

UNIVERSIDAD DE ALCALÁ

DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN



Tesis Doctoral

ANÁLISIS DE MEDIDAS DE POTENCIA EN INTERIORES PARA SU
APLICACIÓN EN SISTEMAS DE LOCALIZACIÓN BASADOS EN LA
TÉCNICA DEL FINGERPRINTING.

Autor:

D. Miguel Ángel Navarro Huerga

Directores:

Dr. D. Oscar Gutiérrez Blanco

Dr. D. Felipe Cátedra Pérez

2010

AGRADECIMIENTOS

Una vez que he sido capaz de llegar hasta aquí, no puedo por menos que mostrar mi más sincero agradecimiento a todos aquellos compañeros y amigos del departamento que de una u otra forma siempre han estado a mi lado, con los que he compartido trabajo y esfuerzo y que sin ellos nunca hubiera sido posible finalizar esta tesis que hoy defiendo y poder cerrar así una importante etapa de mi trayectoria profesional que me va a permitir afrontar nuevos retos.

Ante todo agradecer a Oscar, mi director de tesis, todos esos momentos compartidos y que nos han permitido hilar una gran amistad. Gracias Oscar por tu trabajo, por tu extraordinaria capacidad técnica y por tu calidad humana y tu apoyo en los momentos difíciles.

Cómo no, agradecer a Felipe Cátedra sus expertos comentarios y disposición para la realización de este trabajo.

Mención aparte merece Luis Díaz, amigo, matemático e informático. Sin su interés, apoyo y ánimo no hubiese sido posible el arranque de esta tesis.

Por último, no quisiera olvidar al personal del Politécnico de la Universidad de Alcalá, y de las Consultas del Hospital Príncipe de Asturias por permitirme robarles un poco de su tiempo para realizar las mediciones necesarias que han sido la base de este trabajo.

Dedico esta tesis a mi esposa M^a José, a mis hijos, Jorge y Raúl, y a mis hermanos Javier y José Antonio porque en todo momento habéis estado ahí.

RESUMEN

Esta tesis está relacionada con la técnica de localización en interiores denominada *fingerprinting*. La localización en interiores está sufriendo un gran avance en los últimos años debido a una creciente demanda de los servicios de valor añadido en los terminales móviles. Muchos de estos tienen que ver con el posicionamiento contextual del usuario.

La técnica del *fingerprinting* es una de las más utilizadas en los sistemas de localización en interiores. Esta técnica utiliza la relación entre los niveles de potencia recibidos en el móvil de los diferentes puntos de acceso de una red inalámbrica y los niveles de potencia en una serie de puntos, conocidos como huellas, cuya posición es conocida. Ese conjunto de puntos se conoce como *radiomap*.

Por tanto es de gran importancia tener valores precisos de potencia en las huellas con el fin de tener una buena precisión en la localización. Sin embargo debido a las características de la propagación en interiores esto es complicado debido a la aleatoriedad de la señal recibida.

En esta tesis se ha diseñado una pequeña red de localización constituida por unos pocos puntos de acceso y un dispositivo móvil. Con esta red se han realizado diferentes medidas en diferentes escenarios en diferentes situaciones ambientales. Un factor al que se ha prestado especial atención es la presencia humana en los escenarios de medidas. El objetivo ha sido estudiar la influencia de diversos factores sobre la variabilidad de los niveles de potencia medidos en diversas posiciones o huellas, y por tanto como puede variar la precisión de la localización en las diferentes situaciones.

También se ha implementado algoritmos de clusterización para separar los valores de potencia medidos en grupos o *clusters*. La idea es asociar una huella a cada cluster. Entre las aplicaciones que esto tiene está en comprobar si el sistema de localización está bien diseñado viendo que los *clusters* coinciden con las huellas. También se puede diseñar un algoritmo de localización basado en la identificación de los valores de potencia recibidos en el móvil con un determinado *cluster*, y por tanto con una huella.

Para realizar esto se han implementado los algoritmos k-medias y rek-medias y se han aplicado a un conjunto de potencias medido en un determinado *radiomap*.

ABSTRACT

This thesis relates to the technique of indoor tracking called *fingerprinting*. Indoor tracking is having greater advances in the last years due to the ever increasing demand on services for added values in mobiles. Many of these services have to do with the contextual positioning of the user.

The technique known as *fingerprinting* is one of the most widely used procedures in indoor tracking systems. This technique uses the relation between the levels of power being received in the mobile from different access points in a wireless network and the levels of power contained in a series of points, which are called *fingerprints*, the position of these points is definitely known. This array of points receives the name of *radiomap*.

Therefore it is of paramount importance to possess precise values of power within the traces in order to achieve accurate precision while tracking. However and due to the propagation characteristics found indoors this aim could prove complicated as a result of the randomness of the received signal.

In this thesis a rather small net for tracking formed by few access points was designed and a mobile device used. With this net various measurements at different locations and different environmental scenarios were carried out. A factor upon which special attention was focused was the human presence at the sites where the measurements occurred. The objective was to realize a study about the influence of different factors upon the variability of the levels of power measured at various

positions or *fingerprints*, and by the same token how the tracking precision can vary in different situations.

Clustering algorithms have been implemented in order to separate the values of power measured in groups or *clusters*. The idea behind it is to identify each particular cluster with a particular *fingerprint*. Among the applications this technique offers is the capability to test if the tracking system has been perfectly designed just by observing if the clusters coincide with the *fingerprints*. Another possibility is the design of a tracking algorithm based on the identification of the values of power received in the mobile with a particular *cluster*, and therefore with a *fingerprint*.

For this reason the algorithms k-means and rek-means have been implemented and consecutively applied to a group of powers measured over a small *radiomap*, demonstrating that the algorithm rek-means is therefore the most effective one.

ÍNDICE

Capítulo 1: Introducción	1
1.1.- Objetivos	2
1.2.- Estructura de las Tesis	3
Capítulo 2: Estado del arte y revisión de los sistemas de localización en interiores.	5
2.1.Introducción	5
2.2. Componentes comunes de los sistemas de posicionamiento en interiores	6
2.3 Clasificación de los sistemas de posicionamiento en interiores	9
2.3.1 Tecnología de los sensores	9
2.3.2 Técnicas de medida	11
2.3.3 Propiedades de los sistemas de localización	15
2.4 Sistemas de posicionamiento en interiores	17
2.5 Sistemas de posicionamiento en interiores usando lars sin hilos y la técnica del <i>fingerprinting</i>	20
2.5.1 Entorno de interior.	21
2.5.2 Localización <i>fingerprint</i> .	23
2.5.3 Algoritmo de estimación de localización	26
2.5.3.1 Métodos de vecino más próximo	27
2.5.3.2 Métodos de Redes Neuronales.	32
2.5.3.3 Métodos probabilísticos.	35
2.5.3.4 Métodos de Máquina Vectorial de Soporte.	38
2.5.4. Resumen de las prestaciones de los posicionamientos de interiores existentes.	40
2.6 Referencias.	45

VII

Capítulo 3: Características de la señal de propagación en interiores (Aplicación al método del fingerprinting).	49
3.1.- Introducción	49
3.2.- Modelos teóricos de propagación	50
3.2.1.- Tipos de modelos	52
3.2.1.1.- Modelos matemáticos.	52
3.2.1.2.- Modelos Empíricos o estadísticos	53
3.2.1.3.- Modelos Teóricos	53
3.2.1.4.- Modelos deterministas	54
3.2.1.5.- Modelos estocásticos	54
3.2.1.6.- Modelos de caja negra	54
3.2.2.- Validez de un modelo	54
3.2.3.- Cálculo teórico de la propagación	56
3.2.3.1 Campo eléctrico y señal de radio	56
3.2.3.2 Atenuación en Espacio Libre	57
3.2.3.3 Absorciones y multitrayecto Polarización	59
3.2.3.4 Modelo básico de propagación en interiores	61
3.2.3.5 Modelos empíricos de propagación	62
3.2.3.6 Modelo determinista de Ray-Tracing	65
3.2.3.7 Comparación entre un modelo empírico y otro determinista	66
3.3.- Medidas en escenarios reales	71
3.3.1 Caracterización de los escenarios de medidas	72
3.3.2 Elementos básicos del sistema.	73
3.3.3 Condicionamiento ambiental.	75
3.3.4 Descripción de las sesiones realizadas	84
3.3.5. Medidas realizadas	93
3.4.-Referencias	110

VIII

Capítulo 4: Algoritmo de clustering para la selección de huellas.	113
4.1.- Introducción	113
4.2.- Métodos de clusterización	114
4.2.1. Tipos de agrupación o clusters	115
4.2.2. Técnicas de agrupamiento exclusivas	116
4.2.3. Algoritmo k-medias	119
4.2.4. Algoritmo rek-medias	123
4.3. Resultados experimentales	126
4.3.1. Configuración de medidas	126
4.3.2. Resultados	130
4.3.3. Conclusiones	149
4.3. Referencias	150
Capítulo 5: Conclusiones y futuras líneas de trabajo	153
5.1.- Conclusiones	153
5.2.- Futuras líneas de trabajo	154

IX

1. Introducción.

En los últimos años, debido al gran avance de las comunicaciones móviles y a la popularización de la tecnología de la información, han ido apareciendo diversos servicios añadidos basados en la localización de los usuarios en entornos de interiores. De esta forma se están desarrollando sistemas de comunicaciones que permiten desarrollar innumerables aplicaciones que dependen de la posición del usuario. Por ejemplo algunas de las aplicaciones de más interés comercial se basan en el posicionamiento contextual, permitiendo la interacción con distintos objetos que se encuentran en las proximidades del usuario. Otros servicios están relacionados con la seguridad en red, basada en la localización física del usuario; control de acceso, basado en la identificación del mismo, etc.

Por tanto, posicionar al usuario dentro de un edificio y saber en tiempo real la posición que ocupa, es una parte vital en estos sistemas. Los sistemas de localización basados en redes inalámbricas WiFi están teniendo una repercusión muy importante en estos últimos años. Se han ido desarrollando diversas técnicas, una de las más utilizadas y conocidas es la técnica del *fingerprinting*. Sin embargo estas técnicas aun no han alcanzado la precisión de los sistemas de localización en exteriores, sobre todo el GPS. El principal problema que se plantea es técnico ya que la localización en interiores depende de muchos más factores que la localización en exteriores. También existe un componente económico importante, ya que las redes de localización en interiores necesita de un mayor despliegue de infraestructuras, puntos de acceso, sensores, etc.

La técnica del *fingerprinting* se basa en el nivel de potencia recibido por el sistema móvil de cada uno de los puntos de acceso de la red inalámbrica. Uno de los principales problemas que plantea esta técnica, es la variabilidad del nivel de potencia recibido. Esta variabilidad es debida a muchos factores que dependen del entorno, como pueden ser la humedad, los objetos de la estancia, el tránsito humano, etc.

Existe una línea de de investigación muy importante dedicada al estudio de la influencia de estos factores sobre la precisión de los sistemas de localización. Sobre todo en lo que respecta a la presencia humana en los entornos donde se quiere realizar la localización.

1.1. OBJETIVOS

El objetivo principal de esta tesis se basa en observar los efectos ambientales del entorno sobre los valores de potencia recibidos en el dispositivo móvil. Con este fin se pretende diseñar un sistema de localización básico basado en una pequeña red inalámbrica, compuesta por varios puntos de acceso y un dispositivo móvil que recibe los valores de potencia emitidos por los puntos de acceso.

Este sistema de localización se situará en diferentes escenarios en los que las condiciones ambientales sean cambiantes, prestando mayor atención sobre el tránsito humano en dichos escenarios. Se recolectarán una serie de medidas de potencia y se estudiará la variación de la señal recibida en los diferentes casos estudiados.

Otro objetivo que se plantea en esta tesis es recolectar una serie de medidas de potencia recibida en un conjunto de localizaciones, y agrupar los valores de potencia en grupos o *clusters*. Para ello se aplicarán diferentes algoritmos de clusterización como el k-medias y el rek-medias. Con esto se pretende estudiar la posibilidad de aplicar una nueva técnica de localización

basada en la pertenencia a un cluster en lugar de en el cálculo de una distancia mínima.

1.2. ESTRUCTURA DE LA TESIS

Esta tesis está estructurada en cinco capítulos que se detallan a continuación:

-En el capítulo 1 se presenta la localización en interiores y se definen los objetivos y la estructura de la presente tesis.

-En el capítulo 2 se hace un estado del arte de los sistemas de localización en interiores basados en redes inalámbricas.

-En el capítulo 3 en primer lugar se detallan los métodos existen que permiten predecir el nivel de señal recibido en interiores. Se comentan los más importantes, detallando las ventajas y desventajas de cada uno de ellos; y se comparan algunos métodos realizando simulaciones en un escenario en concreto.

Posteriormente se realizarán una serie de medidas, en diferentes escenarios y con diferentes condiciones ambientales con el fin de estudiar la variabilidad del nivel de potencia recibido en el dispositivo móvil.

-En el capítulo 4 se estudiará la posibilidad de utilizar el método de clusterización con el fin de clasificar las huellas. El objetivo es estudiar la viabilidad de utilizar esta clasificación en *clusters* en la técnica del *fingerprintg*.

-Por último en el capítulo 5 se verán las conclusiones alcanzadas en la presente tesis además de exponer las futuras líneas de investigación que abre este trabajo.

2. Estado del arte y revisión de los sistemas de localización en interiores.

2.1. INTRODUCCION.

En este capítulo se revisa lo publicado acerca de sistemas inalámbricos de posicionamiento en interiores, con la finalidad de proporcionar un trasfondo científico a la presente investigación.

Primeramente, en el apartado 2.2 se describen los componentes de los sistemas de posicionamiento en interiores. Seguidamente, en el apartado 2.3 se discute los diferentes medios para clasificar los sistemas de posicionamiento en interiores. En el apartado 2.4. se detallan los primeros sistemas de posicionamiento en interiores comerciales.

Finalmente en el apartado 2.5. se refiere a las diferentes tecnologías y técnicas que emplean los sistemas de posicionamiento en interiores basados en redes inalámbricas WLAN.

2.2. COMPONENTES COMUNES DE LOS SISTEMAS DE POSICIONAMIENTO EN INTERIORES.

Pahlavan et al [1] plantea un diagrama funcional de los sistemas de posicionamiento inalámbricos. El modelo consiste en un cierto número de dispositivos, sensores de localización, un algoritmo de posicionamiento y un sistema de visualización. La figura 2.1 ilustra estos componentes y sus interrelaciones.

