produce la pérdida de la comunicación total con la estación base, ya que al fin y al cabo pierde conexión con toda la ramificación a la que pertenece. Sin embargo, en esta topología este problema se agrava ya que, en caso de perderse la conexión en alguno de los nodos intermedios de la ramificación, se perdería la conexión con ese nodo y con todos los nodos que se produzcan aguas abajo de él.

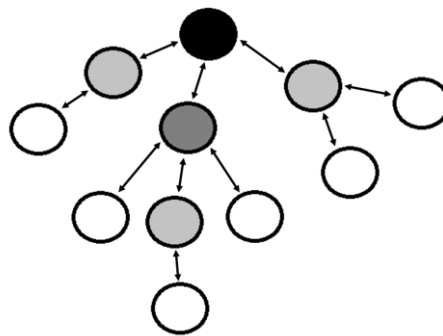


Figura 2-4. Esquema topología en árbol.

2.4.3 Topología en Malla

En la topología en malla, todos los nodos se encuentran interconectados unos con otros, directa o indirectamente, por lo que es posible la transmisión de datos entre todos los nodos mediante una comunicación multisalto. Así, un nodo puede enviar o recibir datos de otro nodo que se encuentre fuera de su radio de alcance de comunicación al existir nodos intermedios interconectados entre ellos. Además de esto, la comunicación a la estación base se sigue manteniendo con todos los nodos, ya sea si se encuentran conectados con ella o transmitiendo la información mediante nodos intermedios.

Con esta topología, se soluciona uno de los principales problemas de las topologías comentadas con anterioridad, la escalabilidad. Cuando un nodo falla, los nodos interconectados a él no pierden la información porque pueden seguir manteniendo la conexión con los nodos vecinos a los que se encuentran interconectados. Además, es muy tolerante a fallos ya que la información se puede transmitir por diferentes rutas en caso de que algún nodo intermedio falle, evitando así la pérdida de información.

Por el contrario, este sistema requiere de un gran consumo de energía debido a la gran cantidad de interconexiones existentes entre los nodos, suponiendo una gran desventaja ya que la batería de los nodos es limitada. Además de esto, el tiempo de retardo en las comunicaciones pueden ser excesivas debido al alto tráfico de datos producidos entre los nodos.

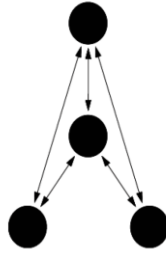


Figura 2-5. Esquema topología en malla.

2.4.4 Topología Híbrida Estrella-Malla

La topología híbrida entre la estrella y la mallada está formada tanto por nodos sensores que tienen la capacidad de comunicación con otros nodos sensores y con la estación base, llamados nodos enrutadores, como por nodos sensores que simplemente tienen la capacidad de comunicación con un único nodo, ya sea con un enrutador o directamente la estación base. Esta topología permite crear una red de comunicación en la que se aúnan los aspectos positivos de ambas topologías en una sola.

Con esta topología se consigue un considerable ahorro en el consumo de energía, ya que no todos los nodos se encuentran constantemente en activo para enviar o recibir información de los nodos vecinos, pero a su vez se mantiene la alta fiabilidad y robustez adquirida de las redes malladas al encontrarse todos los nodos interconectados entre ellos mediante una serie de nodos con la capacidad de intercomunicación con el resto de nodos. En caso de querer ampliar la red de sensores, basta simplemente con agregar el nuevo nodo conectándolo a uno de los nodos encaminadores.

Sin embargo, también existe alguna desventaja. El hecho de que los nodos enrutadores sean limitados y de él cuelguen otros nodos puede originar una congestión y/o retrasos en la comunicación si reciben muchas informaciones de los nodos dependientes de ellos.

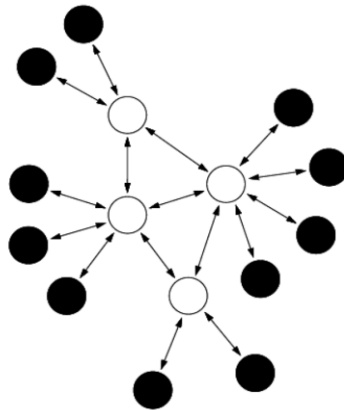


Figura 2-6. Esquema topología híbrida estrella-malla [8].

2.5 Pila de protocolos de WSN

Las redes de sensores inalámbricas actuales difieren de las redes de sensores tradicionales en aspectos como el uso de las baterías que, como se ha comentado anteriormente, al ubicarse los nodos sensores en ocasiones en lugares remotos, no tienen la posibilidad de ser reemplazadas y al no tener una estructura habitual organizada previamente, necesitan de un protocolo de comunicación específico para la transmisión de datos. Este protocolo debe ser capaz de combinar requisitos de potencia para reducir al máximo la energía utilizada, tener conocimientos de los enrutamientos para poder transmitir la información de manera eficiente y promover los esfuerzos cooperativos de los nodos-sensores que permitan mantener la conectividad de la red en caso de cambio de topología debido a la adición, fallos o movimiento de nodos.

El protocolo de comunicación utilizado en las WSN tiene una pila de protocolo que consta de cinco capas, tal y como se puede apreciar en la Figura 2-7. Estas capas son la capa física, la capa de enlace de datos, la capa de red, la capa de transporte y la capa de aplicación.

La capa física transmite y recibe señales a través de un medio de comunicación físico. La capa de enlace de datos es responsable de la multiplexación del flujo de datos, la transmisión y recepción de tramas de datos, el acceso al medio y el control de errores. La capa de red enruta los datos desde la capa de transporte. La capa de transporte es la entrega de los datos requeridos por la capa de aplicación de manera confiable. La capa de aplicación contiene una variedad de protocolos de aplicación para generar diferentes aplicaciones de red de sensores [9].

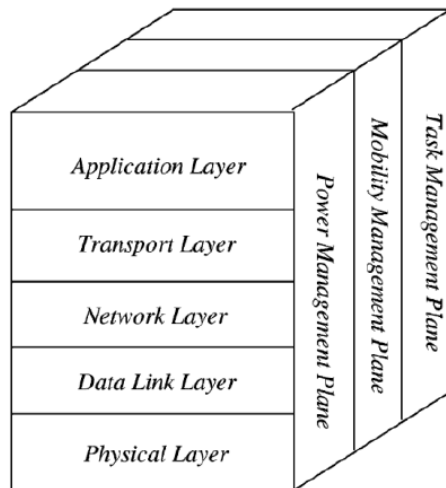


Figura 2-7. Pila de protocolo de una WSN.

2.5.1 Capa física

La capa física es el primer nivel de la pila de protocolos. Es la capa de red más básica y se encarga de convertir el flujo de datos de bits en señales para poder ser transmitido por el medio de comunicación. En esta capa se especifican las frecuencias de transmisión, el esquema de modulación, la detección de la señal y el cifrado de datos.

Debido a que uno de los principales criterios que deben seguir las redes de sensores inalámbricas es el ahorro en consumo de energía, es necesario hacer un estudio para saber qué tipo de comunicación es más óptima entre los nodos; esto es, comprobar si es mejor una comunicación de distancias cortas entre los nodos, pero con un mayor número de saltos de información entre ellos; o bien, una comunicación entre nodos que se encuentran a mayor distancia, pero con menor número de saltos. Por norma general, la solución óptima es un modelo intermedio, en el cual las distancias no sean excesivamente altas y un número de saltos moderados, pero más cercano a la opción de menores distancias y mayor número de saltos ya que, en redes con varios saltos, la agrupación de nodos reduce el consumo de energía en la comunicación, equilibra el tráfico y mejora la escalabilidad cuando aumenta el tamaño de la red.

2.5.2 Capa de enlace de datos

Esta capa es la encargada de la transferencia del flujo de datos. Es responsable de la multiplexación de flujos de datos, detección de tramas de datos, el control de acceso al medio (MAC) y control de errores. De todos ellos, el punto más importante es el MAC. Con él, se busca alcanzar un mayor rendimiento de la red y de latencia de entrega de datos. Esto permitirá que la estructura de la red esté mejor organizada para facilitar la comunicación entre nodos y evitar colisiones en la entrega de la información.

Además, otro de los objetivos del MAC es alcanzar un mayor ahorro en el consumo de energía. El mayor consumo de energía se produce en la colisión, escucha, control de sobrecarga y escucha inactiva. Es por ello por lo que es importante la elección del estándar de comunicación elegido en el protocolo MAC.

2.5.3 Capa de red

La capa de red es la encargada de enviar los datos obtenidos por los sensores desde los nodos hasta el nodo sumidero, ya sea directamente o mediante saltos con un camino intermedio por otros nodos. Como ya se ha comentado, uno de los puntos principales a tener en cuenta es el consumo de energía de los nodos y es por ello que adquiere gran relevancia, ya que es necesario que la información siga las rutas más óptimas o sean capaces de adaptarse a otras nuevas en caso de pérdidas de nodos-sensores. Además de esto, para el diseño de protocolos de enrutamiento para WSN se debe considerar las limitaciones de recursos de los nodos de la red, y la posibilidad de pérdidas y retrasos de paquetes.

Existen diferentes protocolos de enrutamiento. Entre los más comunes encontramos los siguientes:

- ***Flooding o Inundación*** Es el protocolo de enrutamiento más sencillo. En este método, cuando un nodo recibe un paquete de datos, éste se lo envía a todos sus vecinos, éstos a los suyos y así sucesivamente. En este caso, el mensaje sigue todas las rutas posibles y, por tanto, siempre llegará a su destino. Este sistema de inundación no requiere de mantenimiento de la tipología de red ya que aborda todas las rutas; sin embargo, cuenta con una serie de problemas que la hace ser la menos óptima. Entre sus problemas encontramos la superposición de áreas entre nodos, lo que hace que se transmita el mismo mensaje multitud de veces entre los mismos nodos o, el más importante, el uso ineficiente de recursos energéticos, lo que hace que este sistema sea poco eficiente energéticamente hablando [1].

- ***Gossiping***. En él, los nodos, cuando reciben un mensaje, en lugar de reenviarlo a todos sus vecinos, lo envía a un solo nodo vecino elegido al azar y así todos sucesivamente hasta alcanzar su destino. Así, se consigue paliar los defectos de la inundación en lo referente al malgaste de recursos y a la superposición del mensaje, pero cuenta con la desventaja de que el mensaje puede llegar a tardar en recepcionarse una gran cantidad de tiempo, incluso superar el número máximo de saltos posibles en el envío.

- ***SPIN, Protocolo de información vía negociación (Sensor Protocols for Information via Negotiation)***. Este método disemina toda la información en cada nodo a todos los nodos de la red asumiendo que todos los nodos son potenciales estaciones base. Con este sistema de negociación se superan los problemas de superposición de datos resolviendo los problemas existentes en la inundación. Además, este protocolo sí tiene en cuenta el nivel de energía restante en cada nodo por lo que se adapta para minimizar el gasto energético, superando los problemas energéticos y de ancho de banda que tiene el *gossiping*, al enviar copias del mensaje a nodos de forma innecesaria. Para que la negociación funcione, los nodos deben describir los datos usando metadatos que servirán al resto de nodos para saber si las informaciones les será de utilidad o no.

El funcionamiento de este protocolo consiste en tres pasos. En el primero de ellos, el nodo con la información envía un mensaje de aviso a sus vecinos haciéndoles saber que tiene información enviando simplemente metadatos. Aquel nodo vecino que esté interesado enviará al nodo con la información un mensaje de requerimiento de datos y, por último, el nodo con la información enviará a su nodo vecino la información. Esto ocurrirá así sucesivamente hasta llegar el mensaje al destino [1].

- ***Difusión directa***. En este método, cuando la estación pide los datos diseminando por intereses, los nodos intermedios los propagan. Cada nodo que recibe los intereses prepara un gradiente de la información hacia los nodos vecinos y según la fuerza con la que reciban el gradiente mejor o peor será la ruta. Todos los nodos son dependientes de la aplicación, lo que permite ahorrar energía seleccionando las rutas de forma empírica y también, cacheando y procesando los datos en la red. El cacheo de datos aumenta la eficiencia, robustez y la escalabilidad de la coordinación entre nodos. Por el contrario, encontrar datos que satisfagan las peticiones produce una carga extra en los nodos.

2.5.4 Capa de transporte

Los protocolos de transporte se utilizan para mitigar la congestión y reducir la pérdida de paquetes, proporcionar equidad en la asignación de ancho de banda y garantizar la confiabilidad de un extremo a otro. Sin embargo, los protocolos de transporte tradicionales que se utilizan actualmente para Internet, es decir, UDP y TCP no se pueden implementar directamente para WSN debido a dificultades que aparecen en ellos, como son la falta de fiabilidad y de control de flujo y congestión en el caso de las UDP; o bien, las sobrecargas en la conexión, los mecanismos de control de flujo y congestión, la degradación del rendimiento o las

respuestas tardías en las TCP. Para evitar estos problemas, existen diferentes tipos de protocolos de transporte que pueden ser utilizados en las WSN. Algunos de los más utilizados son los siguientes:

- **CODA, Detección y prevención de congestión (Congestion Detection and Avoidance).**

Cuando un nodo sensor tiene una ocupación de su búfer y una carga del canal inalámbrico superior a unos valores umbrales predeterminados, este sistema detecta esa congestión en el nodo y envía un mensaje a sus vecinos para que reduzcan la velocidad de envío de información a través de una contrapresión salto a salto de bucle abierto.

En bucle cerrado de extremo a extremo, el nodo sensor, antes de enviar el paquete, comprueba si el canal por el que desea transmitir la información se encuentra congestionado. Si observan el canal ocupado más tiempo del establecido, los nodos establecen un bit de control en el encabezado del paquete que informa a la estación base del problema. Gracias a este bit, la estación base es capaz de informar a los demás nodos de envío para que reduzcan su velocidad de envío mediante un mensaje ACK. Cuando la congestión está resuelta, el nodo receptor envía un mensaje ACK para que los nodos origen retomen su velocidad de envío de datos.

CODA no tiene mecanismo de confiabilidad, cuanto mayor es el número de nodos de origen peor es su desempeño contra la congestión y sus tiempos de latencia y rendimiento empeoran considerablemente [10].

- **ESRT, Transporte confiable de eventos a sumidero (Event to Sink Reliable Transport).**

Este protocolo tiene principalmente el objetivo de conseguir fiabilidad para evitar congestión y, además, reducir al máximo el consumo de la energía.

En esta tipología, el nodo sumidero está interesado en la información colectiva de los nodos sensores dentro del radio de un evento y no en los datos individuales de los propios sensores. Para comprobar si existe fiabilidad, el *Sink* comprueba la frecuencia de los eventos recibido con éxito dentro de un intervalo de tiempo y lo toma como referencia. El hecho de que el algoritmo se implemente en el nodo destino *Sink*, hace que el consumo de energía de los nodos sensores se vea reducido. En este algoritmo existen 5 estados en los que podemos encontrar las WSN: "sin congestión, baja fiabilidad (NC, LR)", "sin congestión, alta fiabilidad (NC, HR)", "congestión, alta fiabilidad (C, HR)", "congestión, baja fiabilidad (C, LR)", y "región operativa óptima (OOR)". En función del estado en que se encuentren aumentará o disminuirá la frecuencia de eventos según haya congestión y/o fiabilidad, con el fin de intentar operar siempre en el estado OOR, siendo en ella donde se produce una mayor fiabilidad y se consigue adaptar el algoritmo para reducir el consumo energético.

A pesar de esto, el algoritmo ESRT tiene algunas desventajas que afectan a su rendimiento como es que ESRT se encuentra a un solo salto de los nodos sensores, lo que hace que en sistemas donde los nodos sensores no se comuniquen directamente con el *Sink*, sino que transmitan la información entre nodos hasta llegar a él, no pueda aplicarse este sistema. Otro punto negativo es que la frecuencia de envío de eventos se adjudica a toda la red por igual. Debido a que cada nodo sensor en la red pueda tener un tráfico de datos diferente, este hecho puede provocar que el protocolo sea ineficiente o incluso que se produzca congestión [11].

- **Tiny TCP/IP.** Debido a que el sistema TCP/IP no es implementable a las WSN por su capacidad de almacenamiento y su tamaño, se implementó este sistema para poder proporcionar la misma confiabilidad de extremo a extremo y de salto a salto. El Tiny TCP/IP se puede implementar en una memoria flash o en la memoria incorporada de los microcontroladores. Para poder utilizar este protocolo, la ubicación de los nodos sensores debe ser conocida y que cada uno de ellos caiga en alguna de las subredes existentes. Esto se debe a que las direcciones TCP/IP están formadas por 12 números separados en 4 grupos de 3 dígitos. Estos 3 dígitos, escritos de forma binaria, están compuestos por ocho números de valores 0 y 1. Por tanto, las direcciones están compuestas por 4 octetos binarios. De estos cuatro octetos, los dos primeros de ellos son utilizados para identificar la subred a la que pertenece el nodo sensor, mientras que los dos últimos se utilizan para conocer la ubicación del nodo sensor dentro de la subred.

Uno de los aspectos que diferencia el protocolo TCP/IP con el Tiny TCP/IP es que mientras el primero de ellos transmite la información de extremo a extremo, el segundo almacena los paquetes en la memoria caché dentro de los nodos de la subred. Con esto se consigue que en caso de que se pierda la información no sea necesario la retransmisión completa de la información desde el nodo origen, sino que los nodos intermedios tienen almacenada en su memoria caché el paquete, haciendo posible su recuperación en posiciones más avanzadas. Viendo esto con un ejemplo, como el mostrado en la Figura 2-8, se observa que existen nodos intermedios (A y B) entre el emisor y el receptor. El nodo A almacena el primer paquete, y el nodo B almacena el segundo paquete. En caso de que se pierdan ambos paquetes, cada uno se reenvía desde su respectiva caché, es decir, el

paquete 1 del nodo A y el paquete 2 del nodo B. En el peor de los casos, ninguno de los nodos intermedios tiene una copia del paquete faltante y el remitente tiene que reenviar ese paquete [1] [6].

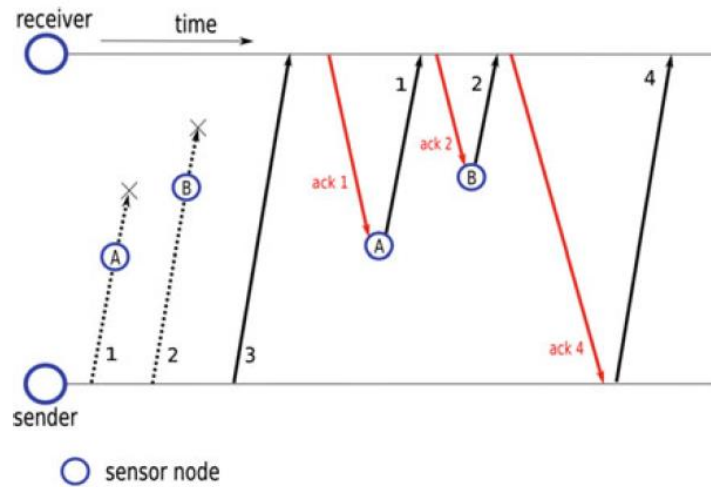


Figura 2-8. Sistema de envío de paquetes Tiny TCP/IP.

Aun así, la utilización de este protocolo también presenta una serie de problemas como son el hecho de que es necesario conocer la ubicación de todos los nodos de la red, saber qué nodo almacena la información en caché hace que el diseño de la red sea más complejo, que la confiabilidad de la red dependerá de la capacidad de almacenamiento de la memoria caché o no tener un control de la congestión definido.

- **GARUDA.** Toma su nombre de un ave mitológico del hinduismo. Es un protocolo de transporte de entrega de datos en sentido descendente que garantiza la fiabilidad para WSN. Esta confiabilidad se define garantizando la entrega en todo el campo de aplicación, en todas las subregiones formadas por los nodos sensores y a un conjunto mínimo de nodos en cada subregión

GARUDA está diseñada con una infraestructura central y con un proceso de recuperación basado en NACK de dos etapas. La infraestructura central se construye utilizando método de entrega del primer paquete. Este método garantiza la entrega del primer paquete mediante un pulso WFP (Wait-for-First-Packet), donde el nodo sumidero realiza una serie de pulsos cortos realizados. Los nodos sensores dentro del rango de transmisión del sumidero recibirán este pulso y esperarán la transmisión del primer paquete. La entrega del primer paquete determina el recuento de saltos desde el receptor hasta el nodo. Los nodos con $3 \cdot i$ (siendo i un número entero) saltos del sumidero se seleccionan como nodos centrales. Los nodos restantes (no centrales) se denominan nodos de segundo nivel. Un nodo central debe conectarse igualmente al menos a otro nodo central cercano.

GARUDA utiliza un mensaje NACK para la detección y notificación de pérdidas. El proceso de recuperación de pérdidas de paquetes, como se ha comentado anteriormente, se realiza en dos etapas. En la primera de ellas, los nodos centrales son los que se encargan de recuperar el paquete. Un nodo central envía una notificación a otro nodo central informando de que se ha producido la pérdida de un paquete para así poder recuperarlo. En la segunda de ellas, son los nodos no centrales los que solicitan a sus nodos centrales la información por pérdida de paquete. Con esto se garantiza la confiabilidad de toda la red [1].

Como contrapartida, GARUDA también tiene una serie de puntos negativos. GARUDA solo realiza la espera por pulsos en el primer paquete emitido, por lo que solo se garantiza la entrega del primer paquete. En grandes WSN, el tiempo de recuperación de paquetes puede ser muy alto. Por último, este protocolo no garantiza el control de la congestión.

- **PSFQ, Bombear lentamente, recuperar rápidamente (Pump Slowly, Fetch Quickly).** Este protocolo se caracteriza por, tal y como indica su nombre, la distribución lenta de los paquetes desde el nodo sumidero hasta los nodos y la recuperación de la información perdida por parte de los nodos de manera rápida e inmediata solicitando los datos faltantes a sus nodos vecinos.

PSFQ utiliza dos temporizadores (T_{min} y T_{max}) [12] para poder realizar sus tres funciones: bombeo, recuperación e informe. En la primera operación, el Sink transmite un paquete a sus vecinos en un tiempo

T_{min} hasta que se haya enviado a todos los nodos que conforman la red. La segunda operación se produce cuando un nodo, en un tiempo próximo pero menor a T_{max} detecta que se ha producido la pérdida de un paquete. Es entonces cuando este nodo envía un acuse de recibo negativo *NACK* a sus nodos vecinos para solicitar la información perdida. El hecho de utilizar el reconocimiento negativo en lugar del positivo (*NACK* vs *ACK*) permite que la comunicación sea más rápida, además de experimentar un ahorro energético al prescindir de acuses de recibos cada vez que se envía un paquete [13]. Por último, la tercera operación consiste en la solicitud del nodo *Sink* a los nodos que conforman la red de un informe con el estado de entrega de los paquetes. Cuando se realiza esta operación, el informe se envía desde los nodos más lejanos hasta llegar al nodo sumidero, mediante un mecanismo salto a salto y escalable. En cada salto producido, el nodo agrega su información al informe recibido, por lo que solo llega al nodo *Sink* un solo paquete.

Este protocolo PSFQ presenta algunas desventajas como son el hecho de que el envío de paquetes se realice de forma lenta, lo que puede provocar un cierto retraso en la información o que este protocolo no contempla un control de la congestión del sistema, por lo que está limitado para el uso de sistemas con bajo volumen de tráfico de datos.

- ***RMST, Transporte confiable de múltiples segmentos (Reliable multi-segment transport).***

Este método está pensado para proporcionar confiabilidad en la transmisión de la información desde los nodos-sensores de la red hasta la estación base.

Como su propio nombre indica, RMST utiliza un método de fragmentación para la transmisión de los paquetes, con el fin de reducir el tamaño de la transmisión. Estos segmentos fragmentados, al llegar al nodo sumidero, son de nuevo juntados y reensamblados. Para asegurar que todos los fragmentos llegan al destino, RMST tiene dos formas de actuación.

La primera de ellas consiste en un esquema salto a salto, donde los nodos intermedios entre el nodo origen y el destino almacenan los fragmentos en su memoria caché. Así, en caso de que se produzca la pérdida de fragmentos, cualquier nodo puede recuperar de los fragmentos faltantes a lo largo del camino.