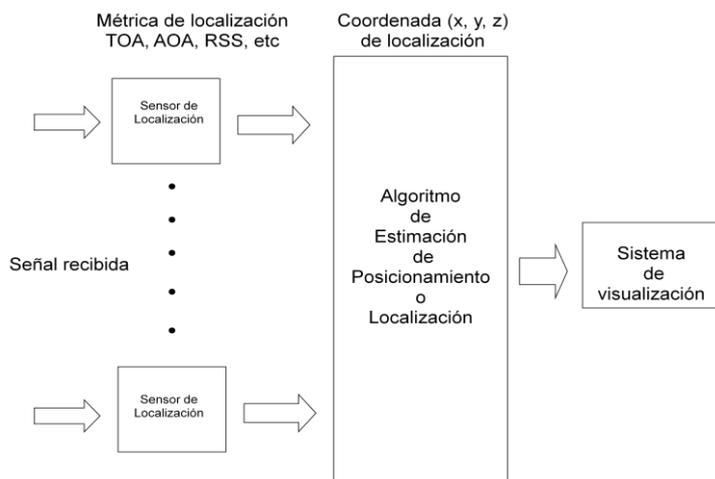


Figura 2.1: Diagrama funcional de un sistema de posicionamiento

Primeramente, los dispositivos sensores de localización detectan las señales transmitidas o recibidas en unos puntos de referencia ya conocidos, usando para ello tecnologías de sensores tales como frecuencias de radio en microondas (RF), infrarrojos o ultrasonidos. La técnica sensora (que puede estar basada en el retardo de la señal, la dirección de llegada o ángulo de llegada, la frecuencia, o el nivel de potencia de la señal) convierte la señal detectada en métricas de localización que son el tiempo de llegada o retardo de la señal (TOA), el ángulo de llegada (AOA), la fase de la señal portadora de llegada (POA) o la potencia de la señal recibida (RSS) [1]. Dado un conjunto de puntos de referencia, la posición relativa de la estación móvil puede ser

deducida a partir de la distancia o de la dirección de estas métricas de localización.

Por otro lado, las características de la señal tales como la RSS en una ubicación concreta pueden formar un patrón único de esta localización. De ahí que el algoritmo de posicionamiento procese la métrica de localización y estime la información de localización utilizando enfoques tales como procesado de señal [2], la distancia Euclidea [3], redes neuronales [4] o enfoques probabilísticos [5]. Finalmente, el sistema de visualización convierte la información sobre localización en un formato apropiado para el usuario final.

Por otro lado, un sistema de localización puede ser contemplado desde el punto de vista de la ingeniería de software utilizando un modelo de capas (análogo al del modelo OSI), que es lo que propone Hightowwer et al [6]. El armazón de estrato de localización es un modelo de ingeniería de software por capas que subdivide el problema de posicionamiento en problemas de investigación más pequeños. Su finalidad es facilitar el desarrollo de futuros sistemas de computación utilizando para ello información de localización. El estrato de localización extraído de [6] en la tabla I está diseñado basándose en las propiedades de los sistemas de posicionamiento. Estos son: los tipos de medida fundamentales, los enfoques de combinación de medidas, las consultas de relaciones entre objetos, la conservación de incertidumbre y la aplicación de las actividades de usuario.

Sin embargo, este modelo abstracto se encuentra en una fase inicial y todavía no dispone de especificación para las interfaces entre capas.

La tabla I resume la descripción de cada capa. En [6] se pueden encontrar más detalles. Basándonos en este estrato de protocolos, este tesis se centra fundamentalmente en la segunda capa.

6. Actividades	Sistema, tal como máquina de aprendizaje, para categorizar en actividades toda información de contexto disponible, incluyendo localización
5. Fusión de Contexto	Sistema para fusionar datos de localización con otros datos de información de contexto distintos a los de la localización, tales como datos personales, color, temperatura, nivel de luz, etc.
4. Disposiciones	Máquina de razonamiento probabilístico sobre las relaciones (proximidad, contiene informaciones geométricas) entre dos o más objetos.
3 Fusión	Método general para combinar conjuntos de medidas y lograr una representación probabilística de posiciones y representaciones de objetos. Exporta una petición (<i>query</i>) o interfaz de eventos que proporciona las posiciones inmediatas de los objetos y las incertidumbres asociadas. Se puede incluir, también, otro tipo de información compleja (velocidad, aceleración, históricos.)
2. Medidas	Algoritmo que transcribe datos primarios de sensor en tipos canónicos de medida conjuntamente con la representación de la incertidumbre, basándose en el modelo de sensor que la ha creado.
1. Sensor	Dispositivos hardware y software del sensor para detectar la variedad de fenómenos físicos y lógicos.

Tabla I: Resumen modelo de capas de localización

2.3 CLASIFICACIÓN DE LOS SISTEMAS DE POSICIONAMIENTO EN INTERIORES.

Se puede categorizar los sistemas de posicionamiento en interiores basándose en la tecnología de sus sensores, las técnicas de medida o las propiedades del sistema.

La tecnología de los sensores se refiere a los tipos de señales usados por los sensores, mientras que las técnicas de medida se refieren a los métodos y métricas usados en los sensores de localización.

Por otro lado, Hightower and Borriello [7] sugieren una clasificación de los sistemas de posicionamiento basada en las propiedades de dicho sistema, que son independientes de las técnicas de los sensores y de las tecnologías de medida. Su clasificación sugiere unas líneas que nos guíen en la evaluación de los sistemas de posicionamiento. Sin embargo, algunas propiedades no son aplicables a todos los sistemas de posicionamiento.

2.3.1 TECNOLOGÍA DE LOS SENSORES.

Basándose en la tecnología de los sensores que se utilizan, los sistemas de posicionamiento heredan ciertas características y limitaciones del tipo de señal del sensor. El retardo de la propagación, la difracción, la reflexión y la dispersión son características básicas que afectan a todos los tipos de señales. El rango efectivo, la anchura de banda disponible, la normativa a aplicar, las interferencias, las restricciones de potencia, la seguridad y el coste son limitaciones tecnológicas [8].

Las señales inalámbricas usadas comúnmente para los sistemas de posicionamiento en interiores de infrarrojos, radio frecuencia y ultrasonidos.

Hay que señalar que otras tecnologías tales como rangos de láser, análisis de escenas y sistemas basados en la inercia son también válidas para sistemas de posicionamiento en interiores pero están fuera del alcance de este estudio.

Seguidamente describimos brevemente las tres principales tecnologías de sensores [8].

Infrarrojos.: La señal de infrarrojos tiene las mismas propiedades que la luz visible. No puede pasar a través de paredes u obstáculos. De ahí que tenga un rango bastante limitado de entornos interiores. Sin embargo, la velocidad de propagación es alta, aproximadamente $3 \cdot 10^8$ m/sg. Por eso requiere una circuitería más sofisticada que las señales de ultrasonidos. La iluminación interior interfiere con este tipo de señal y causa problemas de precisión en la detección. Generalmente tiene un rango de alrededor de 5 metros. Los dispositivos de infrarrojos son generalmente pequeños comparados con los dispositivos de ultrasonidos[9].

Radio frecuencia.: La señal de radio frecuencia (RF) puede penetrar la mayoría de los materiales de la construcción, de ahí que tenga un excelente comportamiento en los entornos de interiores. La velocidad de propagación es también alta. Aproximadamente, $3 \cdot 10^8$ m/sg. Además, hay frecuencias que no requieren licencias disponibles para su uso. Este tipo de señal goza de un amplio rango comparado con los infrarrojos y con los ultrasonidos.

Ultrasonidos.: Aunque los ultrasonidos operan en bandas de baja frecuencia (típicamente 40 KHz) comparado con las otras dos tecnologías de señales, posee una buena precisión para la detección de localizaciones a baja velocidad de propagación del sonido (343 m/s). Las ventajas de los dispositivos de ultrasonidos son su simplicidad y su bajo coste. Sin embargo, los dispositivos de ultrasonidos no penetran muros, se reflejan ante la mayoría de los obstáculos de interiores. Además, estos dispositivos tienen un corto

rango de alcance, entre 3 y 10 m pero tienen 1 cm de resolución en la medida de distancia. También hemos de tener en cuenta que la temperatura de operación influye en el rendimiento de los dispositivos de ultrasonidos [10].

En la tesis [10] se propone un sistema de localización WiFi determinístico basado en un mapa obtenido a priori mediante un proceso de entrenamiento además de un sistema de localización probabilístico que emplea la fusión de la medida de la señal WiFi con la información procedente de los ultrasonidos. Con estos sistemas se consigue un posicionamiento global de un robot desde el primer momento de funcionamiento del mismo sin utilizar sensores adicionales en el entorno ya que la funcionalidad principal de los sensores WiFi es la de establecer una red de comunicaciones WLAN (Wireless Local Area Network, “Red de Área Local Inalámbrica”)

2.3.2 TÉCNICAS DE MEDIDA.

Los sistemas de posicionamiento inalámbrico, además de por la tecnología de los sensores, pueden ser catalogados también por las técnicas de medida usadas para deducir la posición de las estaciones móviles. Las principales categorías de esta clasificación mediante las técnicas de medida están basadas en la medida de la distancia, el ángulo, la huella digital de radiofrecuencia -patrón de localización- o *fingerprint* y cualquier combinación de dichas categorías. La figura 2.3 muestra la clasificación basada en la tecnología y sus tipos.

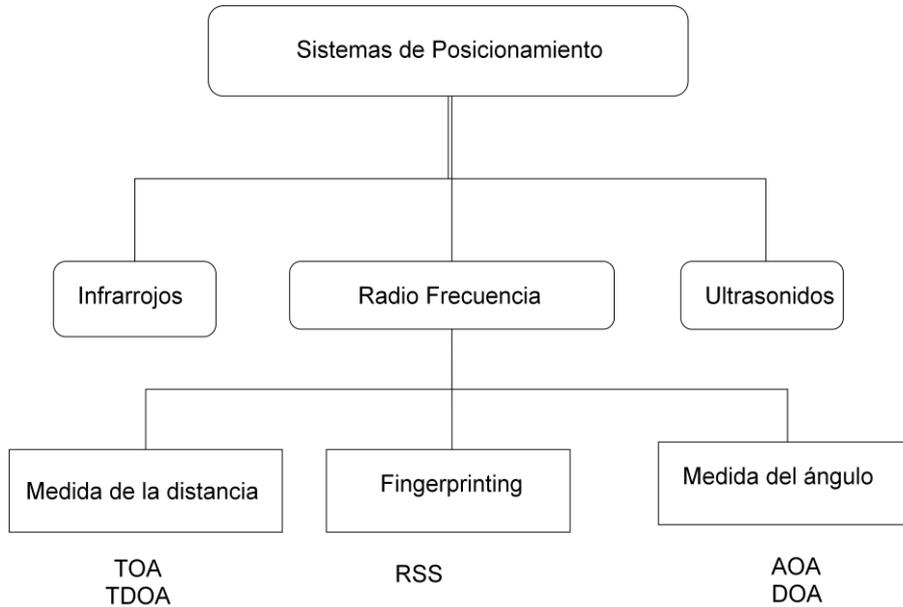


Figura 2.3: Clasificación de los sistemas de posicionamiento

Generalmente, se llama *lateración* a la técnica de medida de distancias, mientras que se llama *angulación* a la técnica de medida de ángulos. Ambas son subcategorías de la *triangulación*, [7] que utiliza la geometría para determinar la localización. Además de estas dos grandes categorías, la proximidad, el análisis de escena y otras características o geometrías tales como el nivel de luz o la temperatura pueden ser usadas como métricas en la localización [8]. Por ejemplo, la técnica de proximidad usa una ubicación conocida, cercana al objeto, para determinar la localización, mientras que el análisis de la escena infiere la localización basándose en la observación pasiva de las características de una escena. La medida de la distancia es la métrica más frecuentemente usada para la estimación de la localización. La distancia ha de ser estimada a partir de la atenuación de la potencia de la señal basándose en la pérdida de camino y el tiempo de vuelo (TOF) de la señal. Las tres técnicas bien conocidas, ángulo de llegada (AOA), diferencia de tiempo de llegada (TDOA) y *fingerprinting* son discutidas en este capítulo.

Las dos primeras técnicas han sido estudiadas extensamente para los sistemas de posicionamiento [2]. Son técnicas adecuadas para sistemas con línea de visión directa, pero presentan problemas o requieren cálculos complejos en canales de radio con ruido, interferencia y multicamino. En entornos de interiores, la estación móvil está rodeada de objetos dispersos que dan como resultado múltiples ángulos de recepción de la señal. Por otra parte, la distancia entre el transmisor y el receptor es generalmente más corta que la resolución del tiempo que puede ser medida por el sistema. De ahí que los enfoques del AOA y del TDOA sean impracticables para entornos de interiores. Posteriormente la técnica del *fingerprinting* ha ganado más importancia debido a su simplicidad comparada con la de los dos primeros sistemas de posicionamiento en interiores.

A continuación pasamos a describir cada una de estas técnicas de medida.

Medida de la distancia basada en el tiempo de llegada: las técnicas de tiempo de llegada (TOA) y la diferencia de tiempo de llegada (TDOA) dependen de la precisión de la sincronización entre transmisor y receptor. Esta técnica está basada en el cálculo del tiempo de vuelo TOF (Time of Flight) o tiempo de ida y vuelta de la señal de propagación, con el fin de estimar la distancia entre el transmisor y el receptor. Este método usa mediciones del tiempo transcurrido en una transmisión entre un sensor y un lector, usando la velocidad de propagación estimada de una señal típica a través del medio. Como consecuencia de que se utilizan valores de tiempo la exactitud del reloj es mucho más importante que en los métodos anteriores. De ahí que una sincronización precisa sea también muy importante en estos sistemas. Este tipo de técnica requerirá un reloj de alta precisión en el sistema de comunicación. Combinando como mínimo tres distancias desde tres posiciones de referencia, la triangulación puede ser utilizada para estimar la localización de la estación móvil. Las técnicas basadas en TDOA son más prácticas [2]. En estos sistemas,

el transmisor emite un pulso acústico (ultrasonido) junto con un pulso de radio, mientras que el receptor compara el tiempo de llegada de ambos pulsos. Conociendo la velocidad de vuelo de cada una de las señales, la diferencia de tiempo entre las llegadas revela la distancia entre el emisor y el receptor. Ejemplos de sistemas de sensores de localización que utilizan el tiempo de vuelo TOF son GPS [7], Active Bats [11] y Cricket [12]. Además de estas técnicas basadas en retardo de tiempo, la distancia entre transmisor y receptor también puede ser determinada a partir de la atenuación de la potencia de la señal y la medida de la distancia directa como *dead reckoning*, método de localización que se utiliza actualmente en los sistemas como GPS, para mantener la estimación de la posición cuando se producen pérdidas de señal de los satélites.

Medida del Ángulo: Las técnicas del ángulo de llegada (AOA) o de la dirección de llegada (DOA) localizan la estación móvil por medio de la determinación del ángulo de las señales incidentes. Usando sencillas relaciones geométricas, la estimación de la localización puede ser calculada por la intersección de dos líneas de orientación (LOBs), las cuales están formadas por una línea radial desde el transmisor al receptor [2]. En un plano bidimensional, al menos se necesitan dos puntos de referencia para la estimación de la localización. Sin embargo, esta técnica requiere el uso de antenas direccionales y arrays de antenas para medir el ángulo de incidencia. En síntesis, es difícil medir el AOA en una estación móvil.

Fingerprinting o correspondencia de patrón de localización: Esta técnica sólo requiere generalmente la medida de la potencia de la señal recibida, u otras características no geométricas, en varias localizaciones. A partir de estos valores se crea una base de datos de localizaciones *fingerprint*. Esta base de datos es conocida en ocasiones como radio-map. Para estimar la localización móvil, el sistema necesita primero medir la potencia de la señal recibida en posiciones concretas y después buscar el patrón de *fingerprint* con

el que concuerde más en la base de datos. Esta técnica no requiere que la estación móvil vea al menos tres estaciones base o puntos de acceso para determinar la localización. La desventaja de esta técnica es que requiere mucho tiempo la recolección exhaustiva de datos de un determinado área.

2.3.3 PROPIEDADES DE LOS SISTEMAS DE LOCALIZACIÓN.

Los sistemas de posicionamiento se pueden clasificar de acuerdo a un conjunto de propiedades que son independientes de los sensores y de las técnicas de medida utilizadas. En la tabla II se listan estas propiedades basadas en la investigación sobre sistemas de localización [7]. Estas propiedades, vistas como otra forma de clasificar, pueden ser utilizadas para caracterizar o evaluar los sistemas de posicionamiento [7]. Se añade una propiedad basada en el tipo de servicios que pueden proporcionar los sistemas de posicionamiento.

Posición física versus localización simbólica	<ul style="list-style-type: none"> - La posición física o la referencia abstracta está basada en el etiquetado analítico de las coordenadas tales como latitud, longitud y altitud. - La localización simbólica o la referencia del mundo real está basada en la proximidad de objetos conocidos o ideas abstractas de localización.
Referenciación absoluta o relativa	<ul style="list-style-type: none"> - Los sistemas de referenciación absoluta comparten parrillas de referencia simples o unificadas. - Los sistemas de referenciación relativa tienen su propia estructura de parrilla de referencia para cada localizador
Computación remota o local	<ul style="list-style-type: none"> - Los sistemas de computación remota estiman la localización (basada en Móvil o Red) de la estación móvil por

Capítulo 2: Estado del arte y revisión de los sistemas de localización en interiores

	<p>red de sistemas de localización o servidor de posicionamiento <i>backhaul</i>. Esto también se conoce como sistema basado en red.</p> <p>- Los sistemas de computación local estiman su propia localización, p.e. autoposicionamiento. También se conoce esto como sistema basado en móvil.</p>
Asistido por Red o por Móvil	Esto indica la parte donde se realiza la computación o estimación de la localización.
Capacidad de reconocimiento	Algunos sistemas de posicionamiento heredan la capacidad de reconocimiento que pueden clasificar o identificar objetos locales tales como el ID global o la denominación.
Precisión y nitidez	<p>- La precisión de la localización es generalmente reportada en metros como error en la distancia en la localización estimada que se desvía de la correcta localización</p> <p>- La nitidez de la localización es generalmente presentada en porcentaje de estimación correcta con cierta precisión</p>
Coste y tiempo	<p>- El coste de despliegue de un sistema de localización consiste en el coste de instalación, el coste de infraestructura, el coste del terminal de usuario o del dispositivo y el coste del tiempo</p> <p>- El tiempo de despliegue del sistema es el tiempo de instalación y el tiempo de para estimar la localización.</p>
Escalabilidad	- El alcance del espacio, tiempo, frecuencia y complejidad del sistema de posicionamiento pueden limitar la escalabilidad.

Seguridad y Privacidad	<ul style="list-style-type: none">- La Seguridad evita el uso no autorizado de la información sobre localización.- La Privacidad asegura la anonimidad del usuario.
Categorías de Servicio	Las aplicaciones basadas en localización pueden ser clasificadas bien por <i>Business-to-Consumer</i> (B2C) o por <i>Business-to-Business</i> (B2B) [13].

Tabla II: *Propiedades de los Sistemas de Localización*

2.4 SISTEMAS DE POSICIONAMIENTO EN INTERIORES.

Se pueden encontrar investigaciones excelentes y exhaustivas sobre sistemas de posicionamiento en [7] y con un enfoque especial para los sistemas de posicionamiento en interiores en [8]. De ahí que en este capítulo no entremos en grandes detalles de cada uno de las vanguardias en este tipo de sistemas de posicionamiento. A modo de ejemplo se revisa un subconjunto de estos sistemas. A continuación se resumen las principales características de estos sistemas.

El sistema de localización de *Active Badge* [9] es uno de los primeros sistemas de posicionamiento de interiores fundamentado en el uso de infrarrojos. Un servidor central determina las localizaciones del usuario usando para ello sensores para recoger periódicamente las señales transmitidas o solicitadas desde las etiquetas (*badge*) portadas por el usuario móvil. La señal de infrarrojos de cada usuario tiene un identificador único. La determinación de la localización está basada en la proximidad de la etiqueta al sensor, de ahí que sólo esté disponible la información para la localización de forma simbólica

dependiendo del tamaño de la habitación. El sistema tiene un rango limitado y la señal de infrarrojos es susceptible a la interferencia de la luz solar o de la luz fluorescente [7].

El segundo sistema de localización, llamado *Active Bat* [11] mejora la precisión sobre el sistema *Active Badge* utilizando señales combinadas de radio y ultrasonidos. La distancia medida (usada para el cómputo de la lateración en el controlador centralizado) se calcula a partir del tiempo de visión de la señal de ultrasonido. El sistema de localización consiste en una serie de receptores montados en el techo que detectan la señal de ultrasonidos desde la etiqueta del *Active Bat* que responde a un paquete de solicitud RF desde el controlador centralizado. Los receptores montados en el techo, que están conectados al controlador centralizado por medio de una red serie cableada, calculan la medida de la distancia comenzando desde el tiempo en el que reciben una señal de reset en la red cableada hasta el tiempo en que reciben un impulso de ultrasonidos desde el sensor móvil(bat). En este caso la precisión y la nitidez son excelentes, se obtiene el 95% de las localizaciones en un radio de 9cm.

El sistema de localización *SpotON ad hoc* [14] es otro sistema de posicionamiento que usa la medida basada en distancia, donde ésta es deducida a partir de la atenuación de la potencia de la señal en lugar de tiempo de visión, como en el caso anterior. Los diseñadores del sistema combinan a la vez las ideas de una red ad hoc y de la localización de objetos. Cada objeto a localizar dispone de una etiqueta RF. La lateración ad hoc se lleva a cabo usando la distancia estimada entre etiquetas, en lugar de utilizar la distancia entre sensores conocidos y estaciones base. De ahí que el sistema pueda proporcionar tanto referencias relativas como absolutas. Un racimo dinámico de etiquetas permite que cualquier nodo participante explote la correlación de múltiples medidas y mejore la precisión de la localización ya que el racimo de etiquetas se hace más denso [14].

El Sistema *Cricket* de soporte de localización [12] es un sistema de localización diseñado con cuatro objetivos: privacidad, descentralización, bajo coste y granularidad a tamaño habitación. Se dice que el sistema es independiente de la tecnología de redes de datos. No tiene servidor centralizado, de ahí que el dispositivo móvil tenga que calcular su propia localización usando tanto tecnologías de ultrasonidos como de RF. El dispositivo móvil mide la señal de ultrasonidos para calcular el rango con técnicas TDOA mientras que la señal RF es usada para sincronización y para identificar el período de la señal de ultrasonido. Cada habitación es equipada con una baliza que transmite un impulso RF con un único ID para una habitación concreta. El enfoque basado en móvil asegura la privacidad. Sin embargo, hay errores potenciales desde las interferencias de la baliza RF que causan confusión entre dos habitaciones adyacentes.

El sistema de posicionamiento *PinPoint 3D-iD* [15] es un producto comercial basado en RF para interiores. Determina la localización de una etiqueta por emisión continua de una señal desde un conjunto de antenas cuyas posiciones son conocidas. Al recibir la señal, la etiqueta retransmite inmediatamente el mensaje por desfase a otra frecuencia de radio y lo codifica con su propio ID. El controlador del sistema mide las distancias múltiples desde la fila de antenas usando el tiempo de retorno(round-trip time) de RF y ejecuta la multilateración. La señal desde la célula del transmisor (llamada célula controladora) es una señal de amplio espectro que opera a 2,4 GHz con una anchura de banda de 40 MHz, mientras que la tarjeta transmite una señal de respuesta a 5,78 GHz. El sistema tiene un rango de 30 m y de 1 a 3 m de precisión. El sistema requiere varios controladores de células por edificio y su hardware es caro.

Estos trabajos, pioneros en esta área, tienen algunas desventajas tales como la limitación de las señales de los sensores de infrarrojos o de ultrasonidos, que no pueden penetrar los muros y puertas habituales de los

interiores de la mayoría de edificios. El coste de la instalación de la infraestructura de sensores y de las fichas o tarjetas para la mayoría de estos sistemas se hace significativo para un edificio con muchas habitaciones u oficinas. Hay que señalar que la medida basada en ángulos o direcciones no fue usada en ninguno de estos sistemas debido a el denso multicamino en el interior de los edificios. Sin embargo hay que indicar que estos sistemas de posicionamiento han demostrado su éxito empíricamente y que adolecen de explicación teórica sobre su sistema y prestaciones.

2.5. SISTEMAS DE POSICIONAMIENTO EN INTERIORES USANDO LANs SIN HILOS Y LA TÉCNICA DEL *FINGERPRINTING*.

Este capítulo revisa los sistemas de posicionamiento en interiores más relevantes, basados en la recepción de una señal RF, que pueden ser usados tanto para localizar objetos estacionarios como para hacer seguimiento de usuarios móviles.

El impresionante crecimiento de los LANs (WLANs) sin hilos basadas en IEEE 802.11 en los últimos años, nos permite augurar un interesante futuro para la técnica de localización basadas en la técnica del *fingerprinting*. Como consecuencia de su gran expansión nos puede permitir ahorrar el coste dedicado a la infraestructura. La mayor parte de las veces, utiliza señales de radiofrecuencia que pueden penetrar la mayoría de los materiales de interior, teniendo como resultado un rango más amplio y reduciendo la cantidad de puntos de acceso requeridos para los fines de localización. Debido a que la amplitud de la señal recibida *RSS* puede ser medida por todas las tarjetas de interfaz de red inalámbrica WLAN, no se requiere tarjeta o etiqueta adicional para algunos de los ordenadores portátiles actuales o PDA ya que disponen de interfaz IEEE 802.11 incorporada. El sistema es bastante flexible debido a que

los diseñadores del mismo pueden elegir entre tener servidor de posicionamiento centralizado o dejar que el móvil determine su propia posición. Sin embargo, la técnica *fingerprinting* requiere una fase de entrenamiento (fase off-line) para recopilar las localizaciones de las huellas de todas las posiciones del área operativa con anterioridad al verdadero despliegue (fase on-line).

Después de un cierto número de estudios empíricos y de viabilidad, tales como los [3, 16], el reciente desarrollo se ha dirigido hacia la mejora de los algoritmos de estimación de la localización y las prestaciones del sistema [17,5, 4]. Las técnicas habituales de máquinas de aprendizaje, tales como redes neuronales y máquinas de soporte vectorial (SVM) han sido introducidas para mejorar la prestación del método del *fingerprinting* basado en la RSS.

La discusión que sigue está dividida de acuerdo a los componentes de sistemas de posicionamiento esbozados en el sección II A. Primero, se tratan los efectos del entorno en las señales RF tales como el canal de radio y la presencia del usuario. Después, se explica la forma común de localización *fingerprinting* y su relación con la posición física. En tercer lugar, se revisa un cierto número de algoritmos de estimación de localización. Para terminar, se comparan las prestaciones de los sistemas de posicionamiento existentes

2.5.1 ENTORNO DE INTERIOR.

El entorno de interior tiene propiedades únicas que influyen en las señales de radiofrecuencia usadas por los sensores de los sistemas de posicionamiento. El fenómeno prominente es el efecto multicamino, que domina el comportamiento de las señales recibidas por todos los sistemas de posicionamiento. Aunque hay varios estudios sobre la propagación de radio en interiores y su modelización, este capítulo discute sólo los estudios enfocados a

los sistemas de posicionamiento en interiores. El estudio [19] de Carnegie Mellon University mostró los resultados de mediciones de la intensidad de señal desde una posición fija en el tiempo de una estación WLAN en el interior de un edificio de oficinas. Se llevaron a cabo diferentes periodos de medida, para determinar la distribución de la potencia de la señal recibida por WLAN. La conclusión señalaba que dado que la media, la mediana y la moda de los datos recogidos en una simple localización estaban muy próximos entre sí, la distribución era lognormal. Además de la distribución, la relación existente entre el rango y la desviación estándar evidencia dependencia lineal con la distancia entre receptor y transmisor, a mayor distancia mayor desviación estándar. Su estudio también mencionaba brevemente el efecto de la hora del día pues aparecían diferencias que se pueden obviar entre las diferentes horas del día y sus correspondientes datos de las señales recibidas.

Otro estudio, [17] también discute brevemente la distribución de la potencia de la señal recibida. Aunque su duración de medida sea algo más corta comparada con la del estudio de Carnegie Mellon, el resultado contradice la creencia tradicional de la potencia de la señal recibida distribuida acorde con una lognormal. Los autores en [17] señalan que la mayoría de las distribuciones de las medidas de la señal son multimodales con un modo dominante y asimétrico.

Otro factor importante que afecta a la señal recibida es el cuerpo del usuario que puede bloquear el paso de la señal durante la operación. El agua, que tiene una frecuencia de resonancia de 2,4 GHz y es parte significativa del cuerpo humano atenúa en gran parte la potencia de la señal WLAN [17, 18]. En algunas aplicaciones basadas en localización tales como la robótica y otros servicios no relativos a lo humano, el efecto de la presencia del usuario puede ser despreciado. El sistema de RADAR [3] sugiere que la orientación del usuario afecta a la media de la señal recibida por el punto de acceso bloqueado por el usuario. De ahí que la orientación deba ser incluida en la información

sobre localización. El movimiento del ser humano en el interior del edificio crea efectos aleatorios de propagación de radio dentro del edificio [16]. Otros factores no controlables como son la temperatura, el movimiento del aire o las interferencias desde otros dispositivos funcionando en la misma frecuencia también son causa de que la señal recibida en una localización concreta fluctúe a lo largo del tiempo [16]. Sin embargo, en lo publicado sobre este tema, no hay una buena caracterización de las propiedades de la RSS relacionadas con las aplicaciones de posicionamiento en interiores.

2.5.2 LOCALIZACIÓN *FINGERPRINT*.

El método de localización *Fingerprint* basado en características de la señal RF tales como RSS es la base de la representación de una única posición o localización. Se crea a partir de la suposición de que cada posición o localización en el interior de un edificio tiene una única firma RF [1]. Generalmente una huella o *Fingerprint* F está etiquetado conjuntamente con una información de localización L . Las localizaciones *Fingerprints* y sus etiquetas (p.e. información de localización) son mantenidas en una base de datos y usadas durante la fase on-line para estimar la localización. La etiqueta y *huella* son generalmente denominadas como la tupla ($F \mid L$). La medida recogida en la base de datos durante la fase on-line es llamada *conjunto de entrenamiento*.

Battini et al [4] señala que la información sobre localización de interior L de localización puede ser grabada en dos formas, ya sea como una tupla de coordenadas o bien como una variable de indicador. La tupla de coordenadas reales puede variar de una dimensión a cinco dimensiones, incluyendo el espacio de tres dimensiones y dos variables de orientación expresadas en coordenadas esféricas [4]. Por ejemplo, una información local de un sistema de dos dimensiones con orientación puede ser expresada como una tripleta:

$$L = \{ (x; y; d) \mid x, y \in \mathcal{R}^2; d \in \{\text{Norte, Este, Sur, Oeste}\} \} \quad (2.1)$$

En el caso de la variable del indicador, el alcance de la localización cubre un área más amplia tal como una habitación. La variable de indicador representa solo una granularidad gruesa tanto si el objeto está en el interior o en el exterior del área. Un ejemplo de ello se encuentra en [4] como $L = \{ -1, 1 \}$. Se dice que los sistemas de localización de interiores que usan coordenadas resuelven un problema de regresión, mientras que los sistemas que usan variables de indicador resuelven problemas de decisión o de clasificación.