La segunda forma de actuación tiene un esquema de extremo a extremo donde los nodos intermedios simplemente transmiten los paquetes sin almacenarlos en memoria caché. De esta tarea se encargan los nodos origen y destino, es decir, la estación base. Así, si se produjera la pérdida de fragmentos en el camino, sería necesario la petición de la información de los nodos hasta llegar al nodo origen para recuperar esa información. En ambas formas de actuación, la estrategia de notificación y recuperación de fragmentos se produce mediante *NACK*, con la diferencia en que en la primera de ellas se produce salto a salto, esto es, el nodo intermedio que haya perdido la información notificará aguas arriba con un *NACK* hasta que alguno le recupere la información; mientras que en la segunda forma de actuación se produce de extremo a extremo, teniendo que propagarse la notificación *NACK* el nodo origen.

RMST está pensado para trabajar con el protocolo de enrutamiento de difusión directa, ya que se sirve de la diseminación de los datos y de la distribución fiable de estos. Por su parte, RMST cuenta con una serie de desventajas como son la limitación de aplicación, ya que solo es adecuado cuando es posible la fragmentación y posterior reensamblado de la información; no ofrece ningún control sobre la congestión y, en especial en el método extremo a extremo, no es eficiente energéticamente [1] [6] [14].

2.5.5 Capa de aplicación

La capa de aplicación es el último nivel de la pila de protocolos en una WSN. Equivale, en su homólogo con el modelo OSI, a las capas de aplicación de presentación y de sesión. Es en esta capa donde se recopilan y gestionan los datos, por lo que es importante garantizar la fiabilidad de estos.

Cuando un usuario utiliza una aplicación o un programa, no interactúa directamente con la capa de aplicación. Es el programa el que interactúa con ella, haciendo que para el usuario no sea complejo el manejo de dichas aplicaciones con la utilización de comandos más específicos.

Al igual que en el resto de capas, según la funcionalidad para el usuario que vaya a utilizar la WSN, existirá una serie de protocolos que se adaptarán mejor a dichas necesidades. Aunque esta capa de la pila es la menos explorada y, por tanto, la que más posibilidades tiene de crecer, existen una serie de protocolos principales, los cuales son el protocolo de gestión de sensores, el protocolo de asignación de tareas y publicidad de datos y el protocolo de consulta de sensores y difusión de datos [15].

- **SMP, Protocolo de gestión de sensores (Sensor Management Protocol).** Los administradores del sistema interactúan con las redes de sensores mediante el uso de SMP. En muchas aplicaciones en las que se utilizan las WSN es necesario el conocimiento de la ubicación de los nodos sensores, como puede ser en el ámbito militar donde, además de conocer la ubicación exacta de los datos recibidos, es indispensable garantizar una estimación de la ubicación ya que sino los enemigos podrían manipular la ubicación de estos nodos. Como las WSN consisten en nodos que no tienen ID globales y, por lo general, no tienen infraestructura, la SMP necesita acceder a los nodos utilizando nombres basados en atributos y direccionamiento basado en ubicación. SMP proporciona operaciones para realizar tareas como intercambiar datos de búsqueda de ubicación, tiempo de sincronización de nodos sensores, mover los nodos, activar o desactivar los nodos, consultar el estado de la red y de los nodos o la autenticación y distribución de claves para la transmisión segura de los datos.
- **Protocolo de asignación de tareas y anuncio de datos (TADAP, Task Assignment and Data Advertisement Protocol).** Se utilizan para manejar los datos que interesen a los usuarios y anuncios de estos datos de los nodos de sensores. Los usuarios consultan los datos que les interesan, enviando su petición a un nodo sensor, un subconjunto de los nodos o la red completa; y los nodos de sensor correspondientes responden anunciando los datos solicitados. TADAP proporciona al software del usuario interfaces eficientes para la difusión de intereses que también apoyan operaciones energéticamente eficientes de nivel inferior.
- **Protocolo de consulta de sensores y difusión de datos (SQDDP, Sensor Query and Data Dissemination Protocol).** Este protocolo proporciona a las aplicaciones de usuario interfaces para emitir consultas, responder a ellas y recopilar respuestas recibidas. Están diseñados para sensores basados en atributos o consulta de ubicaciones, es decir, a peticiones donde, por ejemplo, se desea conocer qué nodos-sensores son los que se encuentran por encima o por debajo de un umbral de un tipo de dato, como podría ser la humedad o la temperatura o se desea conocer los datos de los nodos sensores ubicados en cierta región. El primer ejemplo corresponde a los sensores basados en atributos y el segundo, a los de consulta de ubicaciones. Estos protocolos son útiles para garantizar el intercambio de mensajes entre el usuario y los sensores desplegados en un terreno con condiciones cambiantes [16].

2.6 Estándares de comunicación

La tecnología inalámbrica es la vía esencial que permite la comunicación entre los diferentes sensores utilizados en la red. Por tanto, es necesaria la elección del estándar más acorde a las necesidades que se vayan a tener. Estos estándares utilizan las bandas *ISM* (Instrumentation, Scientific and Medical radio bands), las cuales son 902–928 MHz (EEUU), 868 – 870 MHz (Europa), 433.05–434.79 MHz (EEUU y Europa) y 314 – 316 MHz (Japón) y la banda de GHz de 2.4 – 2.4835 GHz (universalmente aceptada).

2.6.1 Bluetooth

La tecnología Bluetooth, la cual corresponde al IEEE 802.15.1, proporciona un área personal de red inalámbrica, *WPAN*, para intercambiar datos entre dispositivos compatibles con Bluetooth de bajo coste por radio, utilizando un modo de bajo consumo en la banda de 2.4 GHz y con una tasa de transferencia máxima de datos de 2 Mbps. Por el contrario, Bluetooth permite solamente la transmisión de corto alcance, con una distancia máxima aproximada de 10 m., por lo que no es óptimo su uso en las WSN, ya que las distancias entre nodos suelen ser bastantes más amplias. Otro aspecto negativo es que, en la estructura maestro-esclavo, Bluetooth solo permite la conexión con 8 esclavos, por lo que limita su configuración [9].

2.6.2 Wi-Fi

La tecnología Wi-Fi, la cual corresponde al IEEE 802.11, es un conjunto de estándares para comunicación de red de área local inalámbrica, *WLAN*, el cual opera en las bandas de radio de 2.4 GHz o en la de 5 GHz. Mientras que la ventaja principal de la banda de 2.4 GHz es la distancia que alcanza su señal y el poco efecto que tiene en ella la ubicación obstáculos como paredes o muros; la de la banda de 5 GHz es que la banda de frecuencia se encuentra menos congestionada, por lo que la velocidad de transmisión es más alta, pudiendo alcanzar hasta 10 Gbps, aunque esta es una velocidad teórica. Esta velocidad máxima también dependerá del

protocolo IEEE 802.11 utilizado, ya que en la actualidad existen diferentes de ellos, como son el los protocolos 802.11a, 802.11b, 802.11n, 802.11g, 802.11ac o 802.11ax, que es el más moderno. Como punto negativo, Wi-Fi carece de un modo de bajo consumo y es poco integrado. Por lo tanto, un WSN de baja potencia y altamente integrado no puede usar esta tecnología. Este protocolo es ampliamente utilizado en el ámbito residencial, industrial y comercial [9].

2.6.3 UWB

La tecnología Ultra Wide-Band, *UWB*, utiliza el estándar IEEE 802.15. Es una tecnología que tradicionalmente se ha utilizado en el ámbito militar, ya que permite una localización muy precisa por radar con gran resolución.

Su característica principal es la banda de radio, ya que consta de impulsos de entre 1 y 2 nanosegundos y ocupan al menos el 25% de la frecuencia central. El rango de frecuencias de estas tecnologías oscila entre los 3.1 y los 10,6 GHz. A su vez, la potencia consumida por los chips es muy baja, lo que permite una gran autonomía, mayor incluso que con Bluetooth o Wi-Fi. Otro rango característico de esta tecnología es su interferencia con otros sistemas, el cual es bajo debido a su gran amplitud de ancho de banda. Su seguridad es otro aspecto para destacar, ya que es posible el cifrado de las transmisiones con el fin de excluir posibles escuchas innecesarias o exteriores.

Pese a tener una gran cantidad de aspectos positivos, el principal punto negativo lo encontramos en su radio de alcance puesto que, al igual Bluetooth y Wi-Fi, su rango es demasiado corto, alrededor de unos 20 m., lo que los hace poco eficientes para su uso en estos ámbitos [17].

2.6.4 WirelessHart

Este protocolo se originó a partir del protocolo cableado HART (Highway Addressable Remote Transducer) para compatibilizarlo con los dispositivos cableados ya existentes. La tecnología WirelessHART es un estándar que se basa en el IEEE 802.15.4-2006, utilizando la banda de 2,4 GHz y una velocidad de transferencia de 250 kbps. Utiliza la modulación de espectro ensanchado de secuencia directa (DSSS) que es tolerante y discreta a otros sistemas inalámbricos de espectro ensanchado. La comunicación inalámbrica se coordina con TDMA (Time Division Multiple Access), que sincroniza los dispositivos en etapas de 10 ms. Esta tecnología utiliza la red de malla o en estrella, donde la información se va transmitiendo entre los nodos vecinos.

Existen una serie de criterios que deben seguirse a la hora de realizar una red con WirelessHART. Aunque estos criterios son óptimos para ser poder ser implantados en una WSN, suponen un impedimento más a la hora ser seleccionado como el estándar a utilizar. El primero de ello es la existencia de al menos 3 vecinos por dispositivo. El segundo de estos criterios es que deben existir al menos cinco dispositivos dentro del alcance en una propia red. El tercero es que, si hay más de 5 dispositivos dentro de la red, al menos debe tener alcance cada dispositivo con el 25% de los vecinos. Finalmente, el último requisito es que si existen dispositivos con tasas de actualización superiores a dos segundos deben estar dentro de dos veces el alcance efectivo de los dispositivos para maximizar la velocidad [18].

Además de estos requisitos, esta tecnología presenta otros problemas como son la alta latencia de paquetes y el consumo de energía en la red, que podrían limitar la escalabilidad del dispositivo.

2.6.5 Zigbee

La tecnología ZigBee es un estándar de comunicaciones basado en el estándar IEEE 802.15.4 para redes *WPAN*. En la actualidad, es el protocolo más utilizado en las WSN, ya que por sus características permite ser el idóneo para aplicaciones tales como la automatización de edificios, la asistencia médica o el IoT, el internet de las cosas. Con él se busca la simplicidad, el bajo coste, la baja velocidad de transferencia de datos y el bajo consumo de energía. Respecto a la comunicación, Zigbee tiene una velocidad de datos de 250 Kbps a 2,4 GHz, 40 kbps a 915 MHz y 20 kbps a 868 MHz, con un rango de operación de entre 10 y 75 m [19].

La tecnología Zigbee permite la conexión de los nodos en tipología estrella, malla o árbol cuya configuración básica, al igual que en Bluetooth, es maestro-esclavo, pero a diferencia de este, permite hasta la conexión de 254 nodos.

Respecto al bajo consumo, Zigbee permite que los dispositivos puedan encontrarse en modo suspensión, y conectarse o despertarse bien cada cierto tiempo preestablecido o bien cuando se le necesite. Este implicaría que el consumo de ellos se vería reducido mientras se encuentra “durmiendo”, ya que la secuencia de despertar, enviar los datos y volverse a suspender es de tan solo 15 ms.

En la tecnología Zigbee se pueden distinguir tres dispositivos de red: los dispositivos de función completa, que pueden funcionar como enrutadores, dispositivos finales o ambos; los coordinadores de Zigbee, que almacenan la estructura de la red; y los dispositivos de función reducida, que son simplemente los sensores y actuadores que se encargan de enviar y recibir la información.

2.6.6 6LoWPAN

Este protocolo hace referencia a la versión 6 del Protocolo de Internet (IPv6) sobre redes de área personal inalámbricas de baja potencia (6LoWPAN) el cual transmite paquetes también basándose en IEEE 802.15.4, utilizando la banda de radio de 868/914 MHz o 2,4 GHz [1].

En lo que se refiere a este protocolo, sus nodos pueden tener una estructura mallada o en estrella, en la que la velocidad de comunicación entre ellos se encuentra en un rango de entre 20 y 250 kbps [20].

La capa de enlace está formada por una supertrama de 16 intervalos, en los que las diferentes regiones se pueden dividir en activas e inactivas. Cuando está inactivo, se entra en modo suspensión, lo que le permite reducir el consumo energético. Cuando está activo puede existir un periodo de acceso a disputas y otro libre de ellas. Cuando está en el de acceso a disputas debe lidiar con el resto de dispositivos para poder ejercer la comunicación. Otra situación posible es la de modo escucha de bajo consumo con la que es posible reducir el consumo durante el periodo de actividad.

Como punto negativo a resaltar es la distancia máxima de comunicación entre nodos, la cual es únicamente de 10 metros. Este hecho hace que, como otros de los nombrados anteriormente, no sea óptimo para las redes de sensores inalámbricas, ya que las distancias suelen ser mucho mayores. Es por ello que, el uso de esta tecnología está orientada a entornos domésticos o de oficina, ligada al conocido como Internet de las cosas, IoT, donde las distancias entre los elementos son relativamente pequeñas.

2.7 Campos de aplicación

Gracias a los avances tecnológicos y de investigación aplicados a las WSN, estas han adquirido una gran flexibilidad para resolver problemas en una gran cantidad de ámbitos de la vida, lo que las capacita para poder ser implementado en multitud de campos de aplicación [14].

Aplicaciones militares

Ya que las WSN pueden estar compuestas de gran cantidad de nodos, tienen una gran tolerancia a fallos y son capaces de adaptarse a una nueva reorganización de forma rápida, las redes de sensores inalámbricas tuvieron sus inicios en el ámbito militar y es uno de los campos donde más investigaciones y avances se siguen realizando en la actualidad. Esto es debido a que su uso hace posible la exploración y reconocimiento del terreno de batalla, permite avistar objetivos militares, pueden ser ubicados en zonas de difícil acceso que permiten, por ejemplo, mediante sensores acústicos o de presión, conocer la proximidad del adversario. Otro uso en este ámbito de estas WSN es el espionaje mediante la utilización de micrófonos que permitan conocer los próximos movimientos del enemigo.

Como se ha comentado, la capacidad de reorganización estructural y comunicativa entre los nodos los hace muy efectivos también este ámbito, incluso tras la pérdida de alguno de los nodos por ataques enemigos.

Aplicaciones ambientales

El uso de las WSN está ganando cada vez más relevancia en dentro de las aplicaciones ambientales ya que suponen un gran apoyo para la lucha contra el cambio climático, los desastres naturales o la contaminación. En función del entorno a estudio y la escala de este los nodos sensores utilizados serán fijos o móviles. Así, las aplicaciones principales están relacionadas con el control meteorológico, geológico y la contaminación.

Monitoreo meteorológico

El objetivo de esto es, mediante el uso de nodos sensores inalámbricos, medir condiciones ambientales como la temperatura, presión ambiental, velocidad del viento, humedad o radiación que permitan controlar y estudiar las condiciones físicas y atmosféricas del entorno. La utilización de este tipo de redes inalámbricas permite la ubicación de nodos sensores en zonas más estratégicas al no necesitar de cableado, la instalación de una mayor cantidad de estas al tener un bajo coste y la adquisición y almacenamiento de mayor número de datos que por métodos más tradicionales no serían posibles obtener.

Con estas WSN, resulta más fácil y efectiva la predicción meteorológica, del clima, de las precipitaciones y de tormentas; la predicción de posibles huracanes, ya que los nodos ofrecen información de las velocidades del viento que permiten analizar la posibilidad de que estos surjan; la predicción de inundaciones al poder obtener información en tiempo real de las cantidades y niveles de lluvia; el análisis y predicción de sequías al disponer de nodos sensores que permiten un estudio de las condiciones del terreno; o la desglaciación, que mediante el estudio de la temperatura en estas zonas y la contabilización de las cantidades de agua líquida que se registra nos permite conocer la velocidad con la que se produce este fenómeno y sus consecuencias.

El uso de las WSN para este tipo de monitoreo ha ganado gran relevancia en los últimos años debido a que nos ayuda a conocer las causas y consecuencias del cambio climático.

Control geológico

Permite el control y supervisión de magnitudes geológicas posibilitando la predicción de catástrofes naturales. Dentro de estas catástrofes naturales destacan los tsunamis, movimientos sísmicos de la tierra o activación de volcanes. El hecho de poder instalar nodos sensores en zonas de difícil acceso y en zonas subterráneas de la tierra permite cuantificar y recoger datos de movimientos de la tierra o del mar que permitan predecir la posibilidad de que se produzcan estas catástrofes al poder realizarse fácilmente este seguimiento.

Contaminación

El hecho de poder disponer de una tecnología inalámbrica para la recolección de información hace que los nodos se puedan estructurar y comunicar entre ellos de forma que, ubicándolos en cualquier zona, permitan obtener datos de diferentes zonas para conocer la contaminación del aire, agua, tierra o acústica, entre otros. Por ejemplo, la ubicación de nodos sensores en zonas de cultivo permitirá conocer el grado de contaminación del terreno existente para así saber su fertilidad. Otro ejemplo es el utilizado en aguas cercanas a zonas industriales, donde la ubicación de WSN permitirá conocer el grado de contaminación química existente en las aguas para así controlar y preservar la vida marina existente en ella [9].

Aplicaciones médicas

Las WSN adquieren gran importancia en este tema ya que es posible su aplicación para monitoreo de pacientes, tratamiento médico o diagnósticos tempranos de problemas derivados con la salud.

Para personas con enfermedades físicas o mentales, las tecnologías inalámbricas permiten la conexión en todo momento de los pacientes con una asistencia médica en caso de ser requerida, por ejemplo, mediante nodos sensores con botones de ayudas que los conecta directamente con el personal médico.

En lo referido a personas con enfermedades fisiológicas en las que puedan intervenir órganos internos, el hecho de que exista esta tecnología permite que puedan ser implantados dentro del organismo con el fin de tener un control exhaustivo del órgano en cuestión. Un ejemplo de ello es la posibilidad de incluir marcapasos y un pequeño desfibrilador subcutáneo que previene la muerte súbita del paciente, como el instalado en agosto de 2020 por el Hospital Clínico de Valladolid [21].

El hecho de que la información se pueda recolectar y transmitir de manera inalámbrica permite que los pacientes, según las condiciones a las que esté expuesto, puedan seguir más fácilmente sus tratamientos y revisiones desde su propio hogar sin la necesidad de estar en el hospital al estar en todo momento controlado.

Aplicaciones agrícolas

El uso de estas tecnologías está cada vez más extendido en la agricultura de precisión permitiendo un monitoreo constante de las tierras de cultivo. Mediante la instalación de sensores que permitan medir condiciones climáticas, como la temperatura, humedad o radiación solar, u otros agentes fitosanitarios se puede conocer el estado de las plantaciones en todo momento. Con esto, es posible automatizar los tiempos de

regadíos en función de las necesidades del cultivo o conocer si existen agentes patógenos en ellos. Esto provoca tanto un ahorro económico al agricultor en tratamientos de control de plagas o de consumo de agua como un ahorro de tiempo, ya que hace necesaria su presencia en los campos de cultivo de forma constante.

Aplicaciones domóticas

Las redes inalámbricas utilizadas en los edificios, oficinas u hogares permiten una reducción de costes al permitir un monitoreo de la temperatura, la luz y la humedad de las diferentes zonas del recinto. Esto hace que se pueda realizar actuaciones de control que regulen los sistemas de calefacción o del aire acondicionado y de la intensidad lumínica de la habitación.

Estos sensores son fáciles de instalar al no necesitar del cableado tradicional convirtiendo la zona en lo comúnmente llamado hogar inteligente. En este ámbito de aplicación es en el que se centrará este proyecto.

Aplicaciones industriales

En el campo industrial, las WSN adquieren gran utilidad debido a su bajo coste, su tamaño reducido y a ser inalámbricas. Estos factores permiten que la cantidad de nodos y, por tanto, de sensores pueda ser mayor y más accesible para realizar un monitoreo del estado de las maquinarias. Al no ser necesario el uso de cableado como en las tecnologías tradicionales, se pueden introducir un número mayor a un menor coste y por su pequeño tamaño pueden ubicarse en zonas de la maquinaria con difícil acceso que no impidan el buen funcionamiento de esta. Gracias a ellas se evitan o conocen con mayor rapidez fallos en los equipos que permiten ahorrar en mantenimiento y servicio.

Aplicaciones en estructuras

El uso en el campo ingenieril de las redes inalámbricas ubicadas en puentes o edificios permite realizar un control técnico de posibles problemas que puedan surgir en las propias estructuras. Así, con sensores que recojan los movimientos sísmicos de las estructuras, se puede conocer a qué tipo de vibraciones y condiciones atmosféricas se exponen estas estructuras y saber si suponen algún peligro para las personas al producirse vibraciones no recomendadas o incluso encontrar pequeñas grietas que puedan causar problemas.

Aplicaciones en el tráfico

El monitoreo de las carreteras mediante el uso de estas tecnologías permite conocer en tiempo real el estado de las carreteras, ya sea por afluencia de coches, tráfico o accidentes, sin necesidad de realizar cortes innecesarios del tráfico debido a tratarse de redes inalámbricas. Con los datos obtenidos por estos sensores, las agencias estatales de tráfico de cada país tienen la posibilidad de utilizarlos para poder descongestionar ciertas carreteras, realizar desvíos hacia otras cuando sea necesario o controlar los semáforos para evitar atascos.

3 MÉTODOS DE INTERPOLACIÓN ESPACIAL

Una vez estudiada la arquitectura software y hardware de las redes de sensores inalámbricas, el siguiente paso es conocer cómo se realiza el estudio de los datos obtenidos por estas redes inalámbricas. Al tratarse de datos obtenidos mediante sensores que se encuentran distribuidos en diferentes zonas del área de trabajo, será necesaria la utilización de algún método estadístico de interpolación que permitirá poder estimar cuál es el dato que se obtendría en una posición específica determinada.

Interpolar consiste en calcular el valor aproximado de una magnitud en un intervalo cuando se conocen algunos de los valores que toma a uno y otro lado de dicho intervalo. Cuando los valores conocidos se tratan de datos tomados de la naturaleza, ya sea temperatura, precipitación, contaminación de aire, contaminación de suelo, etc., se habla de una **interpolación espacial**. El objetivo al realizar la interpolación es la creación de un mapa de superficie continua de datos. Normalmente, el análisis espacial se realiza con un Sistema de Información Geográfica (SIG) o más conocido por sus siglas en inglés *GIS, Geographic Information System.*, que proporciona herramientas de análisis espacial para calcular estadísticas de las entidades y realizar actividades de geoprocésamiento.

Según los criterios utilizados para realizar la estimación de los datos, los métodos de interpolación se pueden clasificar de diferentes formas. Si el estudio se realiza en función de la cantidad de datos que intervienen en la estimación, estos pueden ser globales o locales. **Los métodos de interpolación globales** son aquellos métodos donde se utilizan todas las muestras de los datos registrados de la zona espacial para realizar las estimaciones. Por su parte, **los métodos de interpolación locales** son aquellos donde se utilizan simplemente la serie de puntos de muestreo más cercanas al punto de estimación. Entre los métodos de interpolación utilizados en este proyecto serán estos últimos, los métodos locales, en los que se basarán las estimaciones realizadas, los cuales son los más utilizados para las casuísticas planteadas en proyectos de este estilo.

Si el estudio se realiza en función de las técnicas de interpolación utilizadas, se pueden clasificar en métodos de interpolación deterministas o geoestadísticos. **Las técnicas de interpolación deterministas** crean superficies a partir de los datos obtenidos de los puntos donde se han realizado mediciones, basándose en la extensión de la similitud o en el grado de suavizado. Por otro lado, **las técnicas de interpolación geoestadísticas** generan superficies continuas a partir de las propiedades estadísticas de los datos de partida cuantificando la autocorrelación espacial entre los puntos medidos.

En este proyecto se realizarán interpolaciones mediante el uso de ambas técnicas de interpolación ya que, aunque los datos tengan origen geoestadístico, con datos obtenidos del ámbito ambiental y geográfico, las técnicas deterministas pueden también producir posibles resultados válidos, por lo que se estudiarán ambas opciones.

3.1 Métodos de Interpolación Deterministas

Estos métodos de interpolación crean superficies continuas simplemente haciendo uso de las observaciones puntuales obtenidas de la muestra. Por tanto, estos métodos son óptimos cuando la muestra de datos es suficiente como para obtener una estimación óptima de todo el terreno que se aborda. Pese a no tratarse de métodos especialmente dedicados para su uso en ámbitos geográficos, es posible la utilización de estos siempre y cuando, como se ha comentado, las muestras obtenidas sean suficientes. En lo referido a este proyecto, tal y como se explicará en el Capítulo 4, la cantidad de muestras obtenidas respecto a la superficie tratada permiten el uso de estos métodos de interpolación para poder realizar un estudio comparativo, por lo que alguno de ellos se tendrá en cuenta este proyecto.

3.1.1 Polígono de Voronoi

Es el método de interpolación más sencillo de entre todos los existentes. Consiste en la asignación del valor conocido más cercano al punto de estudio. Así, al realizar el mapa de continuidad, el terreno genera una serie de polígonos conocidos como *Polígonos de Voronoi*, que delimitan cada área y a cada uno de estos polígonos se le asigna en su totalidad el valor del punto a partir del cual se ha generado.

Este hecho hace que este método sea muy inexacto ya que los datos, aunque se hayan obtenido a partir del valor del más cercano, éste podría encontrarse relativamente alejado de algunos puntos.

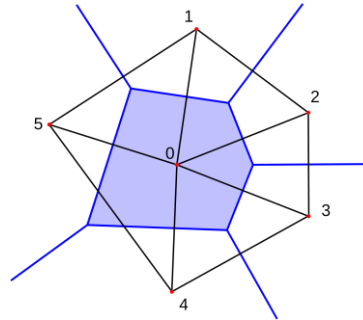


Figura 3-1. Polígono de Voronoi. [22]

3.1.2 Vecino Natural

La interpolación de Vecino Natural, también denominada interpolación de Sibson, parte de la interpolación del polígono de Voronoi. Este método se encarga de realizar nuevos polígonos de Voronoi superponiéndose sobre los ya inicialmente creados en función de la cercanía a estos, hallando el conjunto de muestras de entrada más cercano a un punto y aplicando ponderaciones. La proporción de superposición entre este nuevo polígono y los polígonos existentes nos facilita los pesos establecidos.

En la Figura 3-2, el valor del punto estrella se obtiene a partir de la siguiente fórmula:

$$x = \sum_{i=1}^n \varphi_i(x) \cdot x_i \tag{3-1}$$

Con x_i igual al valor en el polígono, y $\varphi_i(x)$ el peso ponderado de dicho polígono, siendo la suma de todos ellos igual a 1.

Igualmente, el peso ponderado se obtiene:

$$\varphi_i(x) = \frac{A_i(x)}{A(x)} \tag{3-2}$$

Con $A(x)$ siendo el área de polígono de Voronoi inicial y siendo $A_i(x)$ el Área del nuevo polígono que forma parte de la inicial.

Este método, aunque sea más preciso que el anterior, sigue siendo inexacto ya que parte del Polígono de Voronoi y, por tanto, iniciándose a partir de los datos absolutos de los puntos [23].

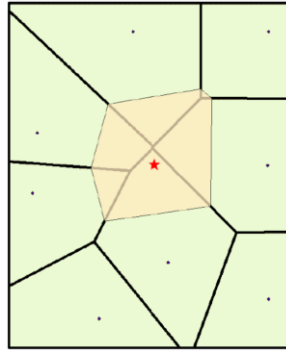


Figura 3-2. Método de interpolación Vecino Natural.

3.1.3 Spline

Este método estima valores usando una función matemática que minimiza la curvatura general de la superficie, realizando una curva con suavizado que pasa por los puntos conocidos. Spline [24] utiliza funciones polinómicas por tramos en lugar de utilizar un único polinomio. Esto permite un mayor ajuste en cada uno de los tramos y una curva más exacta y suave en cada uno de ellos.

Existen dos tipos diferentes de Spline; regularizado y de tensión. El primero de ellos crea una superficie suave en la que no tiene en cuenta los valores máximos y mínimos de la muestra, por lo que el gradiente de valores obtenidos puede estar fuera de este rango; por su parte, el de tensión crea superficies más rígidas que el tipo regularizado pero su rango de valores no se encuentra nunca fuera de la muestra.

La ecuación de Spline es la siguiente:

$$S(x, y) = T(x, y) + \sum_{i=1}^N \lambda_i \cdot R(r_i) \quad (3-3)$$

siendo $S(x, y)$ la solución al sistema, N la cantidad de datos de la muestra, λ_i son los coeficientes obtenidos del sistema de ecuaciones, r_i es la distancia del punto (x, y) a la muestra, y $T(x, y)$ y $R(r_i)$ son valores que dependen del tipo de Spline obtenido.

Cuando la opción utilizada es la regularizada, $T(x, y)$ y $R(r_i)$ tienen la siguiente forma:

$$T(x, y) = a_1 + a_2x + a_3 \cdot y \quad (3-4)$$

$$R(r_i) = \frac{1}{2 \cdot \pi} \cdot \left\{ \frac{r^2}{4} \cdot \left[\ln \left(\frac{r}{2 \cdot \tau} \right) + c - 1 \right] + \tau^2 \cdot \left[\frac{K_0 \cdot r}{\tau} + c + \ln \left(\frac{r}{2 \cdot \pi} \right) \right] \right\} \quad (3-5)$$

Cuando la opción utilizada es la de tensión, $T(x, y)$ y $R(r_i)$ tienen la siguiente forma:

$$T(x, y) = a_1 \quad (3-6)$$

$$R(x, y) = \frac{1}{2 \cdot \pi \cdot \tau^2} \cdot \left[\ln \left(\frac{\tau \cdot r}{2} \right) + c + K_0 \cdot (\tau \cdot r) \right] \quad (3-7)$$

En la primera ecuación, a_i siendo los valores de los coeficientes hallados de un sistema de ecuaciones lineales. En la segunda, r es la distancia entre el punto y la muestra, K_0 es la función Bessel modificada, c es una constante igual a 0,577215. La variable τ^2 es el parámetro *Peso* y varía según el modelo. Para el regularizado los valores más comunes son 0; 0,001; 0,01; 0,1 y 0,5. A mayor valor, más suavidad. Para el de tensión los valores más comunes son 0, 1, 5 y 10. A mayor valor, más grosor en la superficie.

3.1.4 IDW

El método de interpolación de la Distancia Inversa Ponderada [25], más conocido por las siglas en inglés *IDW* (*Inverse Distance Weighting*), es aquel que, como su propio nombre indica, utiliza la asignación de pesos a los datos de entorno en función de la inversa de la distancia que los separa. Según las características del método, éste asume que los datos medidos están relacionados entre sí, siendo los más cercanos al punto que se desea conocer más importante que los más lejanos, al igual que ocurría con el método del vecino natural.

Este tipo de interpolación se suele utilizar cuando la muestra conocida es grande ya que, si la muestra no es lo suficientemente importante o los datos se encuentran demasiado concentrados, la estimación obtenida podría no ser correcta.

Este método tiene dos formas de uso según los criterios que se especifiquen. Si el criterio utilizado es un radio específico, el número de muestras que abarca cada punto que se desea conocer será variable, ya que según la cantidad de datos que abarque dicho radio de búsqueda, el cual siempre será fijo, habrá más o menos muestras. Si el criterio utilizado es un número fijo de muestras más cercanas, entonces el que variará será el radio de búsqueda específico.

La expresión utilizada para este método es:

$$Z(S_o) = \sum_{i=1}^N \lambda_i \cdot Z(S_i) \quad (3-8)$$

donde $Z(S_o)$ es el valor que deseamos conocer, N es el número de muestras alrededor del punto a predecir, $Z(S_i)$ son los valores medidos y λ_i son los pesos asignados, cuya suma de todos ellos es igual a 1.

Cada peso se obtiene a partir de la siguiente ecuación:

$$\lambda_i = \frac{d_{io}^{-p}}{\sum_{i=1}^N d_{io}^{-p}} \quad (3-9)$$

donde d_{io} es la distancia entre el lugar de predicción S_o y el lugar muestral S_i , p es el exponente de ponderación que controla cómo el peso disminuye con la distancia [26].

3.2 Métodos de Interpolación Geoestadísticos

Por Geoestadística se entiende a la rama de la estadística que realiza el estudio de los datos geográficos que se aplica para obtener datos precisos de cálculos en lugares donde no se tiene información, es decir, tiene como objetivo interpretar el comportamiento de los datos que están distribuidos espacialmente.

El uso de la Geoestadística surge en los yacimientos mineros de oro en Sudáfrica, donde el ingeniero de minas Daniel G. Krige y acompañado por Herbert Sichel, pusieron en práctica un método de predicción de yacimientos de oro a partir de los datos obtenidos de los yacimientos existentes durante la década de 1950.

Posteriormente fue Georges Matheron el que, en el Centro de Investigación de Fontainebleau, creado para él por la escuela de Minas de París, siguió avanzando en los métodos implementados por Krige, llegando a desarrollar nuevas teorías basadas en las variables regionalizadas. A estas nuevas teorías de estimación Matheron las denominó *Kriging*, en honor a Daniel G. Krige. En la actualidad, los llamados métodos de interpolación geoestadísticos son variantes del método Kriging, el cual se explica en el Apartado 3.2.3.

Los métodos geoestadísticos, para obtener los mejores resultados, deben pasar por cuatro etapas.

3.2.1 Estudio exploratorio de los datos

Esta primera etapa consiste en conocer de forma general la distribución de la variable regionalizada para definir zonas de estudio, detectar errores, etc. Algunas herramientas de análisis exploratorio de datos, presentadas con sus respectivos objetivos son [28]:

- Los histogramas, que sirven para conocer la distribución estadística de los datos y así corroborar la existencia de datos erróneos o dispersos.
- Los Gráficos cuantil-cuantil, más conocidos comúnmente como gráficos q-q, que permite comprobar las diferencias entre la distribución de probabilidad de los datos y una distribución usada para la comparación.
- Las nubes de correlación, que permiten visualizar los pares de datos obtenidos en un gráfico.

3.2.2 Análisis variográfico

El variograma es una función matemática que permite comprobar el comportamiento de una variable en una zona específica, midiendo la relación existente entre los diferentes valores ambientales medidos en el terreno.

Los datos de las variables ambientales obtenidas se rigen por dos componentes: una determinista, que se deben al simple hecho de obedecer a las leyes de la naturaleza; y otra aleatoria, que se produce al juntarse varios procesos que conforman el entorno, el cual puede provocar evoluciones y resultados tan complejos que hacen que se tomen como variables aleatorias. Este último no se puede representar en sí como una función ya que se deben a procesos aleatorios, pero al tratarse de datos ambientales que suceden en espacios relativamente cercanos, se espera que estos datos sean similares, estando así los valores obtenidos relacionados.

La variable aleatoria se puede representar como:

$$Z(x) = m(x) + \varepsilon(x) \quad (3-10)$$

Siendo $m(x)$ la media y $\varepsilon(x)$ una cantidad aleatoria con una media de cero.

Los parámetros que intervienen en los procesos aleatorios son en general la media, la varianza, la covarianza y el variograma teórico.

$$m(x) = E[Z(x)] \quad (3-11)$$

$$\text{var}[Z(x)] = E\{[Z(x) - m(x)]^2\} \quad (3-12)$$

$$C(x, x + h) = \text{Cov}[Z(x), Z(x + h)] = E\{[Z(x) - m(x)] \cdot [Z(x + h) - m(x + h)]\} \\ = E \cdot [Z(x)] \cdot [Z(x + h)] - m(x) \cdot m(x + h) \quad (3-13)$$

$$2 \cdot \gamma(x, x + h) = \text{var}[Z(x) - Z(x + h)] \quad (3-14)$$

siendo h la distancia entre el punto a medir y el punto de referencia.

En estos supuestos, se entiende también que existe estacionariedad en los datos, es decir, que no dependen ni del tiempo ni de su posición absoluta, sino de la posición relativa entre ellos. Al suponerse estacionariedad, se tiene que la media y la varianza son constantes y que la covarianza simplemente depende de la separación entre los puntos a considerar, por lo que:

$$m = E[Z(x)] \quad (3-15)$$

$$\sigma^2 = \text{var}[Z(x)] \quad (3-16)$$

$$C(h) = \text{Cov}[Z(x), Z(x + h)] \quad (3-17)$$

$$\gamma(h) = \frac{1}{2} \cdot \text{var}[Z(x) - Z(x + h)] = \frac{1}{2} \cdot E[Z(x) - Z(x + h)]^2 = \sigma^2 - C(h) = C(0) - C(h) \quad (3-18)$$

Por tanto, se puede entender el variograma¹ teórico como el valor promedio de la diferencia al cuadrado de los valores en dos puntos separados por una distancia h .

Por su parte, de la representación de los valores del variograma frente a la distancia de los pares de puntos para

¹ Realmente hace referencia al semivariograma, pero por comodidad y convenio se le denomina simplemente variograma

una serie de datos para todas las parejas de datos, obtenemos el variograma experimental. La fórmula utilizada en el variograma experimental es la siguiente:

$$\gamma^*(h) = \frac{1}{2 \cdot N(h)} \sum_{N(h)} [z(x_i) - z(x_j)] \quad (3-19)$$

donde $N(h)$ es el número de pares de puntos (x_i y x_j) separados por una distancia h [29].

La representación de variograma se realizaría con valores de distancia h , $2h$, $3h$... en situaciones donde los medidores se encuentren a distancias fijas y equidistantes unos de otros, formando una malla regular, lo cual permitiría realizar el gráfico utilizando estas distancias constantes. Sin embargo, la realidad es que esto no suele producirse. Lo habitual es encontrar un muestreo en una malla irregular, es decir, que los valores se encuentren dispersos unos respecto de los otros y no se encuentren los valores en las distancias y/o direcciones asignadas.

Para estos casos, al realizar la representación del variograma, se puede que tener en cuenta simplemente la distancia o una combinación de esta con la dirección.

- Si solo únicamente se tiene en cuenta la distancia, se asigna una tolerancia Δh a cada distancia h , la cual es del 50% del valor de h . Así, no se superponen los valores.
- Por el contrario, si se realiza un variograma combinado, se asigna una tolerancia Δh a cada distancia h y otra tolerancia $\Delta \theta$ a cada dirección θ , lo que haría que el variograma fuera más preciso

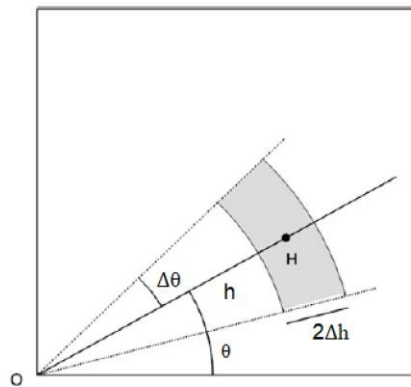


Figura 3-3. Cálculo del variograma según distancia y ángulo.

A la hora de escoger los criterios para representar el variograma primero se realiza un estudio para conocer cuál es el más acorde para cada estudio. En primer lugar, es necesario conocer la isotropía o anisotropía de los valores, es decir, la dependencia existente en las posibles variaciones de estos datos según su dirección.

Si al realizar un estudio visual se aprecia que los valores no tienen una dependencia clara con la dirección de estudio se dice que el variograma tiene un comportamiento isotrópico, donde las variaciones de los valores son igual en todas las direcciones y, por tanto, se realizaría un variograma omnidireccional. En realidad, este variograma omnidireccional realmente tiene un $\Delta\theta=90^\circ$, lo que hace que abarque todo el espectro, y por tanto tan solo se tendría en cuenta la distancia. En este tipo de estudios como el planteado en este proyecto, suele ser el más comúnmente utilizado, ya que factores como la temperatura, humedad, etc., no suelen tener una dependencia de la dirección.

Si, por el contrario, al realizar el estudio observamos que es posible que se produzca esta situación, lo ideal es realizar varios variograma para comprobar cómo se comportan los datos en función de la dirección. Normalmente, se suele utilizar los ángulos de 0° , 45° , 90° y 135° , es decir, los ejes E-O, NE-SO, N-S, NO-SE, respectivamente, utilizando una tolerancia de $22,5^\circ$ para abarcar todo el campo [30].

Una vez realizado el variograma experimental, se realiza un modelo teórico que realiza un ajuste idealizado de entre los diferentes existentes. Para realizar este modelo teórico se tienen en cuenta tres características principales que se pueden obtener de lo apreciado en el variograma experimental:

- La **Meseta o Sill**. Es el máximo valor alcanzado por el variograma y, a partir de él, el variograma se hace constante. Esta meseta puede ser finita o no. En el momento en que la gráfica alcanza este valor máximo y se mantiene (por lo que sería finita) significará que los datos dejan de estar correlacionados espacialmente
- El **Rango o Range**. Es la conocida como zona de influencia. Se define como la distancia a la cual el variograma se hace estable, es decir, es el momento en el que se alcanza la meseta. Los valores obtenidos antes de alcanzar el rango se encuentran correlacionadas espacialmente, mientras que aquellos que se encuentran tras él no lo están y, por tanto, pueden llegar a despreciarse.
- El **Efecto Pepita o Nugget**. Es la diferencia del variograma con respecto de 0 en el origen. Teóricamente, este valor debería ser cero, ya que la variación de un valor respecto a si mismo es nula; pero debido a posibles errores de estimación, por ajuste o por diferentes variables aleatorias es probable que no lo sea. También puede ser un indicador de que las varianzas espaciales se encuentren a distancias menores de las utilizadas en el variograma experimental.

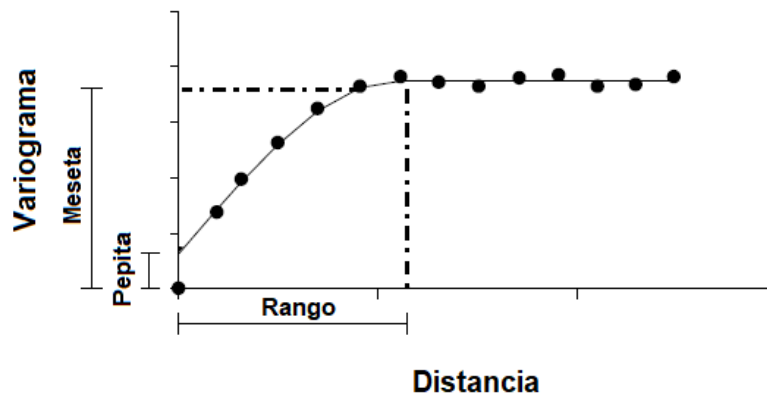


Figura 3-4. Variograma Experimental.

En función de las diferentes características descritas, existen una serie de modelos teóricos admisibles según el comportamiento encontrado en el variograma experimental, que nos permite realizar una aproximación. En las ecuaciones que se plantean a continuación, el efecto pepita está representado por b , la meseta por C y el rango por a . Así, con esto, los modelos más comunes son los siguientes [31]:

1. Modelo Efecto Pepita

Es el modelo más simple. En él, se considera que el variograma presenta una ausencia de correlación espacial de todos los valores estudiados, independientemente de la distancia que los separe.

$$y(h) = \begin{cases} 0 & \text{si } h = 0 \\ C & \text{en caso contrario} \end{cases} \quad (3-20)$$

2. Modelo Esférico

En este modelo se aprecia un comportamiento lineal a distancias de separación pequeñas cerca del origen, es decir, donde los valores presentan una correlación espacial. A medida que se alcanza la meseta, la curva se va aplanando.

$$y(h) = \begin{cases} b + C \cdot \left(\frac{3 \cdot h}{2 \cdot a} - \frac{h^3}{2 \cdot a^3} \right) & \text{si } h < a \\ b + C & \text{si } h \geq a \end{cases} \quad (3-21)$$

3. Modelo Exponencial

Este modelo tiende a alcanzar la meseta de forma asintótica. Al comienzo tiene un comportamiento aparentemente lineal. Al tratarse de un modelo exponencial no se tiene un rango finito, pero se suele definir como $3a$, que es aproximadamente el valor alcanzado del 95% de la meseta.

$$y(h) = b + C \left(1 - e^{-\frac{h}{a}} \right) \quad (3-22)$$

4. Modelo Gaussiano

Al igual que el modelo exponencial, alcanza la meseta de forma asintótica. Tiene la misma estructura que el exponencial, pero a diferencia de este el rango se define en $\sqrt{3}a$ y su comportamiento en el origen es cuadrático.

$$y(h) = b + C \left(1 - e^{-\frac{3 \cdot h^2}{a^2}} \right) \quad (3-23)$$

5. Modelo potencial

Este modelo no presenta ni meseta ni rango, crece de forma indefinida. El valor del exponente varía entre 0 y 2. Cuando el exponente es igual a 1, se le conoce como **modelo lineal**; mientras que cuando es igual a 0 correspondería al **modelo efecto pepita**. Es un modelo acorde para fenómenos no estacionarios, es decir, que varían en función del tiempo.

$$y(h) = b + C \cdot h^\alpha \quad (3-24)$$

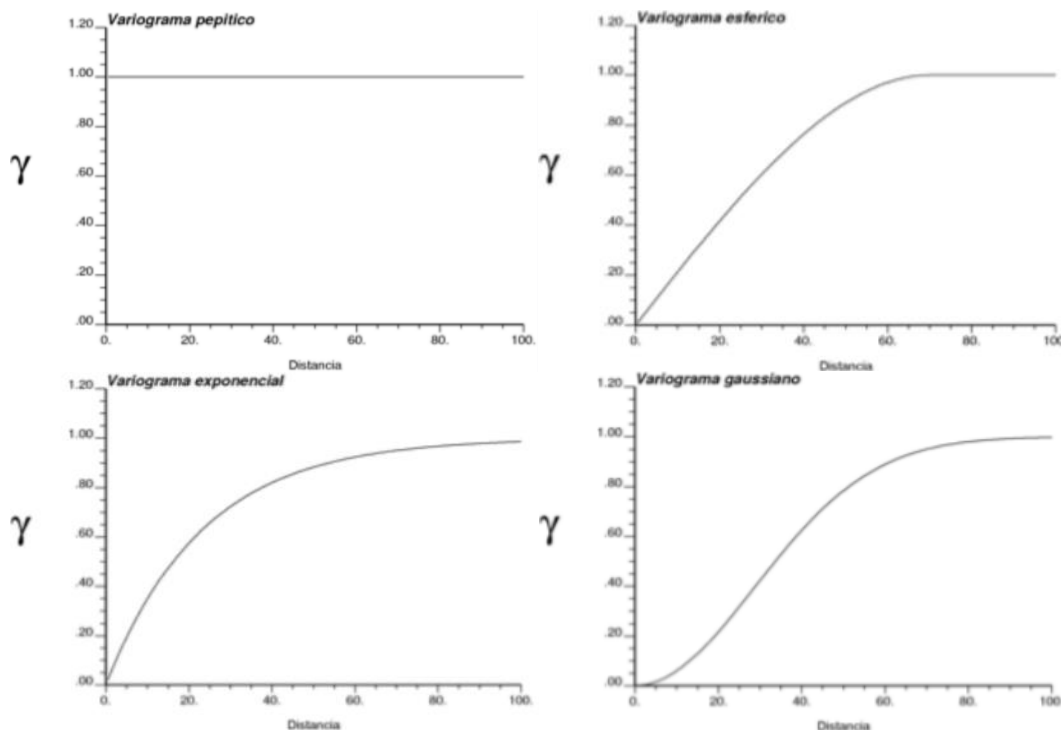


Figura 3-5. Modelos de Variograma Teóricos [28]

3.2.3 Kriging o Krigeaje.

Tras la obtención del variograma se realiza Método Kriging. Es un método geoestadístico de interpolación de estimación de puntos, que proporciona el mejor estimador lineal insesgado de los datos a estudiar basándose en el modelo de variograma seleccionado para la adquisición de datos. Kriging calcula los pesos que debe tener cada punto de referencia utilizado, de ahí el término *lineal*, y estima que el valor del error medio residual es cero, lo que lo hace *insesgado*. Es un método que presupone que la variación espacial de los datos es mantenida respecto a un patrón determinado.

Kriging supone que se dispone de n de valores observados a partir de los cuales se desea obtener el estimador lineal para un punto $Z^*(x_0)$ específico, lo cual se representa con la siguiente fórmula:

$$Z^*(x_0) = \sum_{i=1}^n \lambda_i \cdot z(x_i) \quad (3-25)$$

donde λ_i son los pesos otorgados a cada uno de los valores observados $z(x_i)$.