Es comúnmente reconocido que el nivel de potencia o RSS es la firma RF más simple y efectiva para la localización *fingerprint* porque su legibilidad está disponible en todas las tarjetas de interfaz WLAN. [3] halló que la RSS era más dependiente de la localización que el ratio señal ruido (SNR) porque la componente de ruido es algo aleatoria por naturaleza. Sin embargo, el mismo RSS fluctúa a lo largo del tiempo para cada punto de acceso y localización. Cada elemento RSS puede ser considerado como una variable aleatoria, de ahí que pueda ser capturado registrando sus parámetros estadísticos descriptivos, aproximando su distribución o manteniendo el conjunto de medidas completo. Estos enfoques de la representación RSS dan como resultado los diferentes procedimientos para los algoritmos de estimación de la localización del siguiente subcapítulo. Independientemente del enfoque, la localización *fingerprint* es generalmente representada como una serie o vector de potencia de señal (variables aleatorias) recibida en cualquier posición del área de servicio basada en la localización. El tamaño del vector está determinado por el número de puntos de acceso que pueden ser escuchados.

Para crear un base fingerprint, o radio-map, tal como la [3, 16], para cada posición se toma una muestra de vectores de potencia dentro de una ventana temporal. Después es calculado la RSS media de cada punto de acceso

y grabada como un elemento de la localización *fingerprint*. Para cada área que puede recibir señales desde los N puntos de acceso, la localización *fingerprint* puede ser expresada como un vector de elementos RSS de media ρ_i

$$F = (\rho_1, \rho_2, \dots, \rho_N)^T \quad (2.2)$$

La información *fingerprint* extra, tal como la desviación estándar para cada elemento RSS, tal como se sugiere en [21], puede ser añadida a la localización *fingerprint* como otro vector de desviación estándar

$$D = (\sigma_1, \sigma_2, \dots, \sigma_N)^T \quad (2.3)$$

Un enfoque alternativo de la localización *fingerprint* es investigado en [17, 5] en el cual se estima la localización en función de la distribución de probabilidad como la firma RSS en una posición dada. La localización *fingerprint* se convierte en una distribución de probabilidad condicional de la forma $P(F | L)$ donde F significa el vector de observación de RSS y L significa la información de localización. La probabilidad condicional $P(F | L)$ es llamada *función de verosimilitud* porque proporciona la probabilidad o densidad de probabilidad de la ocurrencia del vector RSS dada la información de localización conocida [5]. Con esta forma de localización *fingerprint*, se puede utilizar la regla de Bayes para estimar la localización.

Los ejemplos de localización *fingerprint* explicados más arriba sugieren que hay dos formas de modelizar la relación o dependencia entre la información de localización y la firma RSS. La primera forma, en la cual la información de localización está ligada al valor constante del vector RSS medio es llamada *enfoque determinístico*. La segunda forma, en la cual es explotada la dependencia probabilística, recibe el nombre de *enfoque probabilístico*.

Además de la localización *fingerprint*, las muestras de la localización *fingerprint* medidas durante la fase on-line es también importante para el sistema que supervisa el objeto móvil. El intervalo de tiempo y el número de muestras deberían ser seleccionadas apropiadamente para representar la localización *fingerprint* para la aplicación móvil.

2.5.3 ALGORITMO DE ESTIMACIÓN DE LOCALIZACIÓN.

Los algoritmos de estimación de localización o algoritmos de posicionamiento son procedimientos que explotan la dependencia entre la información de la localización y la base de la localización *fingerprint*, con la finalidad de determinar una posición o localización a partir de las señales RSS. Los ejemplos de algoritmos de estimación de localización más sencillos son el método de selección de la estación base con la intensidad de la señal más fuerte y el método de selección aleatorio. La selección de la estación base con la intensidad de la señal más fuerte considera que la posición actual del usuario está más cercano a la estación base que recibe la potencia de señal más fuerte, mientras que la selección aleatoria reporta la posición del usuario aleatoriamente desde un conjunto de posiciones conocidas [3]. Es obvio que estos dos algoritmos pueden no proporcionar resultados satisfactorios. Los algoritmos más complejos pueden aprovecharse de la dependencia entre la RSS de la localización *fingerprint* y de la información de localización y pueden proporcionar mejor precisión y granularidad en la información de la localización.

Desde una perspectiva de máquina de aprendizaje, los algoritmos de posicionamiento son clasificadores de patrones, porque son procedimientos que pueden separar muestras de patrones en diferentes clases [5]. Cada clase se refiere a una clase de vectores RSS por medio del aprendizaje de los ejemplos previos de las firmas o *fingerprint* RSS dependientes de la localización. Los

datos previos RSS o conjunto de entrenamiento, son utilizados para calibrar los modelos de estimador que pueden relacionarse automáticamente con las localizaciones *fingerprint* y la información de localización. En la documentación revisada, los algoritmos de posicionamiento pueden ser clasificados como de tipo determinístico y probabilístico, basados en los enfoques que modelan la relación entre la localización *fingerprint* y la información de localización tal como se discutió en el subcapítulo previo. El tipo de algoritmo determinístico es aquel que está basado en los clasificadores de mayor vecindad y en los clasificadores a base de redes neuronales. El tipo de algoritmo probabilístico es aquel que está basado en la inferencia bayesiana y en la teoría de aprendizaje estadístico, tal como la máquina de vector de soporte (SVM). A continuación se discute los principales algoritmos de sistemas de posicionamiento en interiores.

2.5.3.1 Métodos de vecino más próximo.

Los métodos de vecindad más próxima son algoritmos determinísticos porque requieren solo un conjunto de localización *fingerprint* que incluye vectores medios y vectores de desviación estándar de RSS. Con la finalidad de determinar la localización, una forma de función discriminante se utiliza para clasificar la muestra de *fingerprint* RSS en una posición. Estos métodos de vecindad más próxima son también llamados métodos *basados en caso* [5] porque clasifican cada posición en un caso o clase.

El vector de media o de esperanza RSS, es un centro de masa que representa cada clase de una localización *fingerprint*. El algoritmo básico para el clasificador de vecindad más próxima es el que selecciona la clase o caso basándose en la proximidad de una muestra *fingerprint* al centro de masa de esa concreta localización *fingerprint*. Supongamos que existe un conjunto de l localizaciones *fingerprint* denominado como $\{ F1, F2, \dots ; Fl \}$ y cada

fingerprint tiene una correspondencia uno a uno con un conjunto de posiciones $\{ L1, L2 \dots ; L_l \}$. Una muestra de un *fingerprint* medido durante una fase on line, llamada S , puede ser otra media o esperanza del vector RSS de una ventana pequeña de muestras RSS.

Considerando que un sistema de posicionamiento de interiores solo considera la media RSS a partir de N puntos de acceso como una localización *fingerprint*, la muestra del vector RSS es $S = (s_1, s_2, \dots, s_N)^T$ y cada localización *fingerprint* en la base de datos puede ser expresada como $F_i = (\rho_1^i, \rho_2^i, \dots, \rho_N^i)^T$.

La métrica más simple y más cercana es una medida de distancia en el espacio de la señal y se denomina como función $Dist(\cdot)$ [16]. De esa forma, el procedimiento del algoritmo de la vecindad más próxima es expresado tomando el *fingerprint* j que tenga la distancia más corta

$$Dist(S, F_j) \leq Dist(S, F_k), \forall k \neq j \quad (2.4)$$

Una distancia ponderada generalizada L_p [16] puede ser utilizada para calcular diferentes formas de distancia en el espacio de señales como:

$$L_p = \frac{1}{\mathcal{M}} \left(\sum_{i=1}^N \frac{1}{w_i} |S_i - \rho_i|^p \right)^{\frac{1}{p}} \quad (2.5)$$

donde N es el número de puntos de acceso, w_i es un factor de ponderación ($w_i \leq 1$), y p es el parámetro de norma, que empieza por 1. El factor de ponderación w_j es un parámetro de sesgo que puede potenciar o reducir el componente RSS en el *fingerprint* [16]. O bien un cierto número de

muestras de señales, o bien la desviación del *fingerprint* RSS pueden ser usados como factor de ponderación [16].

La distancia euclídea es una métrica de distancia muy conocida usada en [3, 21, 16] para clasificar las posiciones. Para la distancia euclídea, la ecuación 2.5 tiene $p = 2$ y los $w_i = 1$.

Las diferentes medidas de distancias (tales como la distancia Manhattan) son también posibles con $p \neq 2$ [16].

Una modificación al clasificador de vecindad más próxima utilizando información adicional de la desviación estándar del *fingerprint* cae fuera lo que fue estudiado en [16]. Esta modificación permite una clase especial llamada *patrón no clasificable*, el cual no es asociado con ninguna posición en la base de datos. Esto ocurre cuando una muestra *fingerprint* cae fuera de la región de dos desviaciones estándar en cada lado de una media RSS. Consideremos que el vector de desviación estándar para cada localización *fingerprint* j es denominada como $D_i = (\sigma_1^i, \sigma_2^i, \dots, \sigma_N^i)^T$.

Se pueden expresar matemáticamente criterios adicionales para la clasificación de *fingerprint* con vector de muestra S y *fingerprint* F_i tal como se dice en [16]:

$$\begin{aligned}
 \rho_1^i - 2 \sigma_1^i &\leq s_1 \leq \rho_1^i + 2 \sigma_1^i, \\
 \rho_2^i - 2 \sigma_2^i &\leq s_2 \leq \rho_2^i + 2 \sigma_2^i, \\
 &\dots \qquad \qquad \qquad \dots \\
 \rho_N^i - 2 \sigma_N^i &\leq s_N \leq \rho_N^i + 2 \sigma_N^i \qquad (2.6)
 \end{aligned}$$

Ya que hay patrones clasificables, la distancia de error media (la verdadera distancia entre la posición correcta y la incorrecta) es más pequeña [16].

Un clasificador de distancia mínimo, que no ha sido aplicado hasta este momento a ningún algoritmo de posicionamiento, es la *distancia Mahalanobis* [20]. La *distancia Mahalanobis* tiene tres ventajas frente a la distancia euclídea: la contabilización automática para el escalado de los ejes coordenados, la corrección de la correlación entre diferentes características y la permisión tanto de los límites de decisión lineales como no lineales [20]. Estas ventajas son útiles cuando las propiedades estadísticas de la localización *fingerprint* son consideradas explícitamente [20] como muestra la comparación de la distancia euclídea con la distancia Mahalanobis.

De hecho, la distancia euclídea es un caso especial de la distancia Mahalanobis cuando todos los componentes de las señales RSS en la localización *fingerprint* están incorrelados y sus varianzas son las mismas en todas las direcciones [20].

$$L_m^2 = (S - F)^T C^{-1} (S - F) \quad (2.7)$$

Hasta aquí, la discusión de las métricas de distancia está limitada a la determinación de la localización *fingerprint* más próxima. En la práctica, la verdadera localización no está limitada a localizaciones dentro del mapa de radio y puede haber más de una localización *fingerprint* más próxima. Esto proporciona una razón para modificar al método de vecindad más próxima usando para ello k vecindades más cercanas en lugar de solamente una [3,21] o vecindades más cercanas ponderadas [4, 21] , en donde se utilizan los factores w_i de la ecuación 2.5. La posición final estimada es un promedio de las k coordenadas vecinas cercanas más próximas. La razón de utilizar este esquema

es que el promediado de las coordenadas probablemente da como resultado una estimación más próxima de la localización correcta. Bahl y Padmanabhan [3] informaron que para k pequeños hay una pequeña mejora respecto el enfoque de simple vecindad más próxima, mientras que para k grandes aumentan los resultados del error de localización. Phongsak et al [16] informaron que para $k > 8$ el resultado se hizo peor. Todavía hay lugar para mejoras posteriores del método de vecindad más próxima. Algoritmos de búsqueda multidimensional tales como un Árbol-R, Árbol-X y búsqueda por los k -ésimos vecinos más cercanos (*k nearest neighbours*), están entre las posibles mejoras al algoritmo de posicionamiento sugerido en [3].

Los métodos de vecindad más próxima que usa la medida de distancias como funciones discriminantes pueden ser clasificados como un subconjunto de enfoques estadísticos para el reconocimiento de patrones [37]. La separación de la localización *fingerprint* es llevada a cabo bien por fronteras de decisión lineales o no lineales, las cuales, acorde con [23] son determinadas por las distribuciones de probabilidad de los patrones pertenecientes a cada localización *fingerprint*. La ejecución típica de los métodos de vecindad más próxima está basada en el error de la clasificación de o en la bondad en la cual se pueden separar los diferentes localizadores *fingerprint*. Los métodos de vecindad más próxima son rápidos en el despliegue y no requieren casi entrenamiento o sintonización de los algoritmos de posicionamiento. Aunque el cálculo es sencillo y se puede encontrar fácilmente un patrón que sirva, la complejidad aumenta cuando el número de elementos en el patrón (número N de señales de los puntos de acceso) aumenta en el mapa de radio. La escalabilidad de estos métodos no ha sido estudiada en el caso de un edificio grande.

2.5.3.2 Métodos de Redes Neuronales.

La aplicación de Redes Neuronales en sistemas de posicionamiento en interiores presuponen que las huellas RSS son demasiado complejos para ser analizados matemáticamente y pueden requerir funciones discriminantes no lineales bastante sutiles para la clasificación. De ahí que en lugar de encontrar funciones discriminantes adecuadas como la métrica de la distancia mínima, mediante este enfoque se utilice generalizadamente una estructura llamada *neurona*, cuya unidad de proceso vemos en forma de caja negra en [16].

La neurona consiste en un conjunto de enlaces de entrada que son ponderados con pesos sinápticos, un sumador que totaliza las entradas ponderadas y una función de activación que limita la amplitud de la salida de la neurona [16]. La función de activación se presenta generalmente en la forma de una función no lineal tal como una función sigmoideal. La función sigmoideal $f(x) = 1 / (1 + e^{-x})$, se dice que es adecuada para los problemas de clasificación “si/no” [4]. La figura 2.4 ilustra un modelo de neurona.