Respecto a la insesgación del método, hace referencia a que la esperanza sea nula, es decir, el promedio entre los valores estimados y los reales es cero. Esto se representa de la siguiente forma:

$$E[Z^*(x_0) - Z(x_0)] = 0 \quad (3-26)$$

Otra de las condiciones existentes en el método de Kriging es que la varianza del error del estimador debe ser mínima, lo cual suele conseguirse con la restricción de insesgidez:

$$var[Z^*(x_0) - Z(x_0)] \text{ es mínima}$$

Existen diferentes modelos de Kriging, clasificándose entre lineales y no lineales. Los más importantes son los modelos lineales, de entre los que destacan el Kriging Ordinario (*OK*), el Kriging Simple (*SK*) y el Kriging Universal (*UK*).

3.2.3.1 Kriging Ordinario

El Kriging ordinario, conocido por sus siglas inglesas *OK*, es el más general y utilizado de los diferentes métodos. Utiliza promedios ponderados para predecir puntos vecinos, proponiendo que el valor de la variable se puede predecir como una combinación lineal de las n variables aleatorias.

En este modelo de Kriging, se asume que el valor medio constante es desconocido por lo que, para que no exista sesgo, la suma de estos pesos debe ser igual a 1.

$$E[Z(x_0)] = m \quad (3-27)$$

$$E[Z^*(x_0) - Z(x_0)] = \sum_{i=1}^n \lambda_i \cdot m - m \quad \rightarrow \quad \sum_{i=1}^n \lambda_i = 1 \quad (3-28)$$

Por su parte, la optimalidad de la varianza mínima se produce según la siguiente forma:

$$var[Z^*(x_0) - Z(x_0)] = var \left[\sum_{i=1}^n \lambda_i \cdot Z(x_i) - Z(x_0) \right] \quad (3-29)$$

Generalizando, se obtiene:

$$var \left(\sum_{i=1}^N \lambda_i \cdot Z_i \right) = \sum_{i=1}^N \sum_{j=1}^N \lambda_i \cdot \lambda_j \cdot cov(Z_i, Z_j) \quad (3-30)$$

Sustituyendo y realizando una serie de cálculos, finalmente se llega a la siguiente expresión:

$$var[Z^*(x_0) - Z(x_0)] = \sum_{i=1}^N \sum_{j=1}^N \lambda_i \cdot \lambda_j \cdot C(x_{ij}) + C(0) - 2 \cdot \sum_{i=1}^n \lambda_i \cdot C(x_{i0}) \quad (3-31)$$

Para poder minimizar la expresión es necesario primero cumplir con la condición de insesgo, es decir, la suma de los coeficientes debe ser 1. Para lograrlo es necesario la inclusión del multiplicador de Lagrange, μ .

$$var[Z^*(x_0) - Z(x_0)] = \sum_{i=1}^N \sum_{j=1}^N \lambda_i \lambda_j C(x_{ij}) + C(0) - 2 \sum_{i=1}^n \lambda_i C(x_{i0}) + 2\mu \left(\sum_{i=1}^n \lambda_i - 1 \right) \quad (3-32)$$

Minimizando esta expresión se obtiene:

$$\sum_{j=1}^n \lambda_j \cdot C(x_{ij}) + \mu = C(x_{i0}) \quad (3-33)$$

O escrito de forma matricial:

$$\begin{bmatrix} C_{11} & \dots & C_{1n} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ C_{n1} & \dots & C_{nn} & 1 \\ 1 & \dots & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ \mu \end{bmatrix} = \begin{bmatrix} C_{10} \\ \vdots \\ C_{n0} \\ 1 \end{bmatrix} \quad (3-34)$$

Como el variograma es equivalente a la covarianza mediante la relación $\gamma(h) = C(0) - C(h)$, tal y como se vio en el Apartado 3.2.2., se puede reescribir de la siguiente forma [32]:

$$\begin{bmatrix} \gamma_{11} & \dots & \gamma_{1n} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma_{n1} & \dots & \gamma_{nn} & 1 \\ 1 & \dots & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ -\mu \end{bmatrix} = \begin{bmatrix} \gamma_{10} \\ \vdots \\ \gamma_{n0} \\ 1 \end{bmatrix} \quad (3-35)$$

3.2.3.2 Kriging Simple

El Kriging Simple, *SK*, de manera análoga a como se define en modelos lineales, el modelo establecido en este caso es igual a la media más un error aleatorio con media cero. La diferencia es que, en este caso, los errores no son independientes.

En este modelo, se produce la estimación de que tanto la media como la covarianza son conocidas.

$$E[Z(x_0)] = m \quad (3-36)$$

$$Cov[Z(x_0), Z(x)] = C(h) \quad (3-37)$$

En este caso modelo, según las condiciones anteriores, el estimador lineal adquiere la siguiente forma:

$$Z^*(x) - m = \sum_{i=1}^n \lambda_i \cdot [z(x_i) - m] \quad (3-38)$$

donde m es la media de valor conocido, y λ_i son los coeficientes de ponderación.

Y, la condición, de insesgado, se da con el siguiente valor esperado del error de estimación:

$$E[Z^*(x_0) - Z(x_0)] = m - \lambda_0 - \sum_{i=1}^n \lambda_i \cdot m = 0 \quad (3-39)$$

Por su parte, la optimalidad de la varianza mínima tiene la misma forma que el Kriging Ordinario, esto es:

$$var[Z^*(x_0) - Z(x_0)] = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \cdot \lambda_j \cdot C(x_{ij}) + C(0) - 2 \cdot \sum_{i=1}^n \lambda_i \cdot C(x_{i0}) \quad (3-40)$$

El mínimo de esta expresión se consigue anulando las derivadas parciales con respecto a las incógnitas λ_i , por lo que finalmente se obtiene el siguiente sistema de ecuaciones:

$$\sum_{j=1}^n \lambda_j \cdot C(x_{ij}) = C(x_{i0}) \quad (3-41)$$

O lo que es lo mismo, escrito de forma matricial:

$$\begin{bmatrix} C_{11} & \dots & C_{1n} \\ \vdots & \ddots & \vdots \\ C_{n1} & \dots & C_{nn} \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{bmatrix} = \begin{bmatrix} C_{10} \\ \vdots \\ C_{n0} \end{bmatrix} \quad (3-42)$$

con lo que se determinan los valores de los coeficientes de Kriging λ_i [33].

3.2.3.3 Kriging Universal

El Kriging Universal, *UK*, es una variante del Kriging Ordinario, pero donde no se presenta un carácter estacionario. En este método, se produce la combinación de funciones, una formada por una variable determinista y otra por una componente estocástica estacionaria de media cero. Esto es:

$$Z(x) = m(x) + \varepsilon(x) \quad (3-43)$$

$$E(\varepsilon(x)) = 0 \quad (3-44)$$

$$E(Z(x)) = m(x) \quad (3-45)$$

Siendo $m(x)$ la función determinística que describe la tendencia y $\varepsilon(x)$ la componente estocástica.

La función determinística se puede expresar como:

$$m(x) = \sum_{i=1}^K a_i \cdot f_i(x) \quad (3-46)$$

donde a_i son coeficientes estimados de los datos, f_i es la función que describe la tendencia, y K es el número de funciones utilizadas en la deriva. Así pueden obtenerse derivas simples, lineales, cuadráticas, etc. [34] [35].

El estimador de Kriging Universal tiene la misma forma que el Kriging Ordinario, pero éste será insesgado si se cumple las siguientes condiciones:

$$\sum_{i=1}^N \lambda_i \cdot f_i(x_i) = f_i(x_0) \quad (3-47)$$

Por su parte, la optimalidad en la varianza se consigue con el siguiente criterio:

$$\sum_{j=1}^N \lambda_j \cdot \gamma_{ij} + \sum_{l=1}^K \mu_l \cdot f_l(x_i) = \gamma_{i0} \quad \text{con } i = 1, \dots, N \quad (3-48)$$

Que escrito de forma matricial sería tal que:

$$\begin{bmatrix} \gamma_{11} & \dots & \gamma_{1n} & f_{11} & \dots & f_{1n} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \gamma_{n1} & \dots & \gamma_{nn} & f_{1n} & \dots & f_{pn} \\ f_{11} & \dots & f_{1n} & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ f_{p1} & \dots & f_{pn} & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ \mu_1 \\ \vdots \\ \mu_n \end{bmatrix} = \begin{bmatrix} \gamma_{10} \\ \vdots \\ \gamma_{n0} \\ f_{10} \\ \vdots \\ f_{p0} \end{bmatrix} \quad (3-49)$$

3.2.4 Validación Cruzada

Una vez escogido y realizado el método de Kriging más oportuno, es necesario realizar una validación de los modelos para comprobar si todos los pasos anteriores han sido correctamente elegidos, es decir, validar la vecindad escogida para el modelo de variograma, validar el modelo de variograma escogido y validar el modelo de Kriging escogido [30].

Para conseguir todo ello, se opta por descartar uno de los valores observados conocidos y calcular el nuevo variograma sin ese dato. Una vez calculado se estima por Kriging el valor en ese punto eliminado a partir de ese variograma. Como se trata de un valor conocido, es posible comparar el valor obtenido con el valor real y comprobar si es óptimo todo lo calculado inicialmente. Si esto se realiza con cada uno de los valores conocidos se puede obtener una estimación bastante ajustada donde poder observar la robustez del modelo, es decir, comprobar que es óptimo en todas las zonas de estudio y estimación.

4 RECREACIÓN DE LA RED DE SENSORES

Una vez realizado el estudio teórico de cómo funcionan las WSN y cuáles son los métodos de interpolación más utilizados para el tratamiento de los datos obtenidos de la red de sensores y sus estimaciones en zonas sin información, el siguiente paso es pasar a la práctica. El hecho de que se realice una representación de esta se debe a que, aunque originariamente el proyecto estaba pensado para realizarse con motas reales, no ha sido posible su utilización debido a que éstas o estaban deterioradas o sufrían una falta de stock que hacía imposible poder recrear un escenario acorde a las condiciones necesarias para poder realizar estimaciones de temperatura, ya fuese con uno o con varios métodos de interpolación.

Así pues, este capítulo está destinado al proceso de recreación de cómo habrían actuado los nodos dentro de una red de sensores inalámbricos. Este proceso comprende la obtención de los datos, la ubicación de los nodos, la topología de red empleada o los diferentes modos de funcionamiento que puede tener la red creada en la simulación, entre otros. Toda la demostración e interfaz gráfica se ha realizado con el software informático Matlab [36], el cual permite recrear mediante la herramienta GUIDE (Graphical User Interface Development Environment) incluida en el paquete de Matlab, el caso práctico de este proyecto consistente en una sala de laboratorio sobre la cual se reparten los diferentes nodos-sensores disponibles.

4.1 Origen de los datos

Como se ha comentado en la introducción del capítulo, no ha sido posible realizar un muestreo de datos con motas in situ por el usuario, por lo que ha sido necesario buscar una base de datos real que se adaptara a las necesidades del proyecto. Ya que el objetivo del proyecto es realizar un estudio de la distribución espacial de la temperatura, era necesario que dicha base de datos contuviera información sobre la temperatura que nos permitiera recrear una situación como la expuesta.

Al tratarse del interior de una sala de un edificio, es decir, un espacio cerrado, ha sido necesario buscar una situación parecida a esta. Esta situación nos la ofrece el laboratorio de investigación de Intel Berkeley [37], quien, implantando un total de 54 nodos sensores dentro del mismo, realizó el registro de la temperatura, humedad, luz y voltaje en periodos de 31 segundos de manera constante entre el 28 de febrero de 2004 y el 5 de abril del mismo año.

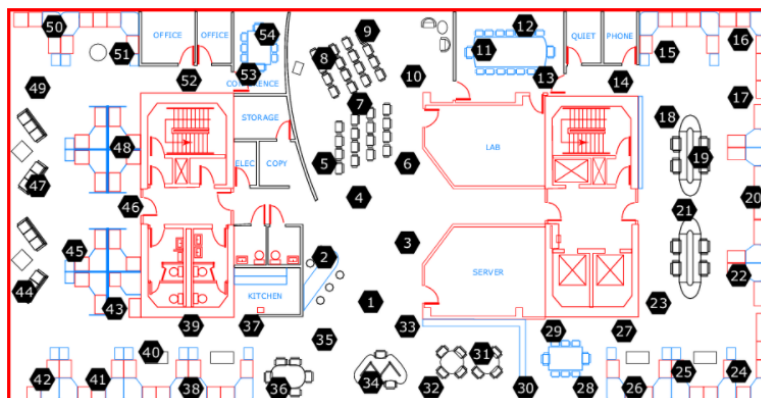


Figura 4-1. Esquema de la sala de laboratorio de Intel Berkeley.

En total, entre el periodo que comprende los datos y todos los nodos-sensores distribuidos, se almacenan alrededor de 2.3 millones de datos. Esta cantidad de datos es muy elevada, incluso para un archivo Excel, los

cuales pueden llegar a almacenar poco más de 1 millón de filas. Además de este inconveniente, tras realizar un estudio previo de los datos de los que se disponen, se puede observar que existen muchos valores que podrían provocar errores en las estimaciones debido a que los nodos, al ir perdiendo voltaje, comenzaron a realizar tomas inexactas. Para evitar los datos erróneos se han eliminado aquellos cuyas temperaturas eran poco acordes a la realidad, lo que ha hecho que se reduzca el periodo de días en los que se recoge información. Al realizar el proyecto, se utilizan como datos aquellos recogidos entre el 28 de febrero y el 17 de abril, es decir, un período de 19 días (2004 fue un año bisiesto), siendo estos los días donde las baterías estaban más cargadas y, por tanto, no producían errores groseros en la toma de información, especialmente en la toma de temperatura. Con estas reducciones, se pasa de un total de 2.3 millones a aproximadamente 1.5 millones de tomas.

Teniendo en cuenta todas las variaciones realizadas, finalmente se ha almacenado toda la información de los nodos en dos archivos diferentes, ya que la totalidad de los datos era superior al límite máximo disponible en Excel: el primero incluye la información de los nodos desde el día 28 de febrero de 2004 al 9 de marzo de 2004, ambos inclusive; mientras que el segundo, contiene la información recogida desde el día 10 de marzo hasta el día 17 de marzo, ambos inclusive.

Respecto a las dimensiones de la sala y a la distribución de los nodos, además de los datos recogidos, el laboratorio recoge las ubicaciones exactas de los nodos dentro de la sala. Como se puede observar en la Figura 4-1, la sala de laboratorio está dividida en dos superficies separadas por un pasillo central. Los nodos-sensores se encuentran distribuidos en los exteriores de estas dos superficies y en el pasillo, dejando las dos zonas centrales, formada una de ellas por el laboratorio y los servidores y la otra por la zona de cocina y almacenaje, sin tomas de medidas. Además, la sala completa del laboratorio tiene unas dimensiones aproximadas de 42x32m. En lo que se refiere a la simulación realizada, para evitar alteraciones en las estimaciones que se vayan a realizar, se ha decidido mantener el mismo esquema de la sala, aunque haya zonas amplias donde no existan nodos-sensores. Igualmente, se ha utilizado la misma medida de la sala para ser lo más fiel a la realidad. Además de esto, para realizar una recreación de una WSN más completa, se ha incluido un *Nodo Sink* o puerta de enlace, representado en ella con un cuadrado amarillo, y un Servidor en el equivalente al área del servidor en la realidad el cual, en la representación, tiene una forma geométrica de un rombo verde.

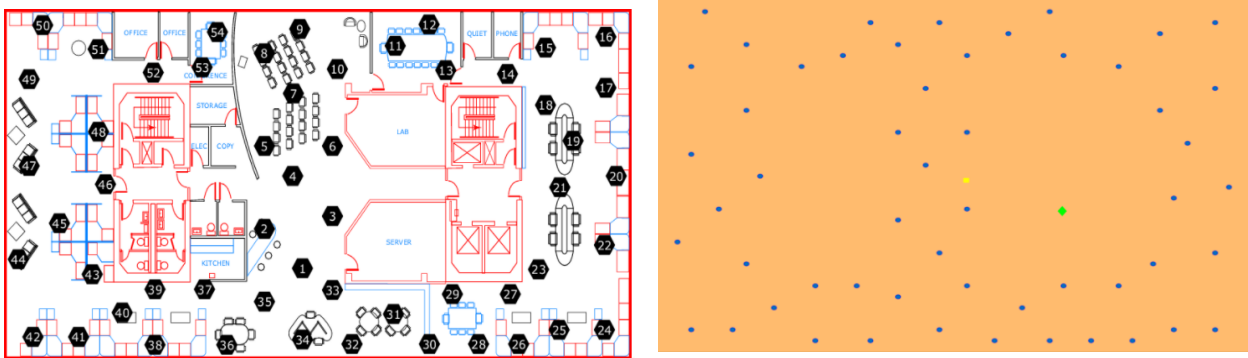


Figura 4-2. Comparativa del laboratorio y la recreación de la red de sensores.

Aunque teóricamente los nodos-sensores deberían registrar los datos cada 31 segundos, ninguno de ellos lo hacen siempre en ese periodo ni todas las veces, por lo que cada vez que se realice una toma de registros, realmente se escogerá el registro de cada uno de los nodos más próximo en el tiempo a la hora en la que nos encontremos en ese momento en la representación. Como en la primera toma registrada en la base de datos no todos los nodos realizan dicha toma, se ha editado la base de datos de manera que los datos registrados en los momentos más cercanos al minuto en el que el último de los nodos realiza su primera toma, sea el primer dato de cada uno de los nodos. Esto es, tras observar el primer registro de cada uno de los nodos, es el nodo 49 el que recoge por primera vez información de su ubicación de forma más tardía, produciéndose a la 01:18:47 del día 28 de febrero. Por su parte, por ejemplo, el nodo 1 hasta ese instante había realizado un total de 18 registros, siendo el último de ellos a las 01:18:16, por lo que se han excluido todas las tomas anteriores y se ha asignado como primera toma esta última y asignándole la misma hora que la del nodo 49. Así se ha realizado con todos y cada uno de los nodos.

4.2 Funciones la recreación de la WSN

Como se ya ha comentado, la red de sensores simulada consta de 54 nodos repartidos por la sala del laboratorio de 42x32m, tal y como se realizó en la distribución original. En una posición aleatoria en un radio máximo de 5 m del centro de la sala, se ha ubicado el nodo Sink o Gateway, que será el encargado de transmitir las peticiones a los nodos y de recolectar la información recibida de estos para enviarlas al servidor. Este servidor estará ubicado en la zona reservada por el propio laboratorio original para los servidores haciendo a la simulación lo más fiel posible a la realidad.

Debido a que la red de sensores no es muy extensa, con tan solo 54 nodos, y que la sala donde se realiza el experimento tampoco excesivamente grande, la topología de red utilizada en la simulación para la comunicación entre los nodos con el Nodo Sink es la topología en estrella, es decir, la comunicación se produce directamente entre cada uno de los nodos y el nodo sumidero, enviando este último, a su vez, toda la información al servidor. Igualmente, cuando se precisa la información de varios nodos en el mismo instante, estos envían sus datos todos a la vez, pero la información llega al nodo Sink de manera ordenada, por lo que en ningún momento se produce colapso ni pérdida de información.

Una vez conocidas las ubicaciones y la forma en la que operan los nodos-sensores, se procede a la explicación de la codificación de la simulación. En primer lugar, es necesario realizar una lectura de los datos utilizados procedentes del laboratorio, los cuales son necesarios guardarlos en variables para poder trabajar con ellos. Al realizar la lectura del archivo se comprueba que existen tres tipos de datos: numéricos, fechas y horas. Como lo más cómodo es estar trabajando con valores numéricos, las fechas y horas, al ser registradas, se harán convirtiéndolas en números. Para las fechas, el número 1 corresponde al 1 de enero del año 0 d.C. por lo que, al primer día de registro, el día 24 de febrero de 2004, le corresponde el valor 732005. Por su parte, las horas al convertirse en números quedan registradas en valores entre 0 y 1, siendo 0 las 00:00:00 y 1 las 23:59:59. Así, a la primera hora registrada, la 01:18:47, le corresponde el valor de 0.0547 siendo redondeado a 4 decimales.

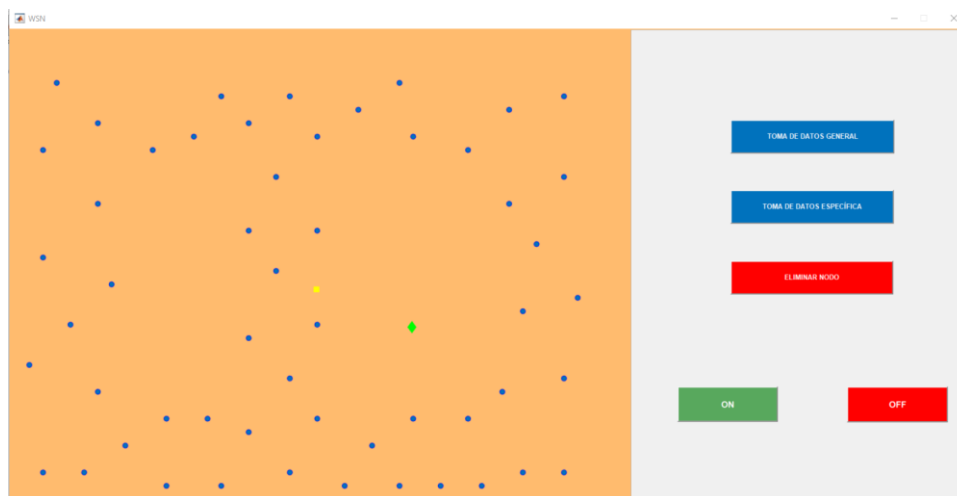


Figura 4-3. Interfaz de la demostración de la WSN en Matlab.

Tras el registro de la información de todos los nodos en variables dentro del software se procede a lanzar la interfaz, la cual tiene una forma como la mostrada en la Figura 4-3. En ella se puede apreciar tantos botones como funciones diferentes permite realizar la aplicación, siendo un total de cinco: ON, OFF, TOMA DE DATOS GENERAL, TOMA DE DATOS ESPECÍFICA y ELIMINAR NODO.

Al comienzo del lanzamiento de la aplicación, crean dos archivos Excel con formato “.xlsx”, llamados *datos extraídos* y *datos zona*, los cuales serán los encargados de almacenar la información cada vez que se realice una toma de datos por parte de los nodos sensores. Según la petición y, por consecuente, función pedida, se almacenará en uno u otro archivo, tal y como se verá en cada uno de sus subcapítulos.

Como punto de interés, hay que destacar que una vez se pone en marcha la simulación de la red de sensores inalámbricas, cada segundo que estén los nodos sin realizar ninguna actividad, ya sea porque no se ha producido ninguna petición o porque se encuentre en un estado de funcionamiento automático el cual las tomas se realizarán cada 10 segundos reales, en la interfaz este segundo corresponderá a 1:30 minutos.

Otro detalle es que no se conocen las ubicaciones de los nodos, es decir, el usuario cuando lance la aplicación no sabe la identidad de cada uno de los nodos, tal y como se puede observar en la imagen. En caso de que algún nodo sea anulado, el usuario no sabrá cuál de ellos es, simplemente sabrá la ubicación de este al desaparecer de la imagen. Si quisiera conocer el nodo anulado deberá comprobar la información extraída en las diferentes tomas y comprobar de cuál de ellos no tiene información. Esto se debe a que no es necesario para el usuario conocer la identidad de los nodos.

Para permitir al usuario conocer la dirección de los mensajes en la simulación, se ha escogido el color blanco como mensajes de salida, mientras que para los mensajes de entrada se ha escogido el color rojo. Esto es, para los mensajes enviados desde el Servidor al *Gateway* y desde éste a los diferentes nodos, la simulación mostrará un recorrido donde los mensajes serán de color blanco. Por su parte, si los mensajes van desde los nodos al *Gateway* y desde éste al Servidor, los mensajes serán de color rojo. Esto se observará mejor en las diferentes imágenes adjuntas de cada una de las funciones.

4.2.1 ON

Esta función es la encargada de dar comienzo a la simulación junto con la de ANULAR NODO. Si no se presiona uno de estos dos botones y, por tanto, no se lanzan sus funciones, no será posible empezar a correr la simulación.

Al presionar el botón de ON, se da comienzo a una toma de datos automática por parte de la simulación. En primer lugar, la función se encarga de realizar la toma de datos de la hora actual, o la última más próxima, de cada uno de los nodos que en nuestra simulación se encuentren activos. La información recogida en cada nodo incluye la fecha y hora original de la toma de datos y la información de temperatura, humedad y luminosidad existente. En caso de que un nodo no se encuentre activo, los datos recogidos serán de 0 en todos sus campos.

Una vez se ha realizado la recolección de datos, estos son enviados al nodo Sink tal y como se muestra en la Figura 4-4, donde los puntos rojos simulan el camino por el que va la información desde los nodos hasta el Sink. Además de esto, para evitar que otras funciones sean llamadas mientras se está procediendo a realizar una recolección de datos, la pantalla de los botones desaparece para dar paso a una ventana de advertencia donde se especifica que los nodos se encuentran en funcionamiento.

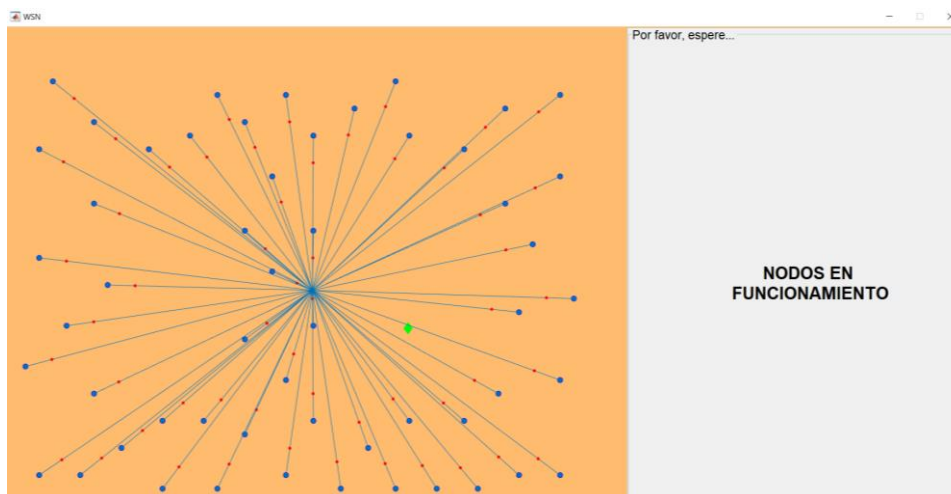


Figura 4-4. Envío de mensajes desde los Nodos al Nodo Sumidero.

Durante el periodo en el que los nodos están transmitiendo su información al nodo Sink, estos se encuentran activos y, por tanto, consumiendo energía de sus baterías. Así pues, los nodos más alejados consumen mayor energía que los nodos más cercanos debido a que su recorrido es mayor y por tanto se encuentran más tiempo en funcionamiento.

Una vez la información, ha llegado al nodo Sink, este recolecta toda esta información de cada uno de los nodos y procede a enviarla al servidor toda a la vez siguiendo el mismo proceso que realizaron los nodos.

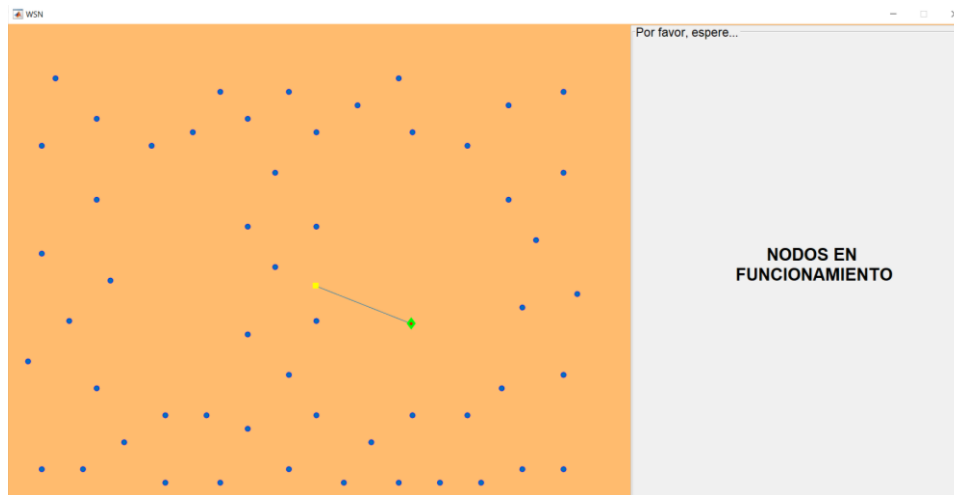


Figura 4-5. Envío de mensajes desde el Nodo Sumidero al Servidor.

Cuando la información ha llegado definitivamente al servidor, éste comprueba que los datos de todos los nodos son correctos, es decir, que de todos ellos se tienen valores relativamente reales. Si no fuera así, ya sea porque el nodo no se encuentra activo porque se ha decidido apagar en la simulación o porque realmente algún nodo se estropee en algún momento de durante el período de 19 días de tomas de datos, como es el caso del nodo 5 el cual, tras comprobar la información existente de Intel, se observa que no tiene recogida ninguna información en todo el periodo, estos nodos quedarán como inactivos y eliminarán de la simulación, tal y como se observa en la Figura 4-6.

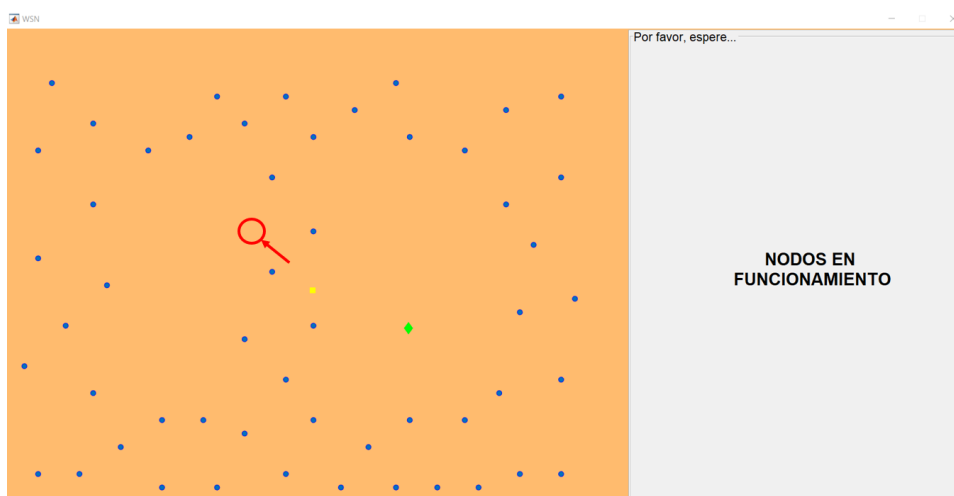


Figura 4-6. Interfaz tras la primera toma de datos automática. Nodo 5 desactivado.

Tras haberse estudiado los datos recogidos, estos son almacenados en el documento Excel “*datos extraídos.xlsx*”. Para ello, primero se comprueba cuánta es la información previa a la hora de la toma de datos incluida en dicho archivo y una vez conocido esto, la nueva información recogida en el servidor es almacenada junto con toda la anterior existente. Para evitar que existan datos erróneos que puedan alterar las estimaciones posteriores, los nodos inactivos y, por consiguiente, su información, no serán recogidos dentro del documento.

Por último, como se trata de una acción que permite un funcionamiento automático, este proceso se irá realizando cada 10 segundos reales, en la simulación equivalen a 15 minutos, mientras no se solicite una nueva petición. Si esto ocurriera, al terminar la petición, se volvería al estado automático volviéndose a esperarse 10 nuevos segundos para realizar una nueva recogida de muestras, por lo que podría darse la situación en la que entre una recogida de muestras y otra se haya producido un lapso mayor de tiempo o incluso menor si la petición realizada se produce al pulsar el botón TOMA DE DATOS GENERAL, como se explicará en el Apartado 4.2.3

Por último, recordar que, como se puede comprobar en la Figura 4-4 y en la Figura 4-5, el color de los mensajes es rojo, por los que la dirección de estos va desde los nodos hacia el servidor única y exclusivamente,

ya que al tratarse de una función en la que la recogida de información por parte de los nodos-sensores se produce cada 15 minutos reales desde la última petición registrada, no es necesario que sea el servidor quien exija estos datos, sino que está todo temporizado automáticamente.

4.2.2 OFF

Esta función, tal y como indica el título del apartado, se encarga de apagar la simulación, es decir, cuando se solicite esta petición la interfaz gráfica se cerrará automáticamente. Esta función, al igual que todas las demás, tan solo podrá ser invocada cuando los nodos no estén en funcionamiento, esto es, justo al lanzar la interfaz, tras la petición de *ELIMINAR NODO* siempre y cuando no se haya comenzado con la función de *ON* o en un intervalo de tiempo en los que el modo automático espera entre tomas de medidas.

En el momento en que se cierre la interfaz gráfica los datos recogidos en los archivos “*datos extraídos.xlsx*” y “*datos zona.xlsx*” serán los definitivos sin la posibilidad de poderse ampliar de nuevo. Si volviese a lanzarse la simulación estos datos se resetearían.

4.2.3 TOMA DE DATOS GENERAL

Esta acción, a diferencias de las anteriores, no puede ser invocada sin haberse adquirido una primera muestra de los nodos sensores realizada en modo automático. Esto se debe a que está pensado para que no se realicen peticiones sin haber comenzado previamente la toma de medidas y se realicen peticiones específicas sin haber entrado en funcionamiento la simulación.

Una vez se puede invocar esta función, el funcionamiento de esta consiste en, al igual que la función *ON*, en una toma de muestra de datos de todos los nodos activos en ese momento. La diferencia con esta radica en que, al tratarse de una petición explícita, es el servidor el que se encarga inicialmente de enviar la notificación al nodo Sink para que sepa que se ha solicitado una petición de tomas de datos. Una vez llega esta información al Gateway, este envía una notificación a todos los nodos sensores para que despierten de su estado de descanso y realicen una toma de medidas.

Como se observa en las Figura 4-7, los mensajes enviados desde el Servidor hasta los nodos tienen ese color blanco que los identifica. Igualmente, también se puede observar que el nodo 5 no se encuentra en la simulación, lo que significa que dejó de estar operativo tras la primera toma automática.

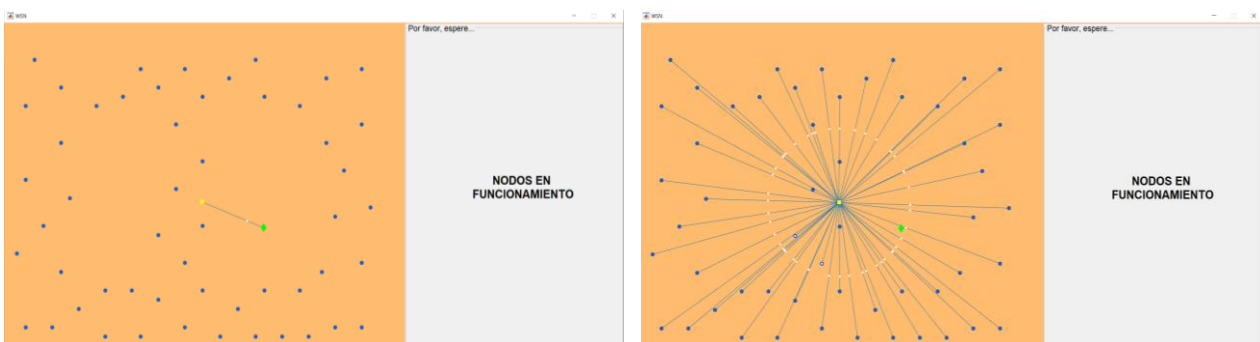


Figura 4-7. Envío de mensajes Servidor-Nodo Sumidero (a) y Nodo Sumidero-Nodos (b)

Cuando esta toma de medidas se ha realizado se produce el camino a la inversa, siguiendo exactamente los mismos pasos que realiza la función de *ON* cuando sus nodos están en funcionamiento, es decir, los nodos envían la información al Gateway y cuando todos los datos se han recolectado son enviados al Servidor; al llegar a él, estos son estudiados y analizados para saber si es necesario desactivar algún nodo que se encontrase activo y posteriormente ser almacenada la información en el archivo “*datos extraídos.xlsx*”.

En resumen, la acción general es la misma que la de *ON*, pero con el matiz de que es el servidor el que pide los datos y con la diferencia horaria, es decir, al tratarse de una petición que interrumpe la espera del modo automático, esta toma de medidas se produce antes de cumplir los 10 segundos de espera, por lo que entre los datos recogidos en el archivo Excel el tiempo transcurrido será siempre inferior a 15 minutos.

4.2.4 TOMA DE DATOS ESPECÍFICA

Este modo de funcionamiento, al igual que TOMA DE DATOS GENERAL, no puede ser invocado si no se ha iniciado previamente el modo automático ya que no parece lógico poner en funcionamiento los sistemas de refrigeración para una zona específica sin haberse puesto en marcha inicialmente para un modo general.

Cuando se realiza el lanzamiento de la función al pulsar este botón, la interfaz, en primer lugar, muestra una ventana en la que advierte al usuario de que debe escoger el punto donde desea conocer la información. Esta ventana podrá ser cerrada pulsando en *OK* o lo hará automáticamente cuando transcurran 4 segundos. Una vez ocurra una de estas dos opciones, el usuario podrá hacer clic en el punto donde sea conocer la información.

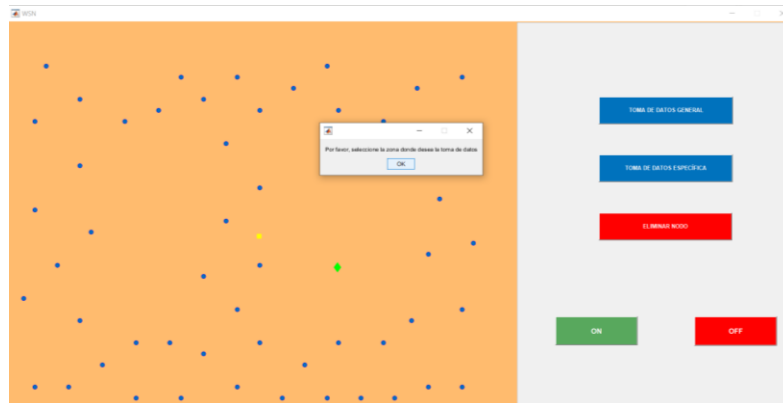


Figura 4-8. Ventana emergente Toma de Datos Específica

Al realizar su elección, la interfaz marcará el punto deseado con *Target*, Objetivo en inglés, y comenzará a realizarse la petición. Este punto objetivo desaparece tras un breve periodo de tiempo. Al igual que en la función anterior, al tratarse de una petición explícita, será el servidor el que realice el primer movimiento, enviando un mensaje al nodo Sink con esta petición. Cuando el mensaje llega hasta el nodo Sink, éste envía la petición a los nodos, pero a diferencia de las demás funciones, al tratarse de una zona específica, tan solo envía el mensaje a los cinco nodos más cercanos al punto objetivo.



Figura 4-9. Envío de mensajes Servidor-Nodo Sumidero y Nodo Sumidero-5 Nodos más cercanos

Una vez llegada la petición a los cinco nodos en cuestión, estos realizarán la toma de medidas y enviarán la información de vuelta al nodo Sink y este a su vez lo hará al Servidor.

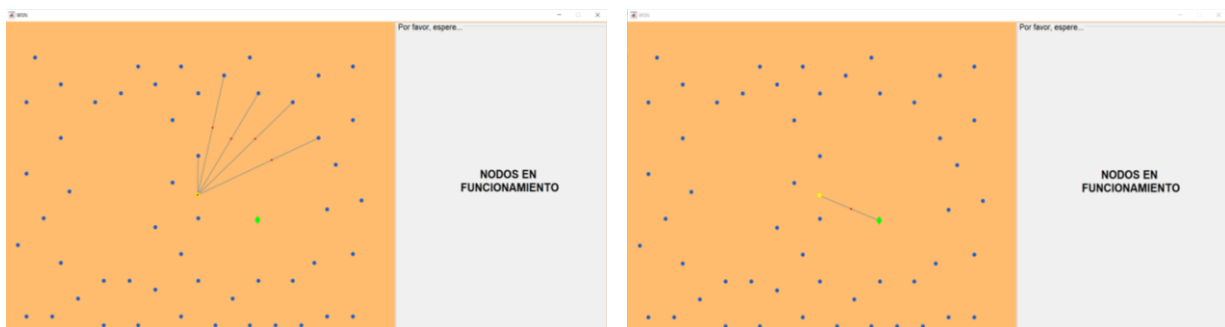


Figura 4-10. Envío de mensajes 5 Nodos más cercanos-Nodo Sumidero y Nodo Sumidero-Servidor.

Cuando se ha terminado de realizar todo el proceso de traspaso de información hasta llegar al Servidor, éste vuelve a analizar la información para conocer si es necesario inhabilitar alguno de los nodos que intervienen en la acción debido a que sus datos no sean correctos y así no registrar datos que puedan entorpecer la estimación en la zona a conocer. Tras analizar esta información recogida de los cinco nodos, será registrada en el archivo Excel “*datos zona.xlsx*” junto con la información de la fecha, hora y coordenadas del punto que se deseaba conocer. Esto se realiza para registrar el punto de referencia donde se desea conocer la estimación.

4.2.5 ANULAR NODO

Esta función, a diferencia de las demás, es la única que permite lanzarse antes de que la simulación de la red de sensores comience a funcionar. Esto es así para que, aunque la red esté estipulada en 54 nodos iniciales, se pueda tener la libertad de reducir el número de nodos antes de que comience a funcionar. Su función, por tanto, es la de inhabilitar o desactivar el nodo que se desea de la red de sensores.

Cuando se presiona este botón, al igual que con el de TOMA DE DATOS ESPECÍFICA, la interfaz lanza una ventana emergente en la que nos indica que seleccionemos el nodo a desactivar. Esta ventana se mantiene abierta durante un máximo de 4 segundos o bien hasta que el usuario cliquee sobre *OK*.

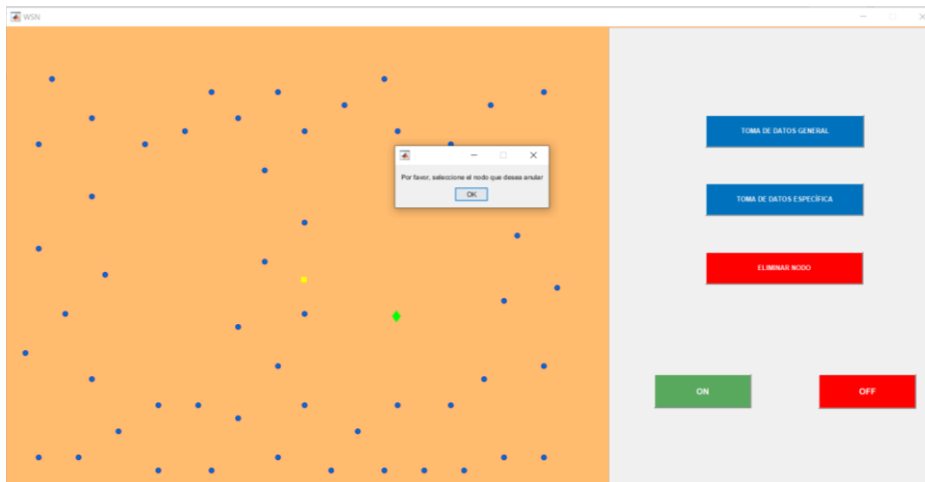


Figura 4-11. Ventana emergente Anular Nodo.

Una vez realizada una de las dos opciones, el usuario escogerá el nodo que desea anular simplemente pulsando sobre él en la interfaz. Para evitar desactivar nodos equivocados, la función tiene un margen de error de error de 3m, por lo que si no existiera ningún nodo en ese radio desde el lugar donde se ha seleccionado, no se inhabilitaría ningún nodo. En el ejemplo, se ha decidido, tras comprobarse que no aparecen sus datos, que el nodo a eliminar sea el nodo 14, lo cual se puede comprobar al comparar la Figura 4-11 y la Figura 4-12.

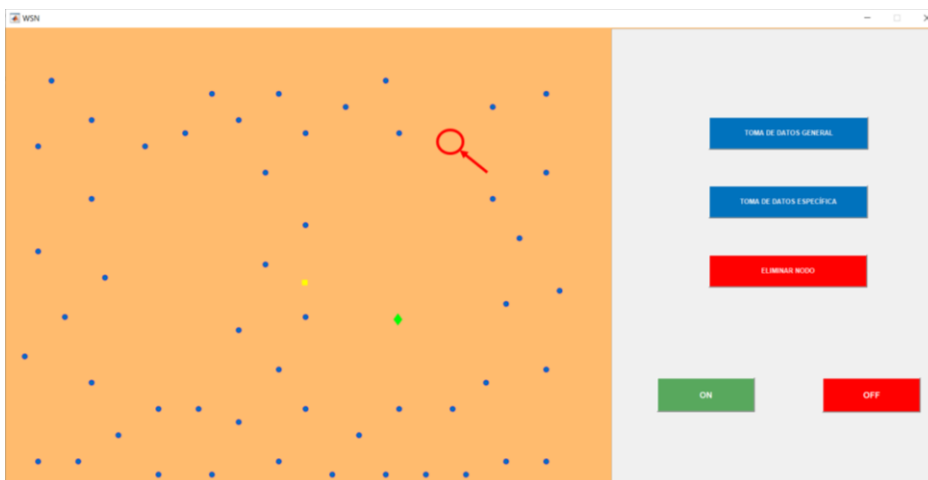


Figura 4-12. Nodo n°14 eliminado.

5 ESTUDIO DE LA DISTRIBUCIÓN ESPACIAL

Tras la realización de la interfaz en la que se muestra cómo sería el comportamiento de la red de sensores en caso de que hubiese sido posible instalar una WSN real en lugar de partir simplemente de los datos del laboratorio de Berkeley, se realizará un estudio del análisis espacial con los datos del mismo.

En primer lugar, se han recogido todos los datos en archivos Excel. Al realizarlo y verificar la cantidad de datos existentes de cada uno de los nodos, se ha comprobado que el nodo 5 no presentaba ningún dato recogido y que el nodo 15 tan solo recogía datos durante los 2 primeros días, por lo que al disponer de tan pocos datos de este nodo que pudiera alterar cualquier tipo de media, se ha decidido prescindir de él en su totalidad. Por otra parte, para realizar las estimaciones, de los 52 nodos restantes se han excluido 5 nodos, los cuales serán nuestra referencia para poder corroborar la efectividad de los modelos que se utilizarán en este proyecto. Ya que la disposición de estos nodos en el laboratorio de Berkeley no es equitativa en todo el campo, sino que lo es aproximadamente en las zonas habitables, tal y como se puede observar en la Figura 4-1, se han seleccionado nodos específicos con el fin de intentar alcanzar un mejor resultado en nuestras estimaciones, ya que excluir nodos en zonas más aisladas podría alterar enormemente los datos obtenidos. Así pues, los nodos escogidos para ser utilizados posteriormente para comprobar los datos son los **Nodos 10, 23, 30, 40 y 49**. Como resultado de este, se nos presenta el escenario de la Figura 5-1, con la disposición de los nodos utilizados, excluidos los nodos 5 y 15, y los nodos reservados en color rojo.

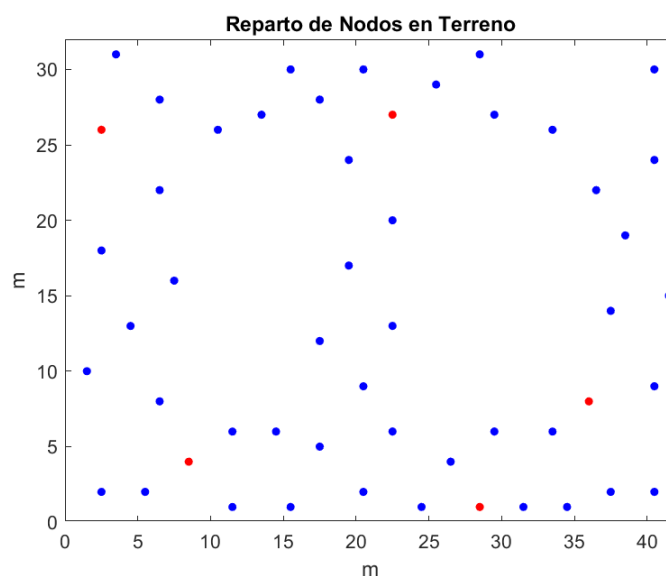


Figura 5-1 . Distribución de nodos utilizados y reservados.