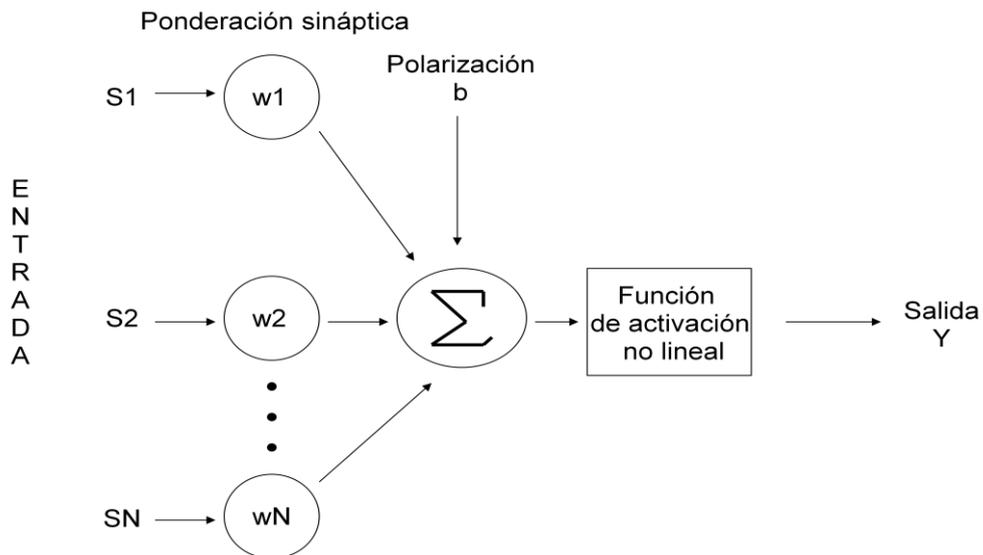


Figura 2.4: Ejemplo de neurona con función de activación no lineal

La interconexión de múltiples neuronas tanto en serie como en paralelo da como resultado una red neuronal de perceptrones multicapa (MLP) [4] en la cual las señales viajan secuencialmente desde la capa de entrada a la capa de salida y desde la salida de una neurona a la entrada de otra neurona. Las capas intermedias entre la salida y la entrada reciben generalmente el nombre de capas ocultas. Una red de neuronas conectadas completamente y con alimentación hacia delante, consistente en una capa oculta, fue utilizada para determinar localizaciones en [16]. Las entradas consisten en tres características, cada una de ellas representando la RSS de tres puntos de acceso. La capa oculta está compuesta de neuronas y la capa de salida tiene 19 nodos correspondientes a las 19 posiciones de su mapa.

Battiti et al [24] también implementaron una red neuronal MLP con una capa oculta que utiliza función sigmoideal y con la capa de salida que utiliza la función de identidad ($f(x) = x$). La arquitectura resultante en [24] fue 3/8/2, que quiere decir 3 entradas, 8 unidades ocultas y 2 salidas de coordenadas 2-D para 194 puntos de medida.

Cada red neuronal MLP puede ser entrenada con muestras de localizaciones *fingerprint* etiquetadas, a fin de calcular interactivamente todas las ponderaciones sinápticas en el interior de las neuronas. El proceso de entrenamiento es interactivo para cada muestra de entrada de una localización *fingerprint* en la cual las ponderaciones sinápticas sean sintonizadas de manera que las salidas estén en la posición correcta. Una técnica de entrenamiento llamada algoritmo de la secante de un paso (OSS) fue utilizada en [24] para ajustar iterativamente todas las ponderaciones sinápticas con la segunda información deducida. Por otra parte, el algoritmo de retropropagación de errores es otro algoritmo iterativo que fue usado en [16] para hallar las ponderaciones sinápticas. En este sentido, estos procesos de entrenamiento crean automáticamente límites complejos para las clases de localizaciones *fingerprint*.

Como modelo flexible, la red neuronal MLP no requiere conocimiento a priori de los parámetros de entorno tales como la localización de los puntos de acceso y las características de construcción [24]. Las prestaciones de nitidez y de precisión de [16, 24] fueron comentadas como mejores que las del método de vecindad más próxima, pero no muy significativamente. Por ejemplo, los resultados de [24] resaltaron que la distancia de error media (prestación de precisión) para la red neuronal MLP fue 1,82 m, mientras que la distancia de error media para la vecindad k más próxima fue 1,81. La desventaja de las redes neuronales recae en la lentitud de su tiempo de entrenamiento, que puede requerir un gran conjunto de entrenamiento para alcanzar una estimación de localización precisa. El problema del sobreentrenamiento, o sobreadaptación, también ocurre cuando el número de iteraciones de entrenamiento es mayor que 3.000, lo cual da como resultado un rendimiento pobre en la estimación de la localización [24]. En general, la prestación de error de una red neuronal no puede ser calculada analíticamente, debido a su complejidad. Sólo se pueden lograr los resultados de un test de medida para datos que no estén en el conjunto de entrenamiento.

Aunque los métodos de redes neuronales parezcan ser diferentes de los enfoques estadísticos o de los métodos de vecindad más próxima, estos métodos basados en redes neuronales tienen su contrapartida equivalente en los enfoques de reconocimiento de patrones estadísticos [22]. Dicho de otro modo, las redes neuronales también crean extremos de decisión como los métodos de vecindad más próxima. Por ejemplo, un perceptron es similar a una función discriminante lineal y los métodos de perceptrones multicapa son equivalentes o bien a los análisis discriminantes no lineales o bien a los estimadores de probabilidades a posteriori [22]. Dado que las redes neuronales no permiten la caracterización de las estadísticas de las localizaciones *fingerprint*, raramente proporcionan información interna del mecanismo subyacente de los sistemas de posicionamiento en interiores.

2.5.3.3 Métodos probabilísticos.

En lugar de representar la localización *fingerprint* con prototipos de los vectores del RSS medio, como en el enfoque de vecindad más próxima, el enfoque probabilístico modeliza la localización *fingerprint* con probabilidades condicionadas y utiliza el concepto de inferencia bayesiana para estimar la localización [5, 17, 18]. Este enfoque presupone un conocimiento a priori de la distribución de probabilidad de la localización del usuario, la cual es mostrada analíticamente para proporcionar una mejor precisión de localización que en un enfoque determinístico tal como el de los métodos de vecindad más próxima [41].

La distribución de localización a priori permite al sistema de posicionamiento mantener información de perfiles adicionales de localización de usuario [5] y puede mejorar la aplicación de seguimiento de localización [17]. Sin embargo, Battiti et al [4] puntualizan que a fin de crear la probabilidad condicional este enfoque necesita conocimiento adicional del modelo de propagación de la señal. El conocimiento puede estar en la forma de distribuciones empíricas de RSS en cada localización basada en conjuntos de entrenamiento medidos o en la forma de un modelo de propagación de radio con parámetros de radio estimados [4]. Sin utilizar conjunto de entrenamiento alguno.

Para cada coordenada de localización L , podemos estimar la función de densidad de probabilidad condicional de la función de verosimilitud $P(F | L)$ a partir de un conjunto de entrenamiento consistente en muestras de localización *fingerprint* y sus etiquetas.

Roos et al [5] sugiere dos métodos para estimar la función de verosimilitud: el método de núcleo y el método del histograma. Dadas n

muestras de RSS desde un punto de acceso en una localización, lo cual es un ejemplo dimensional, el método de núcleo impone una masa de probabilidad tal como una distribución gaussiana en cada muestra de valores de RSS.

Como función núcleo, cada distribución gaussiana tiene un valor medio ρ que es igual a una de las n muestras RSS y una desviación típica σ la cual tiene una anchura de núcleo arbitrariamente ajustable. De ahí, la función de verosimilitud resultante como muestra de RSS en una localización L dada es una suma equitativamente ponderada de todas las n funciones de núcleo gaussiano.

$$P(s|L) = \frac{1}{n} \sum_{i=1}^n \left[\frac{1}{\sqrt{2\pi} \sigma} \exp\left(-\frac{(s-\rho)^2}{2 \rho \sigma^2}\right) \right] \quad (2.8)$$

De la ecuación 2.8, la anchura de núcleo tendrá un efecto suavizador en la estimación de la densidad de probabilidad, si el valor es grande. Para ampliar el método a N dimensiones, Roos et al [5] hacen una suposición libre y multiplican conjuntamente todas las probabilidades condicionales de la siguiente forma $P(F|L) = P(s_1|L) P(s_2|L) \cdots P(s_N|L)$. También puntualizaron que el caso límite de un núcleo gaussiano con anchura de núcleo aproximadamente igual a 0 es equivalente al método de vecindad más próxima [5].

Por otra parte, el método del histograma estima las funciones de densidad continuas por medio de funciones de densidad discretizadas. El histograma es esencialmente un conjunto fijo de barras que representa la frecuencia de ocurrencia de muestras RSS que caen dentro del rango de cada barra. El rango de la barra se calcula a partir de un número ajustable de barras y los valores conocidos del mínimo y el máximo de RSS. El informe [5] se refiere a anchuras iguales a 3, 7 y 27.

Cuanto más grande es el número de barras, mejor puede el histograma aproximar la función de densidad de probabilidad de RSS.

Una forma ligeramente más sofisticada de determinar $P(F|L)$ se presenta en [17], donde los autores calculan dos probabilidades condicionadas diferentes a partir de dos histogramas diferentes y los multiplican conjuntamente. La primera probabilidad condicional representa la frecuencia de las observaciones de un punto de acceso, dada una localización L . En otras palabras, esta probabilidad indica con qué frecuencia el sistema puede encontrar la señal desde ese punto de acceso. La segunda probabilidad condicional representa la distribución de RSS desde ese punto de acceso, dada la misma localización.

Se considera que cada localización tiene una probabilidad a priori $P(L)$ que inicialmente puede ser equiparada para cada localización en el conjunto \mathcal{L} . De esa forma, los algoritmos de estimación de localización basados en el enfoque probabilístico aplican el teorema de Bayes a fin de encontrar una distribución a posteriori de la dicha localización, que es una probabilidad condicional de la localización L dada la localización fingerprint F de la forma siguiente:

$$P(L|F) = \frac{P(F|L)P(L)}{P(F)} = \frac{P(F|L)P(L)}{\sum_{k \in \mathcal{L}} P(F|L_k)P(L_k)} \quad (2.9)$$

De acuerdo con la ecuación 2.9 el enfoque probabilístico clasifica las localizaciones *fingerprint* de acuerdo con la máxima probabilidad a posteriori estimada, o sea que, selecciona la localización *fingerprint* de acuerdo con las funciones de verosimilitud. Por ejemplo, en el caso de dos localizaciones *fingerprint*, el la regla de decisión de Bayes selecciona la

posición A (LA) sobre la posición B (LB) para proporcionar la menor probabilidad de error [23] cuando:

$$P(L_A|F) > P(L_B|F) \quad (2.10)$$

$$P(F|L_A) P(L_A) > P(F|L_B) P(L_B) \quad (2.11)$$

Ya que los métodos probabilísticos contienen información adicional sobre la distribución de la localización, pueden proporcionar mejores prestaciones en la estimación de la localización. La desventaja de este método es que puede requerir un gran conjunto de entrenamiento para estimar más precisamente la distribución de probabilidad condicional. Una vez más, estos métodos probabilísticos requieren conocimiento explícito de las distribuciones de probabilidad de las *localizaciones fingerprint*.

2.5.3.4 Métodos de Máquina Vectorial de Soporte.

Recientemente, la técnica de máquinas de vector de soporte (SVMs) ha sido introducida como clasificador no paramétrico supervisado para enfocar el problema de posicionamiento en interiores, en [4]. El algoritmo SVMs tiene sus raíces en la teoría de aprendizaje estadístico introducida por Vapnik [27] en la cual se combinan a la vez técnicas estadísticas, aprendizaje de máquinas y redes neuronales. Para estimar la dependencia entre la RSS *fingerprint* y la localización desde las observaciones, este enfoque no requiere propiedades detalladas de la dependencia, tales como ocurría en el modelo de propagación o el método probabilístico. La potencia del algoritmo SVMs reside en su habilidad para generalizar clasificaciones que minimizan el error de test o error de clasificación para los datos después del período de entrenamiento. En otras palabras, la máquina de aprendizaje puede ser entrenada correctamente por el aprendizaje de un pequeño conjunto de entrenamiento y creando una estructura

suficiente para la clasificación de los datos sin memorizar la sobreadaptación de las muestras de entrenamiento [4].

El concepto fundamental del algoritmo SVM está basado en el principio de Minimización de Riesgo Estructurado (SRM) que intenta minimizar el extremo de un error funcional o generalizado esperado [4]. El riesgo funcional está definido como un valor esperado de la función de pérdida. La función de pérdida es una medida de cómo difiere la función usada para aproximar la correspondencia del patrón respecto de la correspondencia real. La función de riesgo general muestra que está limitada por la función de riesgo empírica y por el intervalo de confianza de Vapnik-Chervonenkis (VC)

El problema de la clasificación de RSS *fingerprint* puede ser considerado como un caso de clasificación no lineal. Aquí explicamos la operación de clasificación de SVM sin entrar mucho en detalles matemáticos. Primeramente, se hace corresponder los vectores de localización *fingerprint* con un espacio de dimensión superior llamado *espacio de característica* [4] utilizando para ello una función llamada *núcleo* de SVM, para llevar a cabo la transformación del vector. Hay tal variedad de funciones núcleo de SVM que se puede elegir entre funciones polinómicas, funciones de base radial (RBF), núcleos sigmoidales y núcleos de análisis de varianza (ANOVA) [4]. Hay que destacar que Battiti et al eligen una función de base radial (RBF) como núcleo de SVM en [4]. Finalmente, el algoritmo SVM crea un hiperplano de separación óptimo o superficie de decisión en dicho espacio caracterizado y utiliza el hiperplano para llevar a cabo la clasificación. En general, el hiperplano de separación no es único y es óptimo cuando posee la distancia más larga posible desde el punto de entrenamiento más próximo o desde un margen máximo. Los vectores de soporte son los vectores de entrenamiento que son necesarios para definir los hiperplanos [4]; de ahí que la máquina de vector de soporte es el algoritmo de aprendizaje (máquina) basado en vectores de soporte.

Aunque sea novedoso y además sea la técnica más sofisticada usada con éxito en el campo del reconocimiento de patrones, la ejecución de esta técnica para posicionamiento en interiores no difiere mucho de las otras técnicas. Para el problema de la regresión, tiene prestaciones comparables al método de k vecindades más próximas ponderadas. La distancia de error resultante para las k vecindades más próximas ponderadas fue 3,93 metros al 75%, mientras que la distancia de error resultante para SVM fue 3,96 metros al 75%, en [4]. El algoritmo SVM es más adecuado para resolver problemas de clasificación, p.e. para determinar si el área es un espacio interior o exterior [4]. Las prestaciones de los sistemas de posicionamiento dependen del error de aceptación mas que del error de clasificación o error cuadrático medio en los tres métodos previos [22]. Para mejorar las prestaciones de clasificación, se debe seleccionar apropiadamente un núcleo propio de SVM y sus parámetros, ya que hay varias funciones de núcleo entre las que elegir. Desde la perspectiva del modelo teórico de este estudio, el SVM puede ser demasiado complejo para proporcionar información útil en el diseño de un sistema de posicionamiento.

2.5.4. RESUMEN DE LAS PRESTACIONES DE LOS POSICIONAMIENTOS DE INTERIORES EXISTENTES.

Este subcapítulo resume brevemente las prestaciones de los sistemas de posicionamiento en interiores basados en localización *fingerprint* y WLAN. Las principales métricas de prestaciones estudiadas para todos los sistemas se refieren a la precisión del posicionamiento, que es una forma de medida del error. Como se mencionó en el capítulo I y figura en la tabla II, la precisión de la información sobre localización es generalmente reportada como la distancia de error entre la localización estimada y la verdadera localización móvil. Sin embargo, el informe sobre la precisión debería incluir el intervalo de confianza o porcentaje de detección de localización exitosa, que recibe el nombre de nitidez de localización. Las investigaciones recientes en los algoritmos de

posicionamiento han atraído a un gran número de grupos de investigación, con el fin de mejorar esas dos herramientas de reconocimientos de patrones tan distintas, tal como se ha discutido anteriormente. Los parámetros del sistema de los sistemas existentes se resumen en la tabla III y las mejores prestaciones reportadas de estos sistemas figuran reunidas en la tabla IV.