Una vez conocidos los nodos a utilizar y estudiados los datos, se ha realizado el correspondiente estudio utilizando 2 modelos de interpolación diferentes, uno de cada tipo. De entre todos los modelos de Interpolación Deterministas se ha decidido utilizar el modelo de la Distancia Inversa Ponderada, o más conocido como IDW. Por su parte, de entre los modelos de Interpolación Geoestadísticos se ha optado por el modelo de Kriging Ordinario.

A partir de estos modelos se ha realizado un estudio para comprobar cuál de ellos es mejor para la estimación de las temperaturas. Para ello, el modelo IDW realizará en primer lugar las estimaciones en las ubicaciones de los nodos en un instante determinado. El instante escogido aleatoriamente de entre los 19 días que conforman todos los datos de los que disponemos es el día 4 de marzo de 2004 a las 13:16:00. Ya que no todos los nodos tienen registrado datos a esa hora, se escogerá aquel dato anterior a la fecha disponible por cada nodo.

Por otro lado, el modelo de Kriging Ordinario realizará el estudio con la media de cuatro periodos diferentes:

- El primero de ellos ha consistido en realizar la media global de los 19 días para cada uno de los nodos de forma individual.
- El segundo consiste en realizar la media diaria de todos los nodos con los datos de forma que se obtengan medias diferentes de cada nodo para cada día, provocando así que cada nodo tenga 19 medias diferentes, una para cada día.
- El tercer modelo consta de la realización de la media horaria entre las horas en punto, de forma que se obtengan las medias de todos los nodos entre las 24 horas del día.
- Por último, el cuarto periodo está compuesto por los datos de cada nodo del mismo instante en el que fue realizado el modelo IDW, es decir, con los datos del instante de cada nodo en el día 4 de marzo a las 13:16:00.

Cabe destacar que para el modelo IDW tan solo se ha realizado el estudio con un instante debido a que este modelo, a diferencia del Kriging, no necesita de estudios previos para realizar las estimaciones, sino que directamente con los datos in situ es capaz de realizarlos, de ahí que simplemente se trabaje, inicialmente, con un instante concreto.

Tal y como se acaba de comentar, ya que IDW no precisa de ningún estudio previo antes de realizar las estimaciones, nos centraremos en los pertinentes estudios iniciales a realizar para poder realizar el Kriging Ordinario.

5.1 Estudio exploratorio de los datos

En estos estudios previos, tal y como se comentó en el Capítulo 3.2, el primer paso es realizar un estudio exploratorio de los datos. Debido a que exponer los cuatro periodos diferentes conllevaría una gran cantidad de datos y gráficas, se va a visualizar simplemente la media global de todos los nodos que conforman la red.

Para ello, se realiza en primer lugar el histograma. Este histograma nos permite saber cuál es la distribución que siguen los datos recolectados. Para que el Kriging obtenga las mejores estimaciones y, por ende, los mejores resultados, es necesario que los datos sigan una distribución lo más simétrica posible, siendo la mejor la distribución normal, generando unos resultados más óptimos.

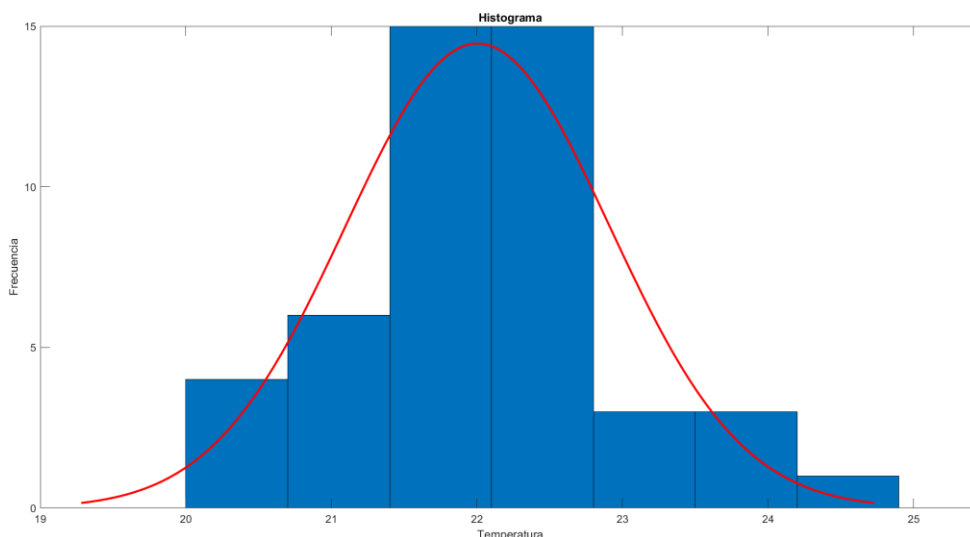


Figura 5-2. Histograma de la media global de los nodos.

Como se puede comprobar en la Figura 5-2, el histograma presenta una distribución normal por lo que presenta una buena perspectiva de futuro para realizar unas estimaciones óptimas. Igualmente, aunque el histograma sea óptimo, no es algo fiable pero tampoco indispensable, ya que al contar con un número bastante reducido de nodos (recordemos que contamos tan solo con 52 muestras), es posible que no sea suficiente con

realizar este estudio exploratorio. Añadir otras pruebas como el gráfico q-q puede ser una buena opción que complementa al histograma. Así pues, realizando dicho gráfico se obtienen los siguientes resultados.

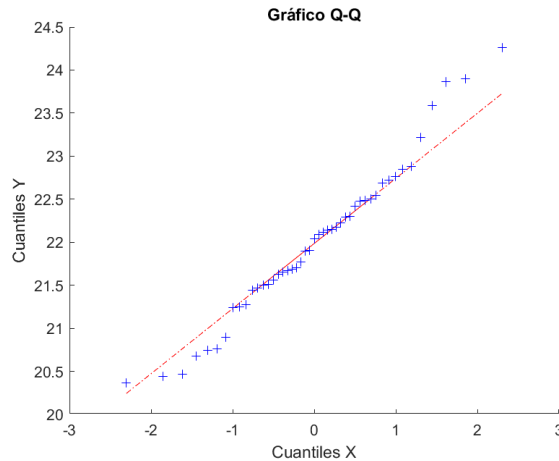


Figura 5-3. Gráfica q-q de la media global de los nodos.

Se puede observar que los resultados resultantes presentan una distribución normal ya que, a excepción de unos pocos puntos, los valores se encuentran aproximadamente alineados en una recta de 45°, lo que hace que la media global de los nodos presente una tendencia que presupone unas estimaciones óptimas.

Aunque esto sea así para este período, no tienen por qué ocurrir lo mismo en el resto de ellos o en alguno de los estudios dentro de cada uno de ellos. Este es el caso de, como ocurre, por ejemplo, los datos del día 8 de marzo, donde claramente se observa que no muestra una distribución normal.

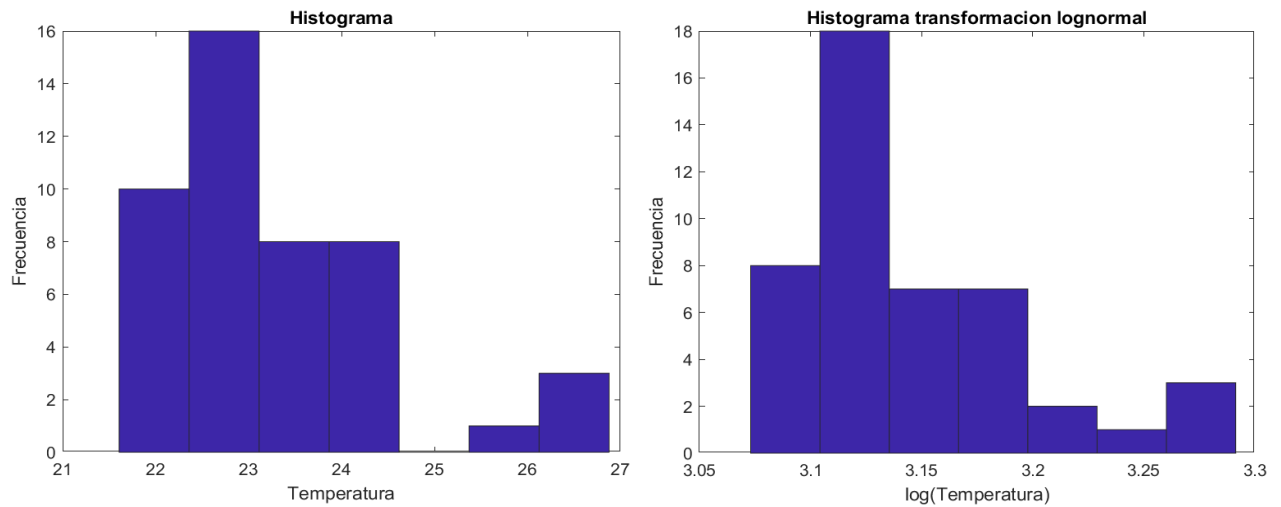


Figura 5-4. Histograma 8 de marzo sin normalizar (izquierda) y normalizado (derecha).

En estos casos, lo ideal es realizar una normalización de los datos para intentar alcanzar una distribución normal. En este proyecto, en los casos en lo que la distribución no seguía una distribución normal, se ha optado por aplicar una transformación log-normal [38]. Algunos periodos mejoran sus resultados, pero en este caso del día 6, la aplicación de dicha transformación no supone una gran diferencia con la normalización inicial, por lo que el estudio previo nos hace suponer que las estimaciones no van a ser del todo óptimas para este día.

5.2 Estimación del variograma

El siguiente paso al estudio exploratorio de los datos consiste en la realización del variograma experimental de los mismos. Igualmente, basando los estudios en el modelo global, se realiza su variograma. Para ello, se ha utilizado la función **variogram**[39], pertenecientes a una de las librerías de Matlab. Para ello, se le ofrece como datos de entrada las ubicaciones de los nodos, los valores de cada uno de ellos y una serie de valores que pueden ser seleccionados por elección del usuario, como es la cantidad de parámetros que se van a tomar, la distancia máxima a la que medir, si deseamos que sea isotrópico o no, entre otros.

Inicialmente es necesario comprobar la omnidireccionalidad del variograma, es decir, comprobar los modelos en las diferentes direcciones y corroborar que la dirección no es un factor determinante para la estimación de las temperaturas, verificando así su isotropía. Para ello, se realizan los variograma con los ángulos de 0°, 45°, 90° y 135°, abarcando así los 360° con una diferencia entre los modelos de 45°.

Para realizar esta estimación del variograma experimental se han utilizado como parámetros de entrada la obtención de 30 valores, que la distancia máxima medida sea de 30 metros, que el modelo sea anisotrópico y que el ángulo, como ya se ha comentado, tenga una variación de 45°. Los resultados obtenidos son:

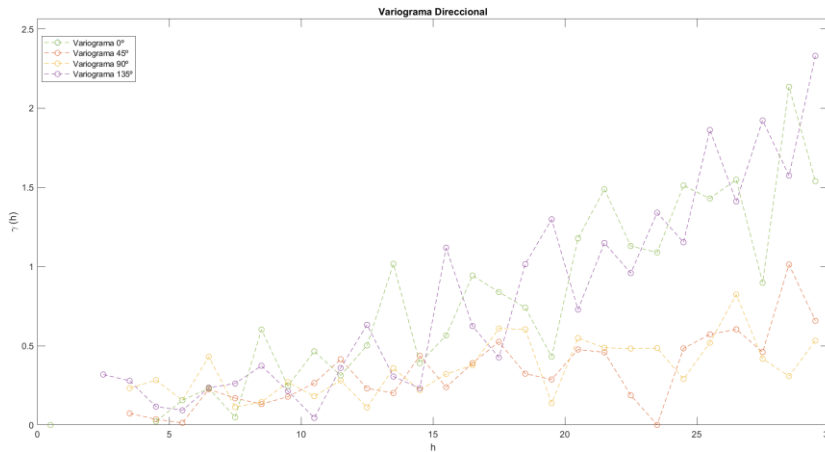


Figura 5-5. Variogramas direccionales de la media global

Se puede comprobar que los 4 variogramas direccionales tienen resultados bastante similares en las menores distancias, mientras que a mayores distancias se comienza a producir una mayor dispersión de los datos, donde se observan diferenciadas dos tendencias. Aunque ocurra esto, debido a que, por las medidas del terreno de 42x32m, la mayor distancia posible entre dos nodos es de 52.8 m. aproximadamente, por lo que los resultados con mayor relevancia a la hora de realizar las estimaciones son los más cercanos, pudiendo interpretarse que el modelo de las medias globales presenta isotropía.

Así pues, utilizando de nuevo como parámetros de entrada la obtención de 30 parámetros diferentes y una distancia máxima de 30 metros, se calcula de nuevo el variograma, diferenciándolo del anterior en que en lugar de ser un que el modelo sea anisotrópico sea ahora isotrópico. Con estos parámetros el modelo de variograma obtenido presenta la siguiente forma:

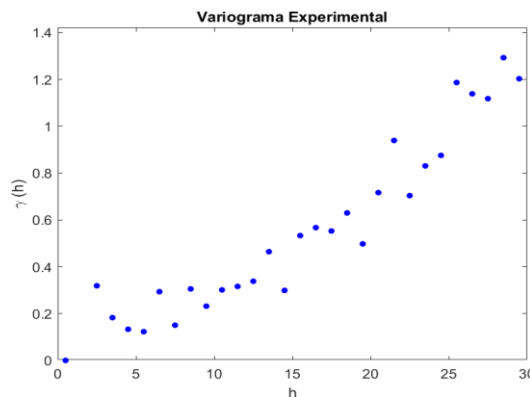


Figura 5-6. Variograma experimental omnidireccional de la media global

Con el variograma experimental realizado, se procede a la modelación del variograma teórico. Para ello, se comprueba qué modelo es el que mejor se ajusta a los resultados obtenidos. De entre los diferentes modelos disponibles, se compararán los resultados con los modelos exponencial y esférico.

Para realizar esto nos apoyamos en la función **variogramfit**[39] el cual, ofreciéndole como entrada los valores γ y de h obtenidos de la función **variogram** y que se pueden observar en la Figura 5-7, realizará la estimación de los valores del rango, la meseta y la pepita. Además de las entradas comentadas, esta función permite escoger el modelo de variograma teórico deseado, siendo el utilizado por defecto, si no se decide escoger ninguno, el modelo esférico.

Así pues, los resultados obtenidos son los siguientes:

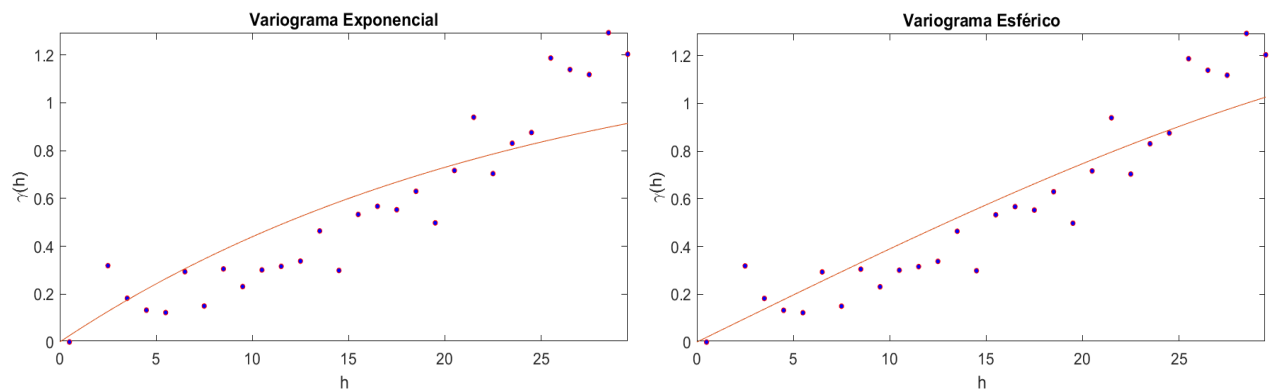


Figura 5-7 Variograma experimental y esférico de la media global

Como se puede observar, el modelo que mejor se ajusta a los datos es el modelo el modelo esférico, pero no se va a descartar ninguno hasta finalizar todo el proceso de estimación de los datos.

Se puede comprobar en ambos modelos que el efecto pepita es nulo. Por otro lado, los valores obtenidos para la meseta son de 1.2931 para los dos modelos. Por último, y a diferencia de los otros dos parámetros, el rango tiene valores diferentes entre los modelos, siendo de 24.0498 para el exponencial, y 49.0262 para el esférico.

5.3 Estimación por Kriging Ordinario

Una vez conocidos los diferentes parámetros del variograma teórico, el siguiente paso consiste realización de las estimaciones mediante Kriging. Como se ha comentado, el modelo escogido para la realización de estas es el Kriging Ordinario. Para ello, nos hemos apoyado en la función **kriging**[39] la cual, requiere como variables de entradas una estructura con todos los datos obtenidos del modelo del variograma, en la que se recoge, entre otros datos, el modelo utilizado, los valores del rango, meseta y pepita, los valores de γ y de las distancias, etc.; las coordenadas de los nodos los que disponemos datos; los valores de estos nodos; y las coordenadas donde queremos realizar las estimaciones, las cuales será todo el terreno del laboratorio.

Con esto, la función devuelve dos parámetros: el primero es una matriz de valores de las estimaciones obtenidas por Kriging en cada una de las posiciones del terreno; mientras que el segundo es otra matriz de iguales dimensiones que la anterior donde se recogen los valores de varianza de Kriging para las estimaciones.

Así, siguiendo con el modelo de las medias globales, las estimaciones y varianzas obtenidas con los modelos exponenciales y esféricos son los mostrados en la Figura 5-8. En ella, de forma visual se han plasmado las estimaciones, permitiendo observar cómo se produce la distribución de la temperatura en el interior de la sala.

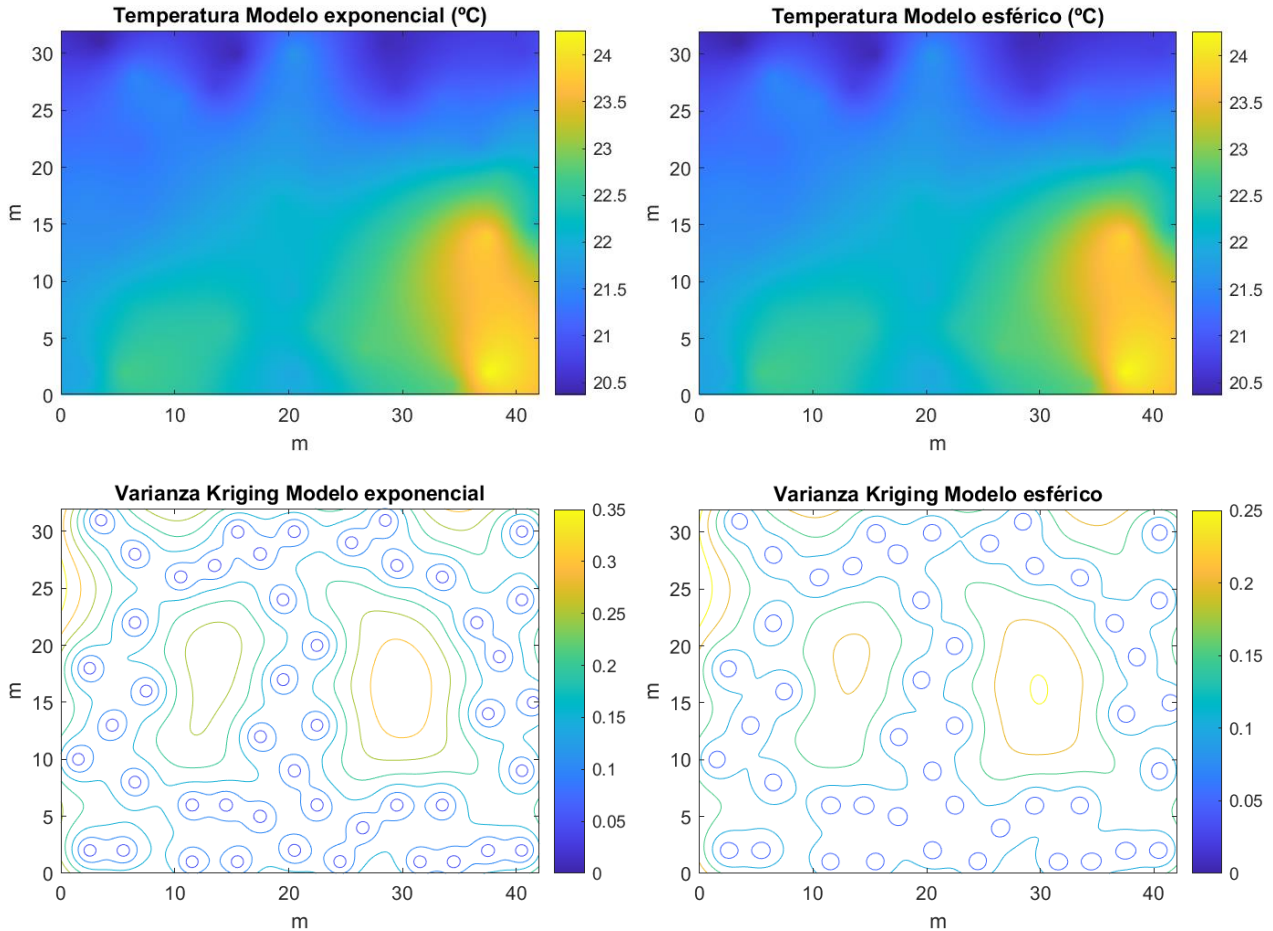


Figura 5-8. Estimaciones y varianza Kriging de la media global con modelo exponencial y esférico

Se comprueba que ambos modelos presentan estimaciones y estructuras muy similares. Por su parte, las varianzas, correspondientes a las dos imágenes inferiores, presentan estructuras parecidas pero mejores valores para el modelo esférico. Tal y como se puede comprobar en la Figura 5-1, la disposición de los nodos presentaba dos áreas sensiblemente afectadas donde no se ubicaba ningún nodo. Esto ha repercutido en la varianza de Kriging de ambos modelos, cuyas zonas donde no se dispone de datos, es decir, las correspondientes a los colores amarillo y naranja presentan mayores varianzas, provocando estimaciones menos fiables que en las zonas próximas a las ubicaciones de los nodos, correspondiente a al color azul.

Aunque las varianzas sean parecidas, se puede comprobar que, tal y como ocurría con el variograma teórico, el modelo esférico se ajusta mejor, ya que sus máximas varianzas son menores que las del modelo exponencial.

5.4 Validación cruzada

El último paso consiste en comprobar la veracidad de las estimaciones realizando la validación cruzada. Esta consiste en eliminar uno de los nodos y comprobar la estimación en ese punto con la ayuda del resto de nodos. Como esta ubicación realmente es conocida, se puede comprobar cómo de fiable es la estimación realizada. Este se realiza igual con cada uno de los nodos, permitiendo conocer los errores en las estimaciones en cada punto conocido donde se ubican los nodos, saber que zonas tienen mejor y peor estimación y poder estudiar el motivo de los errores.

Ya que el método de interpolación por IDW estima el valor en la posición deseada a partir de los datos cercanos, para estimar la temperatura en la ubicación de cada nodo podría considerarse que se engloba dentro de este apartado, ya que para calcular el valor en un punto conocido es necesario excluirlo y utilizar los nodos cercanos. Por tanto, es en este punto donde entraría el método de IDW [40] para la estimación de las temperaturas en el momento que se ha decidido, es decir, el día 4 de marzo a las 13:16:00. Igualmente, tal y

como se ha realizado durante todo el proceso para calcular las estimaciones, se presentarán los datos de forma gráfica del método de interpolación de Kriging ordinario para las medias globales.

Así pues, realizando la validación cruzada para cada uno de los nodos obtenemos la siguiente gráfica, en la que se muestra tanto la validación cruzada con el modelo exponencial como con el esférico.

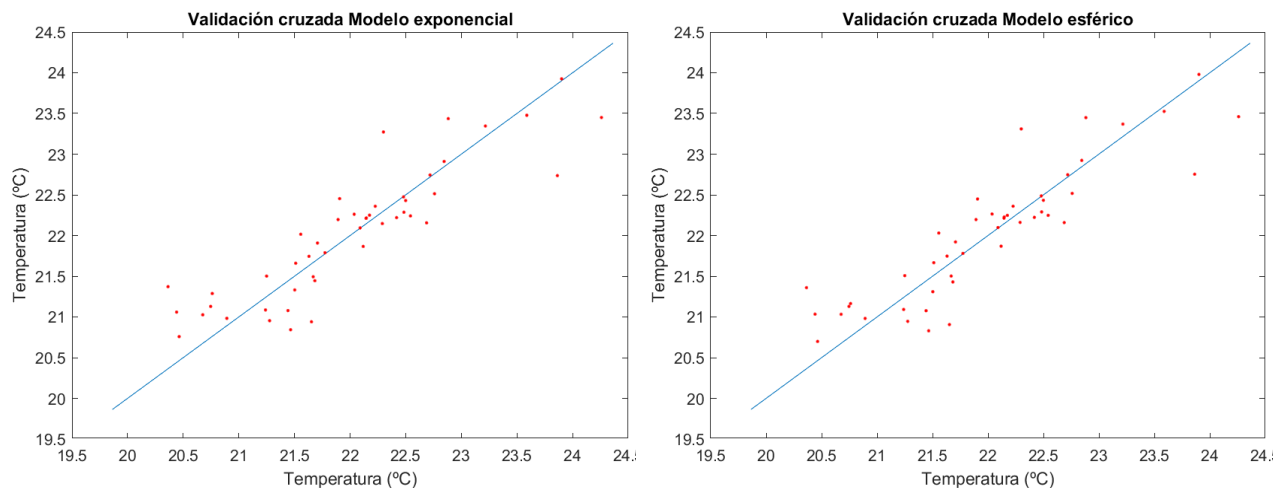


Figura 5-9. Validación cruzada de los modelos exponencial y esférico

Para que la validación cruzada sea óptima, los puntos deben situarse aproximadamente en una recta de 45°. Como se puede comprobar, ambos modelos presentan resultados bastante parecidos, donde la mayoría de los puntos, a excepción de algunos más aislados producidos a mayor temperatura, se encuentran cercanos a esta recta de 45°, lo que implica que las estimaciones son buenas con ambos modelos.

Igualmente, para comprobar los errores cometidos se pueden aplicar diferentes medidores de error para comprobar cuál es el error cometido sin necesidad de deducirlos visualmente. Para ello podemos utilizar medidores como el error absoluto medio (MAE), el error relativo medio (MRAE) y el error cuadrático medio (RMSE), cuyas fórmulas son las siguientes [41]:

$$MAE = \frac{1}{N} \sum_{i=1}^N |Z(x_i) - Z^*(x_i)| \tag{5-1}$$

$$MRAE = \frac{1}{N} \sum_{i=1}^N \frac{|Z(x_i) - Z^*(x_i)|}{Z(x_i)} \times 100 (\%) \tag{5-2}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N [Z(x_i) - Z^*(x_i)]^2}{N}} \tag{5-3}$$

Siendo $Z(x_i)$ el valor conocido en la coordenada x_i y $Z^*(x_i)$ el valor estimado en la coordenada x_i .

Para que los resultados sean óptimos, es necesario que todos los estimadores sean los más pequeños posibles, es decir, que sus valores se acerquen a 0. Cuanto mayor sea esto, mejores estimaciones obtendremos.

Con todo esto, los valores de los diferentes medidores obtenidos para la media global son:

Modelo	MAE	MRAE	RMSE
Exponencial	0,3065	1,3988	0,4113
Esférico	0,3052	1,3920	0,4095

Tablas 5-1. Estimaciones obtenidas por validación cruzada para la media global.

Se puede comprobar que finalmente, tal y como ha ido ocurriendo en el transcurso de los pasos a seguir, ambos modelos tienen resultados prácticamente similares, siendo el modelo esférico el que obtiene mejor resultado, pero siendo la diferencia entre ambos prácticamente indiferente.

Si se realiza todos estos mismos pasos para los otros tres periodos diferentes, realizando a los periodos de las medias por días y por horas las medias de los diferentes valores obtenidos en MAE, MRAE y RMSE, además de realizar las estimaciones para el modelo IDW, obtenemos los siguientes resultados finales:

Período	Modelo	MAE	MRAE	RMSE
Estimación diaria	Exponencial	0.3912	1.7583	0.5411
	Esférico	0.3904	1.7537	0.5404
Estimación horaria	Exponencial	0.4186	1.8889	0.5681
	Esférico	0.4212	1.9001	0.5713
Estimación puntual Kriging ordinario	Exponencial	0.9670	3.5001	1.5528
	Esférico	0.9457	3.4145	1.5512
Estimación IDW	5 vecinos cercanos	0.9706	3.5646	1.5570

Tablas 5-2. Estimaciones obtenidas por validación cruzada.

Se comprueba a partir de los resultados arrojados que el periodo que mejor responde a sus resultados es el de la media global. Esto es coherente, ya que es el modelo que más resultados recoge y, por tanto, los datos de aquellos nodos que durante los 19 días hayan sido más dispersos en relación con la temperatura media de la sala, ya sea por errores en la toma de datos o por perturbaciones que se produzcan en ciertas zonas en determinados momentos, tienen menos efectos, provocando que no tengan una importante relevancia en la estimación media global obtenida.

De todos los resultados obtenidos es preciso hacer hincapié en los obtenidos de forma puntual, ya sea por Kriging Ordinario o por IDW. Sus resultados son el doble o triple peor que el del resto de modelos. Esto se debe a lo comentado anteriormente. Si se observan la Tabla 5-3, la cual contiene los datos de cada uno de los nodos el día 4 de marzo a las 13:16:00, se puede observar que existen temperaturas muy dispares en función de las zonas donde se tomen, ya que si la temperatura ronda los 20-25 °C en la mayor parte de la sala, de los nodos 21 a 32, las temperaturas suben drásticamente pasando a rondar los 32-36 °C. Esta zona, si observamos la Figura 4-1, corresponde la esquina inferior derecha de la sala y cuyos datos, más que a deberse a fallos de medida, ya que esta perturbación involucra a muchos nodos, parece tratarse más bien a agentes externos o actividades realizadas en la sala que interfirieron en la temperatura. Esto, como los resultados finales han reflejado, provoca grandes alteraciones en las estimaciones al realizarse con modelos puntuales, ya que los valores de gamma obtenidos al realizar el variograma se disparan, tal y como se puede comprobar en la Figura 5-10.

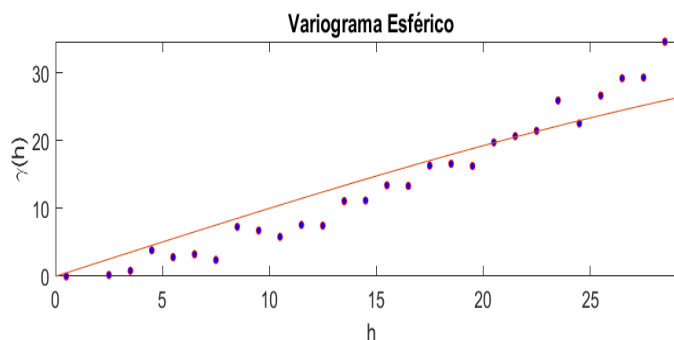


Figura 5-10. Variograma teórico esférico puntual (13:16 del día 4 de marzo).

<i>Nodo 1</i>	26.1918	<i>Nodo 16</i>	22.9872	<i>Nodo 29</i>	33.4050	<i>Nodo 43</i>	22.6728
<i>Nodo 2</i>	24.3494	<i>Nodo 17</i>	24.0064	<i>Nodo 31</i>	32.0424	<i>Nodo 44</i>	22.6246
<i>Nodo 3</i>	25.0648	<i>Nodo 18</i>	26.4760	<i>Nodo 32</i>	33.5320	<i>Nodo 45</i>	22.3208
<i>Nodo 4</i>	23.4086	<i>Nodo 19</i>	28.1028	<i>Nodo 33</i>	27.7696	<i>Nodo 46</i>	21.4976
<i>Nodo 6</i>	23.4772	<i>Nodo 20</i>	27.4952	<i>Nodo 34</i>	25.2314	<i>Nodo 47</i>	20.9782
<i>Nodo 7</i>	22.5168	<i>Nodo 21</i>	35.7468	<i>Nodo 35</i>	24.1730	<i>Nodo 48</i>	20.8312
<i>Nodo 8</i>	23.3400	<i>Nodo 22</i>	33.7770	<i>Nodo 36</i>	25.2804	<i>Nodo 50</i>	20.6548
<i>Nodo 9</i>	23.7418	<i>Nodo 24</i>	35.5606	<i>Nodo 37</i>	24.4572	<i>Nodo 51</i>	20.5764
<i>Nodo 11</i>	22.5364	<i>Nodo 25</i>	36.2956	<i>Nodo 38</i>	26.5936	<i>Nodo 52</i>	22.1444
<i>Nodo 12</i>	22.1836	<i>Nodo 26</i>	32.7286	<i>Nodo 39</i>	25.0844	<i>Nodo 53</i>	22.2326
<i>Nodo 13</i>	21.1056	<i>Nodo 27</i>	32.3854	<i>Nodo 41</i>	26.2506	<i>Nodo 54</i>	22.7030
<i>Nodo 14</i>	23.7516	<i>Nodo 28</i>	33.3948	<i>Nodo 42</i>	24.0064		

Tablas 5-3. Temperaturas recogidas por cada nodo el día 4 de marzo a las 13:16.

5.5 Estimación de la temperatura durante un día concreto.

Aunque inicialmente se ha obtenido que el modelo de variograma para las medias globales obtiene mejores resultados para sus propios datos que el resto de modelos, se va a intentar corroborar que esto sucede también para cualquier tipo de dato, ya que no es comparable realizar la estimación de la temperatura con los datos de las propias medias que, con los datos de instantes diferentes, por ejemplo. Es por ello por lo que, volviendo a utilizar los 5 modelos diferentes, se va a realizar un estudio de la estimación de un día completo, en concreto del día 6 marzo, realizando estimaciones cada 30 minutos, en las horas en punto e y media, obteniendo un total de 48 estimaciones para cada uno de los nodos y para cada uno de los modelos. En esta ocasión, como ya se ha comentado, simplemente se va a utilizar el modelo esférico de variograma para realizar dichas estimaciones.

Con esto se intentará conocer si el modelo de variograma de la media global es realmente el modelo que mejor se adapta para estimar la temperatura o si por el contrario es otro el que lo hace. Como para los modelos diarios y horarios se disponía de diferentes variogramas se van a utilizar los acordes en cada momento, esto es, en el caso del modelo diario, al realizar el estudio para todo el día 6 de marzo, de los 19 variogramas que se habían obtenido se utilizará el específico de ese día; mientras que, para el modelo horario, al realizarse tomas durante todo el transcurso del día, en función de la hora a la que se realice se utilizará uno u otro. Por ejemplo, para las estimaciones que vayamos a realizar a las 8:30 y a las 9:00, el modelo de variograma utilizado será el que recogía todos los datos entre las 8:00:01 y las 9:00:00. Así se dispondrá con las 24 horas del día.

Por su parte el modelo que utiliza los datos in situ, en ese instante concreto, modificará su variograma cada vez que se realice una estimación.

Para conocer mejor el modelo de variograma utilizados para cada una de las horas, se adjuntan los 24 modelos experimentales y teóricos obtenidos, a partir de los cuales se va a trabajar para obtener las estimaciones. Para no excedernos, no se adjuntan los modelos instantáneos que se utilizan cada 30 minutos, ya que eso conllevaría incluir 48 gráficas adicionales.

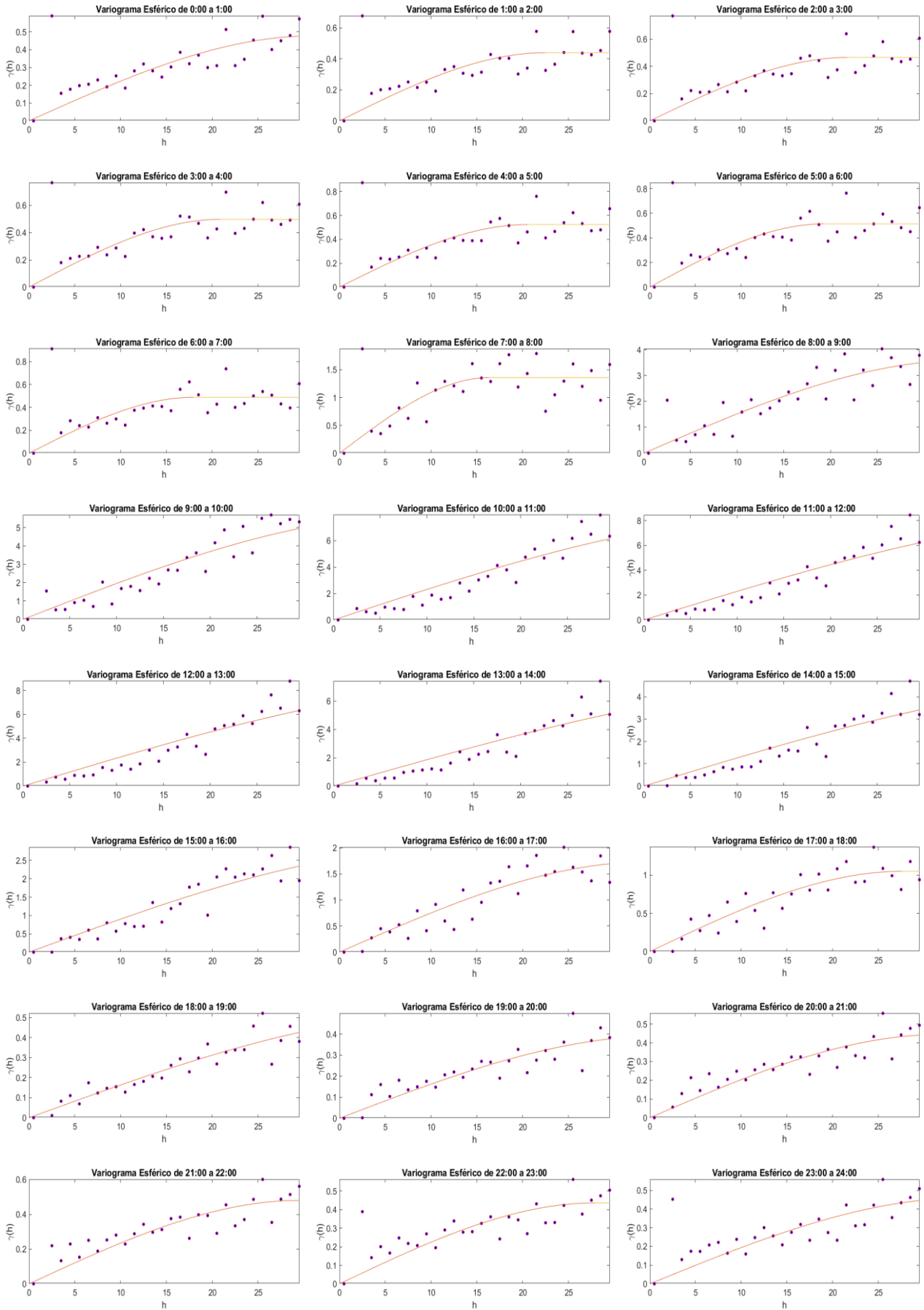


Figura 5-11. Variogramas teóricos esféricos con las medias de cada hora.

	VARIOGRAMA MEDIA GLOBAL			VARIOGRAMA DIA 6 DE MARZO			VARIOGRAMA POR HORAS			VARIOGRAMA INSTANTE			MÉTODO IDW		
	MAE	MRAE	RMSE	MAE	MRAE	RMSE	MAE	MRAE	RMSE	MAE	MRAE	RMSE	MAE	MRAE	RMSE
Estimación a las 00:30	0.2196	1.1168	0.2790	0.2201	1.1191	0.2800	0.2252	1.1451	0.2879	0.2615	1.3310	0.3364	0.2956	1.5062	0.3840
Estimación a la 01:00	0.2232	1.1527	0.2866	0.2237	1.1553	0.2877	0.2325	1.2006	0.2968	0.2677	1.3822	0.3441	0.3063	1.5846	0.3919
Estimación a la 01:30	0.2211	1.1508	0.2837	0.2216	1.1536	0.2846	0.2334	1.2167	0.3091	0.2678	1.3942	0.3448	0.3061	1.5970	0.3886
Estimación a las 02:00	0.2308	1.2078	0.2882	0.2313	1.2105	0.2891	0.2412	1.2637	0.3125	0.2745	1.4370	0.3493	0.3071	1.6117	0.3910
Estimación a las 02:30	0.2270	1.1958	0.2855	0.2275	1.1985	0.2864	0.2395	1.2637	0.3133	0.2700	1.4229	0.3454	0.3047	1.6093	0.3886
Estimación a las 03:00	0.2293	1.2138	0.2922	0.2298	1.2165	0.2931	0.2428	1.2877	0.3213	0.2749	1.4547	0.3549	0.3068	1.6272	0.3929
Estimación a las 03:30	0.2207	1.1707	0.2863	0.2212	1.1732	0.2868	0.2322	1.2340	0.3161	0.2861	1.5159	0.3950	0.2930	1.5575	0.3834
Estimación a las 04:00	0.2482	1.3206	0.3114	0.2487	1.3235	0.3122	0.2604	1.3880	0.3421	0.2872	1.5284	0.3716	0.3151	1.6804	0.4040
Estimación a las 04:30	0.2393	1.2800	0.3044	0.2398	1.2829	0.3051	0.2537	1.3601	0.3346	0.3074	1.6363	0.4181	0.3097	1.6605	0.4029
Estimación a las 05:00	0.2346	1.2595	0.3022	0.2350	1.2617	0.3029	0.2489	1.3400	0.3340	0.2778	1.4929	0.3676	0.3101	1.6696	0.4024
Estimación a las 05:30	0.2325	1.2550	0.2993	0.2329	1.2575	0.2999	0.2476	1.3409	0.3351	0.2969	1.5972	0.4082	0.3079	1.6669	0.4015
Estimación a las 06:00	0.2426	1.3225	0.3102	0.2431	1.3251	0.3110	0.2564	1.4018	0.3451	0.2829	1.5442	0.3757	0.3230	1.7660	0.4184
Estimación a las 06:30	0.2411	1.3228	0.3060	0.2416	1.3252	0.3069	0.2664	1.4659	0.3549	0.2835	1.5576	0.3729	0.3207	1.7647	0.4162
Estimación a las 07:00	0.2405	1.3232	0.3045	0.2410	1.3257	0.3051	0.2638	1.4558	0.3515	0.2826	1.5566	0.3704	0.3172	1.7503	0.4073
Estimación a las 07:30	0.2308	1.2431	0.3147	0.2287	1.2319	0.3127	0.2615	1.4149	0.3789	0.2654	1.4359	0.3841	0.2688	1.4544	0.3695
Estimación a las 08:00	0.2877	1.4810	0.3842	0.2842	1.4633	0.3791	0.3502	1.8068	0.4996	0.2932	1.5094	0.3934	0.3103	1.5931	0.4441
Estimación a las 08:30	0.3380	1.6534	0.4564	0.3337	1.6322	0.4502	0.3531	1.7255	0.4793	0.3406	1.6654	0.4597	0.3817	1.8450	0.5301
Estimación a las 09:00	0.3781	1.7783	0.5055	0.3726	1.7529	0.5004	0.4013	1.8839	0.5315	0.3748	1.7634	0.5021	0.4575	2.1036	0.5928
Estimación a las 09:30	0.3675	1.7048	0.4809	0.3651	1.6925	0.4771	0.3716	1.7239	0.4853	0.3653	1.6935	0.4770	0.4164	1.9064	0.5466
Estimación a las 10:00	0.4458	1.9563	0.5778	0.4448	1.9500	0.5764	0.4471	1.9639	0.5797	0.4447	1.9487	0.5763	0.5179	2.2431	0.6571
Estimación a las 10:30	0.4914	2.0779	0.6233	0.4904	2.0722	0.6227	0.4915	2.0784	0.6234	0.4904	2.0712	0.6228	0.5498	2.3109	0.6967
Estimación a las 11:00	0.5287	2.1666	0.6788	0.5282	2.1628	0.6797	0.5287	2.1669	0.6788	0.5281	2.1616	0.6807	0.5777	2.3597	0.7394
Estimación a las 11:30	0.5663	2.2502	0.7478	0.5685	2.2581	0.7495	0.5668	2.2515	0.7486	0.5702	2.2652	0.7507	0.6163	2.4416	0.7983
Estimación a las 12:00	0.5928	2.3120	0.7979	0.5956	2.3232	0.8004	0.5943	2.3177	0.7991	0.5971	2.3289	0.8018	0.6432	2.4996	0.8577
Estimación a las 12:30	0.6070	2.3398	0.8275	0.6096	2.3497	0.8301	0.6087	2.3462	0.8292	0.6113	2.3561	0.8316	0.6555	2.5289	0.8675
Estimación a las 13:00	0.6050	2.3070	0.8366	0.6080	2.3184	0.8391	0.6071	2.3148	0.8383	0.6098	2.3252	0.8407	0.6455	2.4685	0.8710
Estimación a las 13:30	0.5638	2.1362	0.7573	0.5665	2.1463	0.7589	0.5667	2.1468	0.7590	0.5678	2.1512	0.7599	0.5919	2.2497	0.7937
Estimación a las 14:00	0.5174	1.9587	0.6867	0.5198	1.9677	0.6881	0.5199	1.9681	0.6881	0.5205	1.9704	0.6886	0.5503	2.0957	0.7165
Estimación a las 14:30	0.4780	1.8146	0.6189	0.4802	1.8231	0.6195	0.4780	1.8148	0.6189	0.4803	1.8232	0.6195	0.5394	2.0485	0.6916
Estimación a las 15:00	0.4517	1.6882	0.5712	0.4483	1.6753	0.5693	0.4516	1.6878	0.5711	0.4491	1.6784	0.5695	0.5343	1.9934	0.6882
Estimación a las 15:30	0.4006	1.4994	0.5010	0.3963	1.4829	0.4978	0.4120	1.5420	0.5133	0.3990	1.4935	0.4993	0.4920	1.8330	0.6393
Estimación a las 16:00	0.3182	1.2179	0.4015	0.3144	1.2031	0.3986	0.3297	1.2620	0.4133	0.3163	1.2105	0.3996	0.4058	1.5440	0.5297
Estimación a las 16:30	0.3077	1.1775	0.4723	0.3021	1.1558	0.4686	0.3248	1.2430	0.4920	0.3088	1.1819	0.4727	0.3858	1.4637	0.5945
Estimación a las 17:00	0.2802	1.0758	0.4286	0.2766	1.0612	0.4256	0.2885	1.1102	0.4489	0.2811	1.0793	0.4293	0.3357	1.2748	0.5309
Estimación a las 17:30	0.2844	1.1138	0.4201	0.2828	1.1068	0.4186	0.3019	1.1810	0.4517	0.2844	1.1136	0.4200	0.3295	1.2845	0.4931
Estimación a las 18:00	0.2284	0.9470	0.2977	0.2278	0.9444	0.2966	0.2345	0.9735	0.3223	0.2283	0.9467	0.2973	0.2660	1.0999	0.3551
Estimación a las 18:30	0.2202	0.9433	0.2762	0.2203	0.9436	0.2757	0.2188	0.9382	0.2779	0.2201	0.9430	0.2760	0.2487	1.0679	0.3288
Estimación a las 19:00	0.2125	0.9286	0.2723	0.2126	0.9289	0.2721	0.2110	0.9228	0.2727	0.2125	0.9285	0.2722	0.2537	1.1128	0.3463
Estimación a las 19:30	0.2255	1.0090	0.3034	0.2257	1.0099	0.3032	0.2264	1.0147	0.3108	0.2254	1.0087	0.3033	0.2719	1.2192	0.3802
Estimación a las 20:00	0.2130	0.9616	0.2860	0.2133	0.9629	0.2860	0.2169	0.9805	0.2938	0.2129	0.9614	0.2857	0.2651	1.2000	0.3716
Estimación a las 20:30	0.2108	0.9615	0.2856	0.2111	0.9628	0.2855	0.2131	0.9735	0.2926	0.2107	0.9611	0.2852	0.2730	1.2485	0.3746
Estimación a las 21:00	0.2105	0.9683	0.2785	0.2108	0.9698	0.2787	0.2134	0.9828	0.2861	0.2103	0.9674	0.2779	0.2738	1.2626	0.3713
Estimación a las 21:30	0.2257	1.0485	0.2973	0.2259	1.0498	0.2975	0.2366	1.1011	0.3156	0.2256	1.0481	0.2970	0.2961	1.3793	0.3935
Estimación a las 22:00	0.2286	1.0698	0.3016	0.2290	1.0718	0.3019	0.2403	1.1262	0.3191	0.2296	1.0749	0.3030	0.3013	1.4137	0.3999
Estimación a las 22:30	0.2246	1.0613	0.3019	0.2249	1.06270	0.3023	0.2378	1.1256	0.3239	0.2260	1.0686	0.3045	0.2945	1.3958	0.4041
Estimación a las 23:00	0.2470	1.1891	0.3395	0.2473	1.1905	0.3398	0.2634	1.2694	0.3642	0.2524	1.2162	0.3471	0.3231	1.5581	0.4492
Estimación a las 23:30	0.2500	1.2102	0.3329	0.2503	1.2116	0.3332	0.2534	1.2280	0.3392	0.2578	1.2497	0.3475	0.3213	1.5587	0.4408
Estimación a las 24:00	0.2478	1.2112	0.3354	0.2481	1.2127	0.3357	0.2495	1.2210	0.3392	0.2803	1.3858	0.4627	0.3275	1.6040	0.4483

Tablas 5-4. Resultado de los medidores de error durante el día 6 de marzo

	MAE	MRAE	RMSE
Modelo global	0.3444	1.5342	0.4533
Modelo día 6 de marzo	0.3438	1.5313	0.4526
Modelo por horas	0.3524	1.5736	0.4685
Modelo puntual	0.3539	1.6057	0.4721
Método IDW	0.4075	1.8264	0.5408

Tablas 5-5. Resultado de las medias de los medidores de error durante el día 6 de marzo.