Sistema	Espaciado	Posiciones	Muestras/ Posición	AP	Orientaciones	Entorno
RADAR [3]	No uniforme	70	80 (μsg / muestra)	3	4	Medio camino
Saha et al [16]	Min 3,12 m	19	1200	3	N / A	1 piso
Roos et al [5]	Uniforme 2 m	155	40	10	N / A	1 piso
Battiti et al [4]	N / A	257	N / A	6	N / A	1 piso
Ladd et al [14]	3 m	11	1307 paquetes	5	2	Medio camino
Prasithangaree et al [21]	1,5 m , 3 m	60	40	2-7	4	1 piso
Youssef et al [18]	1,5 m	110	300	4	N / A	Medio camino
Xiang et al [17]	N / A	100	300 (2seg/muestra)	5	4	1 piso

Tabla III: Comparación de parámetros en sistemas de posicionamiento en interiores

Sistema	Tipo de algoritmo	Precisión y nitidez
RADAR [3]	Vecindad mas próxima	Entre 7 pies; 38%
Saha et al [16]	Vec m próx y Red Neuronal	Precisión no especificada; 90%
Roos et al [5]	Bayesiano	La mejor entre 8,27 pies; 90%
Battiti et al [4]	SVM, Bayesiano, Red Neuronal, Vecindad más próxima con ponderación K	Todo entre 16-17 pies; 90%
Ladd et al [14]	Bayesiano	Entre 5 pies; 77%
Prasithangaree et al [21]	Vecindad más próxima con ponderación K	25 pies a 75%; 40 pies a 95%
Youssef et al [18]	Bayesiano	Entre 7 pies; mas de 90%
Xiang et al [17]	Bayesiano con modelo de distribución RSS	Entre 6 pies; 90% (dispositivo estático)

Tabla IV: Comparación de prestaciones de sistemas de posicionamiento en interiores

Aunque las prestaciones de precisión y de nitidez varíen de unos sistemas a otros, tal como se observa en la tabla IV, las diferencias entre ellos no son muy significativas. Es más fácil comparar las prestaciones si se fija una de esas métricas, p.e., a igualdad de nitidez, digamos del 90%, el sistema que consigue la precisión más pequeña es el mejor. Fijémonos en que estas dos métricas de prestaciones son cercanamente dependientes. Cuando queremos incrementar la precisión (acortar la distancia de desviación) la prestación de nitidez debe ser disminuida (decremento del porcentaje de fiabilidad)[26]. Por otro lado, al disminuir la precisión (alargar la distancia de desviación) se puede incrementar la prestación de nitidez (incrementar el porcentaje de fiabilidad). Es de interés fijarse en que los resultados [4] reportaron una prestación de precisión media comprendida dentro de los 3 metros, con los

cuatro algoritmos de posicionamiento. Es importante subrayar que cada sistema tiene diferentes asignaciones de parámetros, de ahí que las comparaciones resumidas no dan buenos resultados. Intuitivamente, esperamos que un sistema que tenga un número mayor de puntos de acceso para funcionar mejor deba tener la dimensión mayor de vectores de localizaciones *fingerprint*, lo cual dará como resultado una mayor separabilidad de patrones. Mas aún, un sistema que tiene posiciones solamente a mitad de camino puede funcionar mejor debido a que tiene menor número de posiciones en con las que el sistema se puede confundir. Sin embargo, esas dos hipótesis no fueron confirmadas en la comparación llevada a cabo anteriormente. Estas observaciones muestran una falta de metodología unificada de comparación de prestaciones y necesita explicación teórica de la relación entre las prestaciones de funcionamiento del posicionamiento en interiores y los parámetros del sistema.

Además de estas dos métricas de prestaciones, otras métricas tales como retardo, capacidad, cobertura y escalabilidad, que fueron estudiadas en apartado 2.3.3. no son consideradas en la mayoría de los estudios actuales. Sin embargo, la complejidad computacional de la fase off-line o fase de aprendizaje y la fase on-line o fase de estimación, es analizada por [4] en los tres enfoques de posicionamiento, a saber, vecindad más próxima k ponderada, enfoque probabilístico bayesiano y red neuronal de preceptron multicapa.

Coincidentemente, la determinación de localización WLAN por medio de racimos y distribución de probabilidad [18] es un estudio que intenta mejorar la precisión de posicionamiento y de reducir los requerimientos computacionales en la estimación de localización de los algoritmos de posicionamiento. Youssef et al [18] propusieron la técnica de Racimo Conjunto (JC) que usa el racimo de localizaciones compartiendo un conjunto común de puntos de acceso, a fin de reducir la complejidad computacional. Sin embargo, su esquema está todavía basado en el enfoque probabilístico con inferencia bayesiana discutido anteriormente. Sin tener que ver con la discusión previa

sobre el reconocimiento de patrones, las técnicas de *fusión de sensores* permiten al sistema de posicionamiento explotar redundancias y contradicciones para reducir la incertidumbre total sobre localización [6]. Esta técnica ayuda a mejorar la precisión y la nitidez de los sistemas de posicionamiento. Se presenta un ejemplo de ello en [17], en donde el sistema combina múltiples observaciones independientes para ganar mejor estimación a partir del algoritmo de posicionamiento. El modelo markoviano oculto (HMM, que es usado en [17] para hacer el seguimiento del usuario, aumenta las prestaciones de precisión y de nitidez. Usando sólo una de las técnicas de sensor de localización se obtienen prestaciones limitadas del sistema y gran incertidumbre en la información de la localización. El enfoque híbrido, que consiste en dos o más técnicas y tecnologías de sensores, puede mejorar también las prestaciones de precisión y de nitidez de los sistemas de posicionamiento en interiores. En [2] se discute un ejemplo de sistema híbrido, en el cual se combinan AOA y TDOA.

2.6 REFERENCIAS

- [1] K. Pahlavan, X. Li, and J. P. Makela, "Indoor geolocation science and technology," *IEEE Commun. Mag.*, vol. 40, no. 2, pp. 112-118, Feb. 2002.
- [2] T. S. Rappaport, J. H. Reed, and B. D. Woerner, "Position location using wireless communications on highways of the future," *IEEE Commun. Mag.*, vol. 34, no. 10, pp. 33-41, Oct. 1996.
- [3] P. Bahl and V. N. Padmanabhan, "RADAR: an in-building RF-based user location and tracking system," in *Proc. IEEE Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM'00)*, Tel Aviv, Israel, Mar. 2000, pp. 775-784.
- [4] R. Battiti, M. Brunato, and A. Villani, "Statistical learning theory for location fingerprinting in wireless lans," *Technical Report*, Oct. 2002. [Online]. Available: <http://rtm.science.unitn.it/~battiti/archive/86.pdf>
- [5] T. Roos, P. Myllymaki, H. Tirri, P. Misikangas, and J. Sievanen, "A probabilistic approach to wlan user location estimation," *International Journal of Wireless Information Networks*, vol. 9, no. 3, pp. 155-164, July 2002.
- [6] J. Hightower, B. Brumitt, and G. Borriello, "The location stack: A layered model for location in ubiquitous computing," in *Proc. IEEE Workshop on Mobile Computing Systems & Applications (WMCSA'02)*, Callicoon, NY, June 2002, pp. 22-28.
- [7] J. Hightower and G. Borriello, "Location systems for ubiquitous computing," *IEEE Computer*, vol. 34, no. 8, pp. 57-66, Aug. 2001.
- [8] J. A. Tauber, "Location systems for pervasive computing," *Area Exam Report*, Massachusetts Institute of Technology, Aug. 2002.
- [9] R. Want, A. Hopper, V. Falcao, and J. Gibbons, "The active badge location system," *ACM Transactions on Information Systems*, vol. 40, no. 1, pp. 91-102, Jan. 1992.
- [10] M. Ocaña, "Sistema de localización global wifi aplicado a la navegación de un robot semiautónomo," *Master's thesis*, Universidad de Alcalá, 2005.
- [11] A. Harter, A. Hopper, P. Steggles, A. Ward, and P. Webster, "The anatomy of a context aware application," in *Proc. ACM International*

Conference on Mobile Computing and Networking (MOBICOM'99), Seattle, WA, Aug. 1999.

[12] N. B. Priyantha, A. Chakraborty, and H. Balakrishnan, "The cricket location-support system," in Proc. ACM International Conference on Mobile Computing and Networking (MOBICOM'00), Boston, MA, Aug. 2000, pp. 32-43.

[13] V. Zeimpekis, G. M. Giaglis, and G. Lekakos, "A taxonomy of indoor and outdoor positioning techniques for mobile location services," Journal of ACM Special Interest Group on Electronic Commerce, vol. 3, pp. 19-27, 2002.

[14] J. Hightower, R. Want, and G. Borriello, "SpotON: An indoor 3D location sensing technology based on RF signal strength," University of Washington, Seattle, WA, Technical Report UW CSE 2000-02-02, Feb. 2000. [Online]. Available: <http://seattle.intelresearch.net/people/jhightower/pubs/hightower2000indoor/hightower2000indoor.pdf>

[15] J. Werb and C. Lanzl, "Designing a positioning system for finding things and people indoors," IEEE Spectr., vol. 35, no. 9, pp. 71-78, Sept. 1998.

[16] S. Saha, K. Chaudhuri, D. Sanghi, and P. Bhagwat, "Location determination of a mobile device using ieee 802.11b access point signals," in Proc. IEEE Wireless Communications and Networking Conference (WCNC'03), New Orleans, LA, Mar. 2003, pp. 1987-1992.

[17] A. M. Ladd, K. E. Bekris, G. Marceau, A. Rudys, L. E. Kavraki, and D. S. Wallach, "Robotics-based location sensing using wireless ethernet," in Proc. ACM International Conference on Mobile Computing and Networking (MOBICOM'02), 2002, pp. 227-238.

[18] M. A. Youssef, A. Agrawala, and A. U. Shankar, "WLAN location determination via clustering and probability distributions," in Proc. IEEE International Conference on Pervasive Computing and Communications (PerCom'03), Dallas-Fort Worth, TX, Mar. 23-26, 2003, pp. 23-26.

[19] J. Small, A. Smailagic, and D. P. Siewiorek, "Determining user location for context aware computing through the use of a wireless lan infrastructure," Online, Dec. 2000. [Online]. Available: <http://www-2.cs.cmu.edu/~aura/docdir/small00.pdf>

[20] J. T. Tou and R. C. Gonzalez, Pattern Recognition Principles, 2nd ed. Reading, MA: Addison-Wesley, 1974.

- [21] P. Prasithsangaree, P. Krishnamurthy, and P. K. Chrysanthis, "On indoor position location with wireless LANs," in Proc. IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC'02), Lisbon, Portugal, Sept. 2002.
- [22] A. K. Jain, R. P. W. Duin, and J. Mao, "Statistical pattern recognition: A review," IEEE Trans. Pattern Anal. Machine Intell., vol. 22, no. 1, pp. 4-37, Jan. 2000
- [23] R. O. Duda, P. E. Hart, and D. G. Stork, Pattern Classification, 2nd ed. New York, NY: Wiley-Interscience, 2000.
- [24] R. Battiti, "Location-aware computing: a neural network model for determining location in wireless lans," Universita degli Studi di Trento, Tech. Rep. DIT-0083, Feb. 2002. [Online]. Available: <http://rtm.science.unitn.it/~battiti/archive/83.pdf>
- [25] M. Pardo and G. Sberveglieri, "Learning from data: A tutorial with emphasis on modern pattern recognition methods," IEEE Sensors J., vol. 2, no. 3, pp. 203-217, June 2002.
- [26] M. Siebert, D. Bultmann, and M. Vorobjova, "On the reliability of location based networking," in Proc. 1st Workshop on Positioning, Navigation and Communication (WPNC'04). Hannover, Germany: Shaker Verlag GmbH, Mar. 2004, pp. 107-112.

3. Características de la señal de propagación en interiores. Aplicación al método del *fingerprinting*.

3.1. INTRODUCCIÓN.

Como se vio en el capítulo anterior, los algoritmos de localización basados en la técnica del *fingerprinting* están basados en el nivel de la señal o RSS recibida en una serie de puntos llamados huellas[1]. Por lo tanto el estudio de las características de propagación en entornos de interiores es una parte importante a la hora de diseñar un sistema de localización en este tipo de escenarios.

En este capítulo se va a estudiar las características de propagación de las señales de radiofrecuencia en interiores de edificios. En primer lugar se hará un repaso de los modelos de predicción existentes para esta clase de escenarios, pasando a continuación a ver las limitaciones de estos modelos. Posteriormente se mostrarán una serie de medidas realizadas en varios escenarios en los cuales se varían las condiciones ambientales.

Los escenarios seleccionados han pretendido recoger un entorno de oficinas variado, con posibilidades de toma de mediciones en diferentes horarios, con y sin personas. Se han tomado mediciones en zonas de un edificio universitario como son un pasillo con varios despachos en ambos laterales y un

laboratorio al fondo; un pasillo y aulas y, en dos zonas de consultas hospitalarias.

3.2. MODELOS TEÓRICOS DE PROPAGACIÓN.

Un modelo de propagación es un conjunto de expresiones matemáticas, diagramas y algoritmos utilizados para representar las características de radio de un determinado entorno [2].

La señal emitida por un transmisor va a experimentar múltiples transformaciones en su señal y camino a lo largo de su trayectoria por el medio de propagación, llegando al receptor solo una pequeña parte. El camino entre receptor y emisor pueden variar en múltiples formas debido a la existencia de diferentes obstáculos. Esto hace muy difícil predecir el nivel de la señal recibida en un determinado punto o analizar el canal de radio.

Los modelos de propagación se han enfocado tradicionalmente en predecir la potencia promedio de la señal recibida; así como la variación de la potencia en la proximidad espacial de un lugar en particular.

Los modelos de propagación de radio se pueden clasificar en dos tipos principales: modelos de propagación en exteriores y modelos de propagación en interiores. Además, ateniéndose al tamaño de la zona de cobertura, los modelos en exteriores se pueden dividir en modelos de propagación en zonas grandes (macrocelas) y de zonas pequeñas (microcelas).

Existen más modelos de propagación para sistema inalámbricos en exteriores, debido a que la comunicación inalámbrica en exteriores se viene utilizando desde hace mucho más tiempo: radio, televisión,...etc. Los modelos

para exteriores no tienen en cuenta las características particulares de la propagación en interiores.

El campo de la propagación en interiores es relativamente nuevo y las primeras investigaciones datan de principios de los 80s. La llegada de las WLAN hace todavía más necesario el disponer de modelos para interiores de predicción de cobertura.

La mayoría de los modelos de cobertura de redes inalámbricas existentes son modelos de caja negra, en los que se ajustan los parámetros del modelo y se aplican las fórmulas directamente para obtener los resultados de cobertura.