Tras la realización de Kriging con todos los modelos y utilizando los medidores de error para cada uno de los instantes, se obtienen los resultados de MAE, MRAE y RMSE recogidos en la Tabla 5-4. En ella se incluyen todos los datos de cada uno de los nodos para los diferentes modelos de variograma en cada uno de los instantes. Posteriormente, se ha realizado la media de cada medidor para cada modelo obteniendo el valor promedio de cada medidor y así conocer cuál de ellos es el mejor, como se muestra en la Tabla 5-5.

De las tablas de resultados se concluye que el mejor medidor es en este caso el del modelo de variograma propio del día 6 de marzo, aunque con diferencias muy pequeñas con el modelo de las medias globales. Si comparamos los resultados con los obtenidos en el primer estudio, se ratifica que ambos modelos también eran los que mejores resultados obtuvieron, aunque en esa ocasión el modelo de variograma global obtuvo bastante mejores resultados.

Por otro lado, aunque la dificultad y el tiempo de ejecución del método de interpolación IDW respecto al método de interpolación de Kriging ordinario es mucho menor, los resultados reflejan que IDW, al menos en el caso que en el presente proyecto nos compete, presenta peores estimaciones que el método de Kriging, inclusive con su homólogo del modelo puntual haciendo que los aspectos positivos de IDW no tengan la suficiente relevancia como para decidir su elección.

Así pues, de entre los 4 modelos de variogramas utilizados con Kriging Ordinario, teniendo en cuenta las primeras estimaciones donde el modelo global era mejor, que las estimaciones durante todo el día 6 de marzo no difieren prácticamente entre el modelo global y el diario y, quizás el aspecto más importante, que en el caso en que en una línea de futuro se decidiera estimar temperaturas en días no recogidos en la base de datos, el modelo diario no podría utilizarse ya que ha podido utilizarse al realizarse el variograma conociendo con anterioridad al proceso de estimación los datos de todo el día, siendo únicamente válido como modelo a posteriori, por lo que se llega a la conclusión de que las estimaciones obtenidas con el variograma de la media global es el más adecuado.

Hay que recalcar que, ciertamente, el periodo total que comprende la base de datos no es muy prolongado, por lo que para épocas del año en las que las temperaturas son notablemente diferentes con respecto a la época del año comprendida en nuestro periodo de 19 días véase, verano, sería necesario o bien realizar nuevas mediciones y obtener un nuevo modelo de variograma de medias globales que sea utilizado durante esas fechas o bien obtener un modelo un modelo que comprenda los datos de todo el año y comprobar su comportamiento ante cualquier día del año.

5.6 Comprobación de la fiabilidad del modelo

Una vez determinado según los resultados obtenidos que el modelo más idóneo para calcular las estimaciones es el modelo de las medias globales mediante el variograma esférico, el siguiente paso es la comprobación de su fiabilidad. Para ello, se realiza la estimación de las temperaturas en las ubicaciones excluidas inicialmente al comienzo de este capítulo, es decir, en las ubicaciones de los nodos 10, 23, 30, 40, y 49.

Estas estimaciones estarán apoyadas en el modelo esférico global. En cada una de las horas se realizará Kriging Ordinario con los valores de los parámetros obtenidos del variograma y con las temperaturas de los 47

nodos restantes, recordemos que, de los 54 nodos, 2 eran inservibles y 5 son los excluidos. Esto, realizado cada 30 minutos durante el día 6 de marzo de 2004, ofrece 48 estimaciones para cada nodo, siendo la media total de los medidores de cada uno de ellos las recogidas en la tabla siguiente.

	MAE	MRAE	RMSE
Nodo 10	0.1229	0.5793	0.1591
Nodo 23	0.3076	1.3860	0.3546
Nodo 30	0.1943	0.7350	0.2895
Nodo 40	0.1508	0.5953	0.2144
Nodo 49	0.2808	1.3901	0.3185

Tablas 5-6. Resultado de las medias de los errores obtenidas de los nodos no utilizados

De los resultados se puede comprobar que los errores cometidos son cercanos a 0, lo que indica que se ha realizado una buena estimación. Tanto es así que los valores obtenidos son incluso mejores que los producidos al realizar la validación cruzada, por lo que se corrobora que el método de interpolación de Kriging ordinario, utilizando los parámetros obtenidos de las medias globales con el variograma esférico, es apto para realizar las estimaciones del laboratorio de Intel Berkeley.

Es cierto que los nodos excluidos inicialmente para corroborar los datos del modelo obtenido son nodos cuya presencia no era, en teoría, estrictamente necesaria para conocer su ubicación, ya que se encontraban rodeados de otros sensores, otorgándoles así un mayor respaldo a la hora de obtener una mejor estimación, a excepción de que tuviera perturbaciones externas a las provocadas por la temperatura la propia sala, que otros nodos cuya eliminación sí pudiera provocar mayores errores. Si se observan las medias de los medidores de error durante el día 6 de marzo de forma individual en cada uno de los nodos al realizar la validación cruzada, se comprueba que existen nodos cuyos errores son más altos. En la Tabla 5-7 se agrupan los valores del medidor del error medio absoluto (MAE) de estos nodos al realizar la validación cruzada.

<i>Nodo 1</i>	0,223	<i>Nodo 13</i>	0,354	<i>Nodo 25</i>	0,469	<i>Nodo 36</i>	0,421	<i>Nodo 47</i>	0,386
<i>Nodo 2</i>	0,259	<i>Nodo 14</i>	0,137	<i>Nodo 26</i>	0,407	<i>Nodo 37</i>	0,254	<i>Nodo 48</i>	0,113
<i>Nodo 3</i>	0,166	<i>Nodo 16</i>	0,327	<i>Nodo 27</i>	0,290	<i>Nodo 38</i>	0,150	<i>Nodo 50</i>	0,618
<i>Nodo 4</i>	0,361	<i>Nodo 17</i>	0,316	<i>Nodo 28</i>	0,585	<i>Nodo 39</i>	0,203	<i>Nodo 51</i>	0,526
<i>Nodo 6</i>	0,215	<i>Nodo 18</i>	0,217	<i>Nodo 29</i>	0,192	<i>Nodo 41</i>	0,326	<i>Nodo 52</i>	0,321
<i>Nodo 7</i>	0,228	<i>Nodo 19</i>	0,138	<i>Nodo 31</i>	0,138	<i>Nodo 42</i>	0,409	<i>Nodo 53</i>	0,603
<i>Nodo 8</i>	0,619	<i>Nodo 20</i>	0,470	<i>Nodo 32</i>	0,122	<i>Nodo 43</i>	0,579	<i>Nodo 54</i>	1,006
<i>Nodo 9</i>	0,702	<i>Nodo 21</i>	0,261	<i>Nodo 33</i>	0,217	<i>Nodo 44</i>	0,467		
<i>Nodo 11</i>	0,260	<i>Nodo 22</i>	0,261	<i>Nodo 34</i>	0,171	<i>Nodo 45</i>	0,189		
<i>Nodo 12</i>	0,626	<i>Nodo 24</i>	0,404	<i>Nodo 35</i>	0,237	<i>Nodo 46</i>	0,293		

Tablas 5-7. MAE obtenidos de la validación cruzada de cada nodo en °C

De entre los resultados de las validaciones cruzadas, comprobamos que la mayoría de los nodos adquieren valores aproximados al MAE medio. Sin embargo, existen otros nodos, entre ellos los casos del nodo 8, 9, 12 y, en especial, el nodo 54 que adquieren valores más disparados. Estos nodos corresponden a los ubicados en la parte superior de la sala en el centro. Comprobando los datos correspondientes a las temperaturas en esos nodos durante todo el día, se observa que la media del nodo 8 y 9 es de unos 21.6 °C, mientras que la del nodo 54 es 1.5°C inferior, es decir, 20.1°C. Tal abrupta diferencia de temperatura se debe a que, si se observa la Figura 4-1, estos nodos se encuentran separados por una pared, lo que provoca que se creen dos pequeños microclimas que en el momento de realizar el variograma este detalle no se tuviera en cuenta, de ahí que a la hora de realizar las estimaciones esta diferencia provoque mayores errores. Para corregir esto la solución más coherente sería la realización de subdivisiones cuadrículas dentro del laboratorio, creando diferentes zonas a las que aplicar el variograma y Kriging, cada uno con el suyo propio.

El problema que ocasionaría esto es la imposibilidad de realizar Kriging en ninguna de las subdivisiones. El modelo de interpolación Kriging está pensado para que sea utilizado con un número mínimo de parámetros de los que partir. Si nuestra base de datos máxima es de 54 nodos realizar, por ejemplo, 6 subdivisiones, provocaría que cada una de ellas estuviera compuesta por unos 8-10 nodos, en función de cómo estén distribuidos, lo que impediría la estimación por este método. En ese caso, parece más conveniente realizar un modelo determinista como el IDW, que basan todo su modelo en el peso que proporcionan los nodos más cercanos, pero como se ha comprobado en este proyecto, en caso de que mejorasen, no lo harían en gran medida ya que de producirse otros agentes externos afectarían en gran medida a sus resultados.

Por tanto, en este proyecto, no parece viable realizar una optimización de la sala, tanto por la imposibilidad de realizar subdivisiones como por la innecesidad de añadir más nodos a una sala cuyas cantidades están acorde a las dimensiones de esta.

Aun así, teniendo en cuenta que los sensores de temperaturas existentes en el mercado tienen una precisión máxima de $\pm 0.1^\circ\text{C}$, estos los errores máximos cometidos en las estimaciones no supondrían gran problema, por lo que las estimaciones realizadas en líneas generales son óptimas.

6 CONCLUSIONES

El objetivo principal de este trabajo era obtener las mejores estimaciones de la temperatura en el interior de una sala gracias a la utilización de una red de sensores inalámbricos para poder realizar un estudio de su distribución. Para ello, ante la imposibilidad de utilizar motas reales que permitieran crear esta red de sensores y realizar las tomas de temperatura a partir de la cuales poder realizar la estimación en el resto de la sala, se ha partido de una base de datos ofrecida por el laboratorio de Intel Berkeley, el cual dispuso de una red de sensores y realizó la toma de diferentes medidas, entre ellas la temperatura, durante un periodo de tiempo prolongado de forma experimental.

Así pues, apoyándonos en esta base de datos y en las ubicaciones de los nodos sensores, se ha realizado en primer lugar una recreación de cómo habría sido nuestra red de sensores si hubiese dispuesto de las motas para crear la red. Después, utilizando los datos disponibles, se ha realizado mediante el uso de los métodos de interpolación Kriging Ordinario e IDW y con las medias de las temperaturas recogidas por los nodos en distintos periodos, un estudio que permitiera estimar cuál es el mejor método para realizar dichas estimaciones y qué periodo se adaptaba mejor al cálculo de las estimaciones. Además, para corroborar las premisas obtenidas, se ha comprobado con estos mismos métodos y modelos obtenidos en cada periodo las temperaturas durante un día completo, realizando los errores en las estimaciones respecto a las temperaturas reales. Por último, con el modelo que mejores resultados obtuvo, se llevó a cabo las estimaciones en ciertas ubicaciones reservadas de algunos de los nodos para comprobar la efectividad real del modelo.

Los resultados arrojaron que de entre los dos métodos de interpolación utilizados, el que mejores resultados ofrecía era el método de interpolación Kriging y que, dentro de los diferentes modelos empleados con este método, el que mejor resultados otorgaba era aquel cuyos parámetros del variograma se obtenía mediante el uso de la media de todos y cada uno de los valores de los nodos.

Comprobando los valores estimados respecto a los nodos excluidos se obtuvieron errores en la estimación poco elevados, llevando a la conclusión de que el método de Kriging es óptimo y válido para predecir las temperaturas en el interior de una sala, lo que puede permitir conocer el estado de confort que se tiene dentro de ella.

6.1 Líneas de mejora e investigación futura.

- Comparar los métodos de interpolación utilizados con otros métodos más avanzados como, por ejemplo, la utilización de redes neuronales, para comprobar la efectividad de cada uno de ellos y realizar una comparativa para saber la relación calidad-coste que tendría el hecho de aplicar sistemas más complejos.
- Realizar el estudio en exteriores, permitiendo conocer si la veracidad del método es aplicable tanto en lugares cerrados como en espacios abiertos.
- Realizar un estudio para conocer el número de nodos óptimo-necesarios para obtener unos resultados que no se vean afectados por la reducción de nodos, con el fin de poder reducir costes en la instalación.
- Utilización de una red de sensores propias y real, en lugar de una representación de otra realizada. Esto permitiría aplicar todos los conocimientos adquiridos teóricamente acerca de cómo se produciría la comunicación entre los diferentes nodos sensores, además de disponer de una base de datos actualizada y realiza en las ubicaciones deseadas.

- Utilizar una red de sensores propia y adquirir nuestros datos podría permitir estudiar el estado de confort obtenido en la sala, pudiendo realizar un control de la temperatura en el interior de la sala in situ. El hecho de que los datos hubiesen sido recogidos de forma previa hacía imposible llevar a cabo esta función, ya que posibles alteraciones en la temperatura ocasionadas por sistemas refrigerantes no se apreciarían en las temperaturas ya registradas.

REFERENCIAS

- [1]. Hossam Mahmoud Ahmad Fahmy. *Wireless Sensor Networks. Concepts, Applications, Experimentation and Analysis*, 2016. Springer. DOI 10.1007/978-981-10-0412-4.
- [2]. Senchun Chai, Zhaoyang Wang, Baihai Zhang, Lingguo Cui & Runqi Chai. *Wireless Sensor Network*, Pte Ltd. 2020. Springer. <https://doi.org/10.1007/978-981-15-5757-6>
- [3]. Roberto Fernández Martínez... [et al.] (Integrantes del Grupo de Investigación EDMANS). *Redes inalámbricas de sensores: teoría y aplicación práctica*, 2009. Universidad de la Rioja.
- [4]. Sebastián Román Maroto Cantillo. *Desarrollo de aplicaciones basadas en WSN*. S.1. 2010. Universidad Politécnica de Valencia.
- [5]. Vidyasagar Potdar, Atif Sharif, Elizabeth Chang. *Wireless Sensor Networks: A Survey*. Digital Ecosystems and Business Intelligence Institute, Curtin University of Technology, Perth, Western Australia. 2009 IEEE. DOI 10.1109/WAINA.2009.192
- [6]. R.R. Selmic, Vir V. Phoha & Abdul Serwaddal. *Wireless Sensor Networks. Security, Coverage, and Localization*. 2016. Springer. DOI 10.1007/978-3-319-46769-6
- [7]. Divya Sharma, Sandeep Verma, Kanika Sharma. *Network Topologies in Wireless Sensor Networks: A Review*. Vol. 4, 2013. Dept. of ECE, NITTTR, Chandigarh, UT, India.
- [8]. M.A. Matin and M.M. Islam. *Overview of Wireless Sensor Network*, 2012. <http://dx.doi.org/10.5772/49376>
- [9]. Ana-Belén García-Hernando, José-Fernán Martínez-Ortega, Aggeliki Prayati, Luis Redondo-López, MsC, Juan-Manuel López-Navarro. *Problem Solving for Wireless Sensor Networks*. 2009. Springer. DOI 10.1007/978-1-84800-203-6.
- [10]. Rahman, Abdur & El Saddik, Abdulmotaleb & Gueaieb, Wail. (2008). *Wireless Sensor Network Transport Layer: State of the Art*. 10.1007/978-3-540-69033-7_11.
- [11]. Yogesh Yogesh Sankarasubramaniam Özgür B. Akan Ian F. Akyildiz. *ESRT: Event-to-Sink Reliable Transport in Wireless Sensor Networks*. 2003, Annapolis, Maryland, USA.
- [12]. <https://present5.com/transport-layer-congestion-control-reliable-transport-data-from/>
- [13]. Miguel A. Labrador, Pedro M. Wightman. *Topology Control in Wireless Sensor Networks. With a companion simulation tool for teaching and research*, 2009. Springer.
- [14]. Kazem Sohraby Daniel Minoli Taieb Znati. *Wireless Sensor Networks. Technology, Protocols, and Applications*, 2007 John Wiley & Sons, Inc. ISBN 978-0-471-74300-2.
- [15]. C. S. Raghavendra, Krishna M. Sivalingam, Taieb Znati. *Wireless Sensor Networks*, 2006. Springer. ISBN 9780387352695.
- [16] Junaid Ahmed Khan, Hassaan Khaliq Qureshi, Adnan Iqbal. *Energy management in Wireless Sensor Networks: A survey*. *Computers and Electrical Engineering*, Elsevier, 2014, Volume 41, January 2015, pp. 159-176. [ff10.1016/j.compeleceng.2014.06.009](https://doi.org/10.1016/j.compeleceng.2014.06.009)[ff.fhal-01283728f](https://doi.org/10.1016/j.compeleceng.2014.06.009)
- [17]. Ramón Jesús Millán Tejedor. *UWB: la revolución de las redes personales*, 2005. IDDG Communications S.A. <https://www.ramonmillan.com/tutoriales/uwb.php#introduccion>
- [18]. Preetha, A Divya & Lakshmi R, Srividhya. (2015). *Secure Communication using WirelessHART in Wireless Sensor Networks (WSN)*. 10.13140/RG.2.2.16990.89923.

- [19]. Sinem Coleri Ergen. ZigBee/IEEE 802.15.4 Summary, 2004.
- [20]. F. Labeau, A. Agarwal and B. Agba, "Comparative study of Wireless Sensor Network standards for application in Electrical Substations," 2015 International Conference on Computing, Communication and Security (ICCCS), 2015, pp. 1-5, doi: 10.1109/CCCS.2015.7374135.
- [21]. <https://www.republica.com/2020/08/07/implantan-por-primera-vez-en-espana-un-marcapasos-sin-cables-y-un-desfibrilador-subcutaneo/#>
- [22]. <https://www.pngwing.com/es/free-png-xtvfo>
- [23]. <https://desktop.arcgis.com/es/arcmap/10.3/manage-data/terrains/hidden-natural-neighbor-interpolation.htm>
- [24]. <https://pro.arcgis.com/es/pro-app/latest/tool-reference/3d-analyst/how-spline-works.htm>
- [25]. <http://www6.uniovi.es/~feli/CursoMDT/Tema1/Page16.html>
- [26]. D. Murillo et al. Comparación de métodos de interpolación para la generación de mapas de ruido en entornos urbanos. Vol. 3. N°1, 2012.
- [27]. R.C. A. Minnitt & W. Assibey-Bonsu. Transactions of the Royal Society of South Africa, 2013, vol. 68, No. 3, 199-202, <http://dx.doi.org/10.1080/0035919X.2013.826748>
- [28]. Gabriela Barrios, José Ignacio Bonzi Rios, Joaquín Galindo. Proyecto número 2: Estimación de recursos minerales mediante Kriging, 2012. Universidad de Chile.
- [29]. Xavier Emery. Geoestadística, 2013. Universidad de Chile.
- [30]. Gallardo A. 2006. Geoestadística . Ecosistemas. 2006/3. URL: http://www.revistaecosistemas.net/articulo.asp?Id=431&Id_Categoria=1&tipo=portada
- [31]. Roberto Díaz. Análisis Variográfico. Mine-class Capacitaciones online
- [32]. Julio A. Valle Poblete. Wireless Sensor Networks and Kriging, 2010. Universidad de Valencia
- [33]. María José Ginzo Villamayor. Análisis Geoestadístico de datos funcionales. Universidad Da Coruña
- [34]. Willian De Jesús Caballero Guardo. Kriging Universal para Datos Funcionales, 2011. Universidad de Colombia.
- [35]. Gmalíel Chávez Hernández. Metodología para el análisis geoestadístico para la interpolación de la precipitación en El Salvador, 2009. Universidad de El Salvador
- [36]. <https://es.mathworks.com/help/>
- [37]. <http://db.csail.mit.edu/labdata/labdata.html>
- [38]. Luis Miguel Molinero Casares. ¿Y si los datos no siguen una distribución normal?... Bondad de ajuste a una normal. Transformaciones. Pruebas no paramétricas. 2003. URL: <https://seh-lelha.org/2003/07/14/bondad-de-ajuste-a-una-normal/>
- [39]. Wolfgang Schwanghart (w.schwanghart[at]unibas.ch), 2010
- [40]. Simone Fatichi -- simonef@dicea.unifi.it, 2009
- [41]. MAE y RMSE: ¿qué métrica es mejor? URL: <https://ichi.pro/es/mae-y-rmse-que-metrica-es-mejor-252933908062525>

GLOSARIO

WSN: Wireless Sensor Network	3
TCP/IP: Transmission Control Protocol/Internet Protocol	3
AC/DC: Alternate Current / Direct Current	4
FPGA: Field Programmable Gate Arrays	4
ISM: Industrial, Scientific and Medical	4
MAC: Media Access Control	4
SPIN: Sensor Protocols for Information via Negotiation	10
UDP: User Datagram Protocol	10
CODA: Congestion Detection and Avoidance	11
ESRT: Event to Sink Reliable Transport	11
NC: Non-Congestion	11
LR: Low Rate	11
HR: High Rate	11
OOR: Optimum Operating Range	11
PSFQ: Pump Slowly, Fetch Quickly	12
ACK: Acknowledgement	13
NACK: Negative Acknowledgement	13
RMST: Reliable multi-segment transport	13
SMP: Sensor Management Protocol	14
TADAP: Task Assignment and Data Advertisement Protocol	14
SQDDP: Sensor Query and Data Dissemination Protocol	14
IEEE: Institute of Electrical and Electronics Engineers	14
WLAN: Wireless Local Area Network	14
UWB: Ultra-Wide Band	15
HART: Highway Addressable Remote Transducer	15
DSSS: Direct Sequence Spread Spectrum	15
TDMA: Time Division Multiple Access	15
IoT: Internet of Things	15
IPv6: Internet Protocol version 6	16
6LoWPAN: IPv6 over Low power Wireless Personal Area Networks	16
GIS: Geographic Information System	19

IDW: Inverse Distance Weighting	22
OK: Ordinary Kriging	27
SK: Simple Kriging	27
UK: Universal Kriging	27
GUIDE: Graphical User Interface Development Environment	31
MAE: Mean Absolute Error	45
MRAE: Mean Relative Absolute Error	45
RMSE: Root Mean Squared Error	45