Los resultados de la predicción de cobertura de un punto de acceso proporcionada por un modelo de propagación en interiores son muy importantes para facilitar las siguientes tareas:

- Predecir el tamaño de las áreas que se pueden cubrir con un único punto de acceso, AP.
- Planificar la ubicación de las celdas de modo que, aún utilizando la misma frecuencia, no se interfieran ni causen errores.

En entornos cerrados los niveles de señal fluctúan en mayor medida que en entornos abiertos. Esta diferencia se explica en el hecho de que en una localización específica, el campo eléctrico se forma por un número mucho mayor de componentes indirectos que en el caso de un entorno abierto. Los modelos de propagación en interiores difieren de los modelos de propagación tradicionales en dos aspectos:

- Las distancias cubiertas son mucho más pequeñas.

- El componente variable del entorno es mucho mayor para separaciones más pequeñas entre transmisor y receptor.

Un modelo científico se puede definir como una explicación de determinados fenómenos naturales representada de una forma determinada. Por ejemplo, Galileo desarrolló un modelo para explicar los patrones de movimiento de los objetos celestes, incluidas las salidas y puestas del sol y de la luna. Los actuales científicos desarrollan modelos para explicar el fenómeno de división incontrolada de células o ausencia de “apoptosis” (muerte programada de células) asociada con el cáncer. Y como veremos más adelante, se desarrollan modelos para explicar el fenómeno de propagación de ondas de radiofrecuencia en redes inalámbricas.

Los científicos usan dibujos, gráficos, ecuaciones, estructuras tridimensionales, o explicaciones en texto libre para comunicar sus modelos (que son ideas, no objetos físicos) a otros científicos.

3.2.1. TIPOS DE MODELOS

3.2.1.1. Modelos matemáticos

Son procedimientos matemáticos que describen el comportamiento del fenómeno modelado. Dependiendo de la complejidad de las fórmulas matemáticas aplicadas y del número de parámetros de las que dependan, pueden llegar a necesitar gran cantidad de potencia de computación para generar los resultados.

3.2.1.2. Modelos Empíricos o estadísticos

Se basan en la extrapolación estadística de resultados a partir de medidas del fenómeno realizadas sobre el terreno.

La principal ventaja de este tipo de modelos es que se tienen en cuenta de forma implícita las influencias propias del entorno en su conjunto, sin ser reconocidas cada una de ellas de manera aislada.

Por el contrario, su precisión depende no sólo de la precisión de las medidas, si no de la similitud entre el entorno donde fueron llevadas a cabo las medidas y el entorno a analizar [2].

3.2.1.3. Modelos Teóricos

Se basan en los principios fundamentales del medio que se intenta modelar (por ejemplo los principios de la física, si se modelan fenómenos físicos). Pueden ser aplicados en diferentes entornos sin que se vea afectada su precisión.

En la práctica la implementación de modelos teóricos requiere enormes bases de datos de características relativas al entorno, las cuales en la práctica son imposibles de obtener en muchos casos, o muy difíciles de obtener.

Los algoritmos usados por los modelos teóricos son generalmente muy complejos y computacionalmente poco eficientes. Por esta razón su implementación se restringe a pequeñas áreas. Por el contrario, si su implementación es correcta, proporcionan gran precisión en su predicción, en comparación con los modelos empíricos.

3.2.1.4. Modelos deterministas

En este tipo de modelos los resultados de la simulación no tienen ninguna distribución de probabilidad, y son valores que sólo dependen de las condiciones de simulación (entradas). Siempre con la misma entrada van a producir la misma salida.

3.2.1.5. Modelos estocásticos

Son modelos en los que los resultados del modelo tienen una distribución de probabilidad. La misma entrada no tiene porqué producir los mismos resultados, en distintas simulaciones del modelo.

3.2.1.6. Modelos de caja negra

Son modelos en los que sólo se reproduce la entrada y salida del fenómeno, sin preocuparse de lo que pase en el interior.

3.2.2. VALIDEZ DE UN MODELO

Queda claro que existen muchos tipos de modelos, lo que puede dar lugar a que para un único fenómeno se puedan plantear diferentes modelos, y para definir si un modelo es correcto se pueden usar varios factores:

1. El modelo puede explicar las observaciones realizadas del fenómeno.
2. Se puede usar el modelo para predecir el comportamiento de fenómeno modelado.

3. El modelo es consistente con otras ideas acerca del funcionamiento del fenómeno modelado. En primer lugar, de un modelo no se dice si es correcto, si no que se dice que es “aceptable”. Y la aceptabilidad del modelo se basa en la capacidad del modelo de cumplir los tres factores expuestos anteriormente: explicar, predecir y ser consistente con otros conocimientos.

En segundo lugar, puede haber más de un modelo que sea una explicación “aceptable” para el mismo fenómeno. No siempre es posible (ni deseable) elegir un único modelo como aceptable.

Los modelos tampoco son inamovibles. Es frecuente que con el aumento de observaciones del fenómeno se cambien los modelos, aparezcan otros más “aceptables”, que son capaces de explicar más casos de comportamiento; que pueden predecir mejor el comportamiento o que son más consistentes con otros tipos de conocimiento.

En resumen:

- Los modelos son representaciones simplificadas de la realidad por medio de un conjunto de restricciones e hipótesis.
- Los modelos son ideas que usan los científicos para explicar patrones de comportamiento que se observan en el mundo real.
- Los modelos se consideran aceptables o no aceptables en base a lo bien que pueden explicar y predecir comportamientos; y si son consistentes con otros conocimientos contrastados.
- Los modelos se usan constantemente para responder cuestiones acerca de determinados fenómenos, y cuando se obtienen nuevos datos son susceptibles de ser revisados o descartados.

3.2.3. CÁLCULO TEÓRICO DE LA PROPAGACIÓN

En los sistemas de localización en interiores basados en señales de radio es imprescindible conocer todos los pequeños detalles posibles de las distintas características de propagación de la señal para conseguir una estimación teórica lo más parecida posible a la estimación experimental. La señal experimenta variaciones de mayor o menor importancia debidas a fenómenos como la atenuación en el vacío, atenuación por absorción de la señal por parte de un obstáculo grande, reflexiones, refracciones o difracciones. También hemos de tener en cuenta la presencia y el movimiento de personas. En el caso de la localización en interiores, todos estos fenómenos hacen que la señal tenga que ser estudiada muy cuidadosamente si se quiere establecer un sistema de posicionamiento basado en este tipo de señales. En este capítulo se van a describir distintos tipos de cálculos de la propagación, y se detallarán las ventajas de unos respecto de otros

3.2.3.1. Campo eléctrico y señal de radio

Las ondas de radio se producen por las fluctuaciones periódicas de un campo eléctrico, y se pueden descomponer en componentes ortogonales. Usualmente habrá una componente horizontal y otra componente vertical, y el cociente entre las dos componentes definirá la polarización de la onda [3]. Por ejemplo, cuando la componente vertical es cero, se dirá que la señal está polarizada horizontalmente.

La potencia recibida procedente de un campo eléctrico en un punto se puede expresar como:

$$P_r = \frac{F G_r \lambda^2}{4\pi L} \quad (3.1)$$

Siendo F el flujo de potencia del campo eléctrico en ese punto, G_r la ganancia de la antena (que depende de su geometría y tamaño y de la frecuencia de la señal emitida), λ es la longitud de onda, y L es la pérdida producida en el receptor. Aunque el término λ aparece elevado al cuadrado en la ecuación (3.1), no es cierto que a mayor longitud de onda de la señal, la potencia recibida sea mayor, ya que la ganancia G_r será menor al crecer la longitud de onda y estos términos se cancelan el uno al otro. De esta ecuación se puede decir que en el vacío la potencia recibida es independiente de la frecuencia y sólo depende del flujo F . Pero en condiciones no ideales, mayor frecuencia significará que se producirán más pérdidas por absorciones, reflexiones, etc. al disminuir la longitud de onda, que hará que los obstáculos parezcan más grandes para la onda.

Por último, los valores de potencia recibida varían en un rango muy amplio, por lo que normalmente se expresan en escala logarítmica. En nuestro caso, la potencia de señal recibida de un punto de acceso se expresa en dBm (decibel milliwatts), que es una medida logarítmica del cociente de la potencia recibida tomando como referencia una potencia de 1mW.

Un inconveniente del estudio experimental de las potencias recibidas surge en cuanto se tienen diversos tipos de dispositivos inalámbricos: cada uno tiene antenas con ganancias distintas así como diferentes sensibilidades. Por ello, una red que para un dispositivo inalámbrico puede recibirse con una potencia de señal tenue, para otro con menor sensibilidad puede no llegar a recibirse en absoluto.

3.2.3.2. Atenuación en Espacio Libre

En una onda que se propaga idealmente, la potencia máxima recibida en una distancia d disminuye por igual en las tres dimensiones del espacio.

Ignorando la posible pérdida de potencia por absorción del medio donde se propaga la onda, la ecuación anterior nos queda del siguiente modo:

$$P_r = \frac{P_t G_t \lambda^2}{4\pi d} \quad (3.2)$$

Siendo P_t la potencia transmitida, G_t la ganancia de la antena y d la distancia entre emisor y receptor [3]. Al combinar las ecuaciones *que define los decibel milliwatts*, $\text{dBm} = 10 \log \frac{P}{\text{mW}}$ y la (3.2) tenemos una versión reducida de la potencia recibida, que normalmente se expresa en decibelios:

$$P_r [\text{dB}] = P_t [\text{dB}] + 10\log(G_r) + 20\log(\lambda) + 20\log(d) - 22.0 \quad (3.3)$$

Esta última ecuación es, al igual que la ecuación (3.2), válida solamente en un entorno libre de obstáculos, donde no se producen distorsiones en la señal provocadas por absorciones o cualquier otro fenómeno de los que se pueden presentar en la señal inalámbrica. Cuando el camino entre transmisor y receptor no está libre de obstáculos y por tanto no hay visión directa, lo que se conoce como LoS (línea de vista, *Line of Sight*), la potencia recibida en el receptor es significativamente inferior a la que habría en un entorno libre de obstáculos. Incluso si hay LoS entre transmisor y receptor la ecuación (3.3) no tiene por qué ser del todo cierta.

Además de LoS, entre transmisor y receptor se debe encontrar sin obstáculos una zona amplia de espacio denominada zona de Fresnel para que no se produzcan frentes de onda secundarios perjudiciales para la señal propagada.

La presencia de un obstáculo dentro de la zona de Fresnel hace que se generen frentes de onda secundarios que son perjudiciales al no estar en fase con el frente de onda principal, atenuándolo. En cambio, si la zona de Fresnel

está libre, no habrá frentes de onda secundarios fuera de fase y la propagación será la misma que la producida en el vacío. Por la naturaleza de los interiores de un edificio, no será posible establecer una zona de Fresnel libre de obstáculos, por ello, no hay mucho que se pueda hacer para evitar atenuaciones. Sin embargo, gracias a estas atenuaciones la señal no posee un alcance elevado y gracias a ello se pueden establecer sistemas de localización con señales Wi-Fi.

3.2.3.3. Absorciones y multitrayecto.

Para llegar de la antena transmisora a la antena receptora, la señal inalámbrica se propaga principalmente por el aire, pero en algunos momentos tendrá que propagarse por otros medios como el vidrio de las ventanas, por las paredes, muebles, etc. Al aparecer obstáculos entre emisor y receptor, como consecuencia del fenómeno de la *absorción*, este absorbe parte de la energía de la onda. Cada vez que la onda se topa con un obstáculo, se produce una pérdida de potencia de γ^d , donde γ es una constante que depende de las propiedades físicas del medio y de la frecuencia de la señal que se propaga por el medio, y d es la distancia recorrida en ese medio. Al tratarse de una función exponencial, esto significa que en nuestra escala geométrica la pérdida por absorción es lineal con respecto a la distancia.

La absorción en interiores es normalmente muy elevada, debido a que se encuentran paredes y toda clase de obstáculos grandes cada pocos metros. Según [4], la atenuación total por absorciones ante frecuencias menores a 10GHz, como las presentes en la señal inalámbrica, oscila entre 1 y 20 dB, dependiendo del número de obstáculos y su material.

La caracterización del canal de propagación en interiores es un fenómeno complejo [5] Los espacios interiores son proclives a las interferencias, sus características físicas provocan reflexiones, refracciones y

dispersión o *scattering* de la señal radio, dando lugar al multicamino. El multicamino es el causante del desvanecimiento (*fading*) de la señal, y el origen de las mencionadas interferencias. Por otra parte, debido al movimiento habitual de personas (y objetos) dentro de los espacios interiores, los canales de comunicación en interiores presentan el problema de su falta de estacionariedad temporal.

Algunos de estas características físicas son fáciles de implementar en modelos de propagación, pero otras son muy complicadas. Es habitual despreciar los efectos de algunas de ellas, porque la incidencia en los resultados finales, no merecen el gasto en computación y complejidad de los modelos de propagación.

En muchos modelos, se pueden implementar todo tipo de obstáculos estáticos, siempre que se definan sus límites, su ubicación, y las características de los materiales que los componen. Pero hay unos elementos que no se tienen en cuenta en los modelos de propagación: las personas.

El ser humano, como elemento compuesto mayoritariamente por agua, supone un obstáculo importante en la propagación de ondas en las frecuencias que se utilizan habitualmente; ya que el agua absorbe las microondas (igual que en los hornos microondas que trabajan en frecuencias similares). Realizar los cálculos de cobertura en áreas sin tener en cuenta las personas que van a estar o circular por ellas, supone que los resultados sean erróneos.

Existe varios estudios que intentan modelar el movimiento de las personas en estos escenarios[1,5,6]. En la presente Tesis no se pretende crear un modelo matemático que tenga en cuenta este fenómeno, pero se harán varias medidas en diferentes escenarios y con diferente afluencia de gente para poder estudiar como varía la señal y si esta se ajusta a algún tipo de función densidad de probabilidad.

Otro aspecto importante en la propagación en interiores es el multitrayecto que surge debido a las múltiples reflexiones y difracciones de la señal. Este fenómeno se puede modelar mediante la siguiente expresión:

$$r(t) = \sum_{l=1}^L \beta_l s(t - \tau_l) \quad (3.4)$$

Al haber multitrayecto, la señal en un punto oscila en el tiempo alrededor de su valor medio. La señal recibida se suele modelizar como una combinación de efectos a gran escala (*large-scale*) y efectos a pequeña escala (*small-scale*). Los efectos o desvanecimiento a gran escala son la atenuación debida a las paredes y mobiliario, y se suele modelizar como una distribución log-normal, sin embargo, al medirse la potencia en unidades logarítmicas la distribución es en realidad una normal. Los efectos o desvanecimiento a pequeña escala son más sutiles, y se refieren a las fluctuaciones debidas al multitrayecto, y se modelizan como una distribución Rician si entre transmisor y receptor hay LoS, y en cambio siguen una distribución de Rayleigh si no hay LoS entre transmisor y receptor.

3.2.3.4. Modelo básico de propagación en interiores.

Un modelo simple que se puede realizar para modelizar la pérdida de potencia de la señal inalámbrica es, según [7]:

$$P_L [dB] = P_L [dB] + 10 n \log \left(\frac{d}{d_0} \right) + X_\sigma \quad (3.5)$$

Siendo P_L la pérdida a una distancia d del transmisor, P_{L0} es la pérdida a una distancia d_0 , n es un valor que depende del entorno (del material del edificio, los muebles que hay en él, personas, etc.) y por último X_σ representa un valor aleatorio en dB, que sigue una distribución normal con desviación típica de σ dB. Aquí hay que notar que la potencia sigue una distribución normal porque la señal sigue una distribución log-normal cuando no se

presenta en una escala logarítmica (dBm) sino en miliwatios. La variable n suele tener un valor entre 1,8 y 2,6, y σ suele tener un valor de 1,5 a 7,2. Estos valores se discuten en [4].

Una cuestión importante, además de ver el valor óptimo de n para cada entorno, sería ver si σ ha de ser constante o es mejor hacerlo variable con la distancia. Esto es porque el multitrayecto es mucho más evidente cuando el receptor está cerca del transmisor.

Los desvanecimientos de pequeña escala no son importantes para nosotros ya que al evolucionar con el tiempo solo nos importa la media de la potencia de señal recibida, y en ella los desvanecimientos de pequeña escala tienen una media de 0 porque a largo plazo tienden a anularse todos los frentes de onda secundarios más débiles.

3.2.3.5. Modelos empíricos de propagación.

Tradicionalmente se han usado métodos empíricos de predicción[1-3], en los cuales se determinan las pérdidas de propagación en función de los muros y pisos a través de los cuales se propaga la onda. El problema de estos métodos es que son aproximados y no pueden predecir fenómenos como el *fast-fading*. Para predicciones más exactas se necesitan métodos deterministas [4, 6]. Los más usados en problemas de comunicaciones en interiores están basados en teoría de rayos con la formulación de Óptica Geométrica y Teoría Geométrica de la Difracción.

Los primeros modelos empíricos desarrollados estaban relacionados con la propagación en el espacio libre, ajustándose ciertos parámetros del modelo de forma estadística a partir de las medidas realizadas [7,10,11]. El problema de estos modelos es la necesidad de realizar medidas previas.

Posteriormente aparecieron modelos más complejos en los que se tenía en cuenta la información específica del edificio [8,9,12].

En primer lugar se definen las pérdidas de propagación como:

$$PL(d)[dB] = 10\log P_r(d_0) - 10\log P_r(d) \quad (3.6)$$

Donde $P_r(d)$ es la potencia recibida a una distancia d y d_0 es una distancia de referencia. En nuestro caso tomamos esa distancia como 1m. Esto se hace para normalizar las pérdidas a dicha distancia de referencia.

Un modelo empírico presentado en [10] calcula las pérdidas de propagación de la siguiente forma:

$$PL(d)[dB] = 10\log \left(\frac{d}{d_0} \right)^n + \sum_{p=1}^P WAF(p) + \sum_{q=1}^Q FAF(q) \quad (3.7)$$

Donde P y Q son el número de muros y pisos que hay entre el transmisor y receptor, $WAF(p)$ y $FAF(q)$ son las atenuaciones por muro y piso respectivamente, y n es el exponente de las pérdidas por trayecto. Para pérdidas en el espacio libre, este último parámetro es igual a 2.

Se ha observado que las pérdidas de propagación, como función de la distancia, tiene dos regiones diferentes [10]. En la primera región las pérdidas de propagación son similares a las del espacio libre. En la segunda región las pérdidas se ven incrementadas a causa de que las ondas electromagnéticas se ven obstruidas por las paredes, techos y suelos. Con el fin de tener en cuenta

este efecto, se considera una distancia de ruptura, en la cual ocurre la transición entre las dos regiones. Incorporando esta distancia a (3.7) tenemos:

$$PL(d)[dB] = 10 \log \left(\frac{d}{d_0} \right)^{n_1} U(d_{bp} - d) + 10 \left[\log \left(\frac{d_{bp}}{d_0} \right)^{n_1} + \log \left(\frac{d}{d_{bp}} \right)^{n_2} \right] U(d - d_{bp}) + 10 \sum_{p=1}^P WAF(p) \cos \theta_q + \sum_{q=1}^Q FAF(q) / \Phi_q \quad (3.8)$$

Donde d_{bp} es la distancia de ruptura, n_1 y n_2 son los exponentes de pérdidas a un lado y otro del punto de ruptura, y $U(\cdot)$ es la función escalón definida como:

$$U(d) = \begin{cases} 0, & d < 0 \\ 1, & d \geq 0 \end{cases} \quad (3.9)$$

Otra diferencia entre las expresiones (3.7) y (3.8), es que en esta última se ha tenido en cuenta el ángulo de incidencia en las pérdidas por muros y pisos. Debido a que las pérdidas por transmisión son mínimas cuando se incide normalmente y se incrementan a medida que el ángulo de incidencia aumenta, hay que tener en cuenta de alguna manera este factor.

Realmente la expresión de atenuación en función del ángulo de incidencia no es trivial, pero se puede aproximar por el coseno de este ángulo, tal como se expresa en (3.8).

3.2.3.6. Modelo determinista de Ray-Tracing.

Otros modelos importantes se basan en el trazado de rayos (*Ray-Tracing*) [9,14,15,16]. La señal de radio en interiores está caracterizada por poseer un fuerte multitrayecto. Sin embargo, en los modelos anteriores no se caracteriza el multitrayecto, con la excepción de la generación de dos caminos alternativos a la línea recta, que si bien no se adecuan demasiado bien a la realidad, dan unas predicciones aceptables de potencia recibida.

Dentro de los métodos deterministas, los más usados en problemas de propagación en interiores son los basados en Óptica Geométrica y Teoría Geométrica de la Difracción y entrazados de rayos.

El programa utilizado en esta tesis, llamado FASPRI[16] está basado en estas técnicas. Utiliza un modelo totalmente 3D con representación geométrica mediante facetas planas, incluyendo para cada una de esas facetas planas las características electromagnéticas del material con el que está construida cada una de ellas. Este código tiene en cuenta efectos de hasta orden cuarto incluyendo múltiples reflexiones y transmisiones.

El problema de los modelos de propagación basados entrazado de rayos, es el elevado tiempo de cálculo requerido. Sobre todo si el modelo geométrico es complejo y exige crearlo con gran cantidad de facetas. Por ello el código FASPRI incluye un algoritmo de aceleración del trazado de rayos, llamado Angular Z-Buffer o AZB [17].

Tal como se puede ver en [16], [17] el programa ha sido validado con medidas, dando resultados bastante satisfactorios.

3.2.3.7. Comparación entre un modelo empírico y otro determinista.

En este apartado se han comparado predicciones obtenidas mediante un modelo empírico y un modelo determinista.

El primer caso que se ha probado es el de un pasillo de 20 metros de largo por 1.5 de ancho, tal como se muestra en la figura 3.1. En esta figura se puede apreciar también la posición de la antena transmisora. Los puntos de observación se encuentran a la misma altura de la antena (1.5m) y con una separación entre ellos de 0.25m. La frecuencia de operación es 2.4GHz.

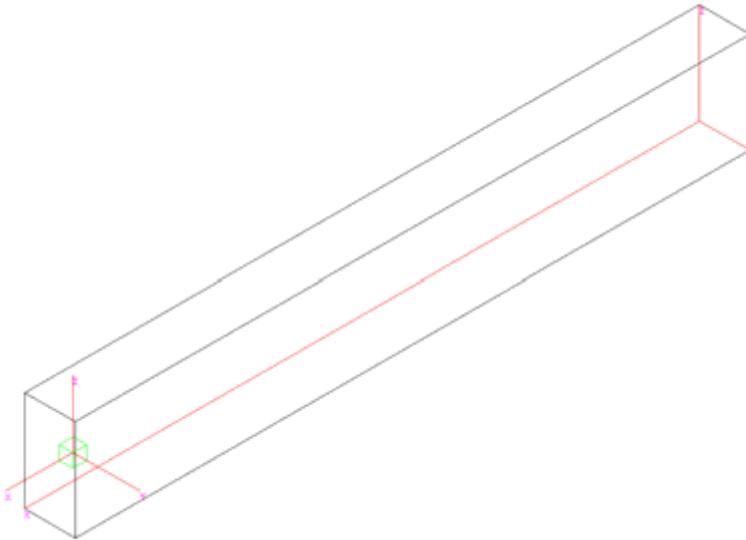


Figura 3.1. Pasillo

En este caso sólo intervienen los mecanismos de propagación por visión directa y reflexiones en los muros y techos. Por lo tanto, en el método estadístico mostrado en la expresión (3), no intervienen los sumatorios de pérdidas por muros. Por otra parte, para esta misma expresión es necesario fijar

los parámetros n_1 , n_2 y d_f . Este último parámetro, la distancia de ruptura depende de a que distancia del foco se obstruya la primera zona de Fresnel. El diámetro de esta última se puede calcular aproximadamente como:

$$Z_f \approx \sqrt{\lambda d} \lambda \quad (3.10)$$

En este caso la distancia de ruptura es de 18m. n_1 como corresponde a la propagación en el espacio libre, su valor debe ser 2. Y para la propagación a través de un pasillo n_2 toma generalmente el valor de 2.5 [8].

En la figura 3.2 se muestran los resultados del FASPRI y del método estadístico respectivamente. En la figura 3.3 se muestra la gráfica de error resultante de restas los resultados de ambos programas.

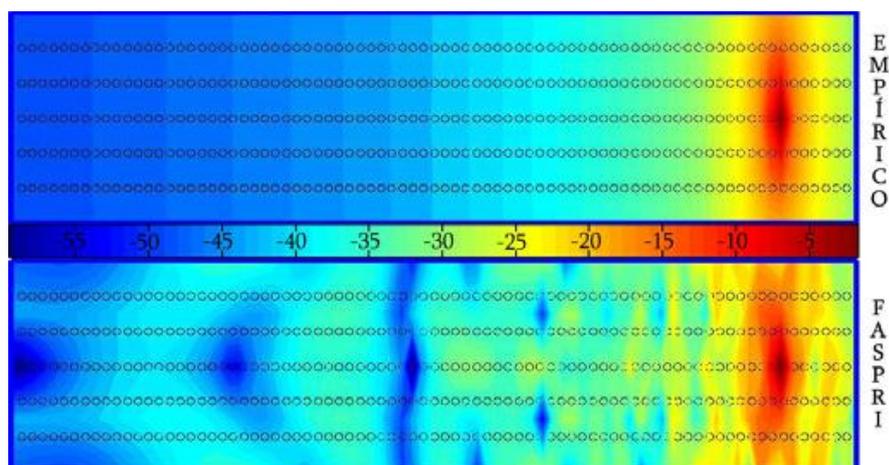


Figura 3.2. Resultados del método empírico y el FASPRI para un pasillo de 20m de largo

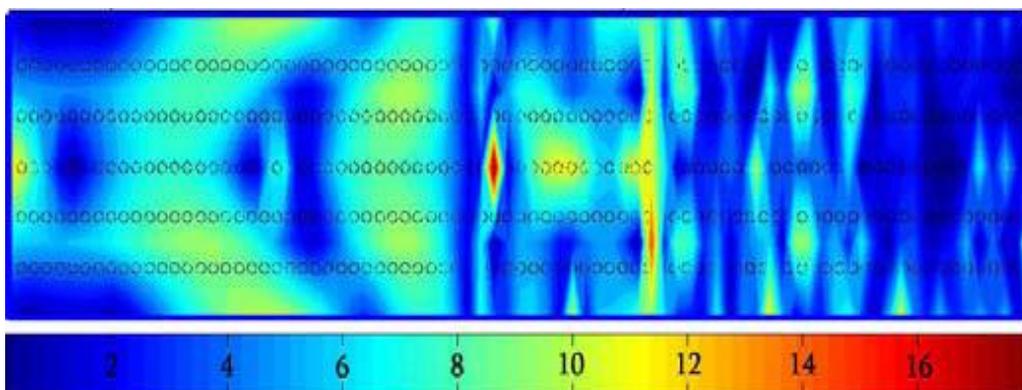


Figura 3.3. Diferencia Método empírico-Faspri en un pasillo.

Se puede observar que, si bien los valores de las pérdidas de ambos métodos son similares, en los resultados del FASPRI aparecen máximos y mínimos locales que no aparecen en los resultados del método empírico. Esto se debe a que en este último método solamente se tienen en cuenta las pérdidas por propagación, mientras que en el método de trazado de rayos se suman todos los multitrayectos con su correspondiente fase. De esta forma, en algunos puntos los rayos se sumaran en fase y en otros en contrafase. En cuanto a la diferencia de resultados entre ambos métodos, se puede observar que salvo en sitios puntuales el error no supera los 10dB. La diferencia media es de 4.89dB con una desviación típica de 1.26dB.

El siguiente caso probado fue un ala del edificio politécnico de la Universidad de Alcalá. Esta es ya una geometría más compleja que la anterior, la cual contiene un pasillo, varios despachos y una sala de reuniones. En la figura 3.4. se muestra el modelo geométrico así como la posición de la antena. Al igual que el caso anterior se ha utilizado un mallado de puntos situados a la misma altura que la antena(1.5m) y con una separación de 0.25 m. La frecuencia de operación, al igual que el caso anterior es de 2.4GHZ.

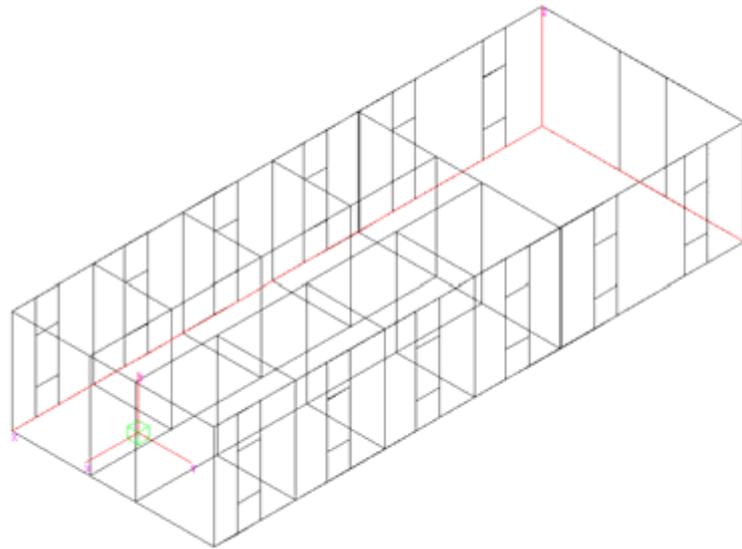


Figura 3.4. Pasillo N22 del edificio politécnico de la Universidad de Alcalá

Para este caso interviene como mecanismo de propagación, aparte de las reflexiones, las transmisiones. Se ha tomado, como material de los muros, el ladrillo. Introduciendo esta información en el FASPRI, las pérdidas por transmisión por muro resultantes, cuando se incide normalmente, son de 6dB aproximadamente. Este valor es el que se introduce como pérdidas a la matriz WAF de la expresión (3.8). En este caso para la distancia de ruptura se toma un compromiso entre el pasillo y la sala de espacio más amplio. El valor tomado es de 30 m.

En la figura 3.5 se muestran los resultados del FASPRI y del método empírico respectivamente, mientras que en la figura 3.6. muestra, en cada punto de observación la diferencia entre ambos programas.

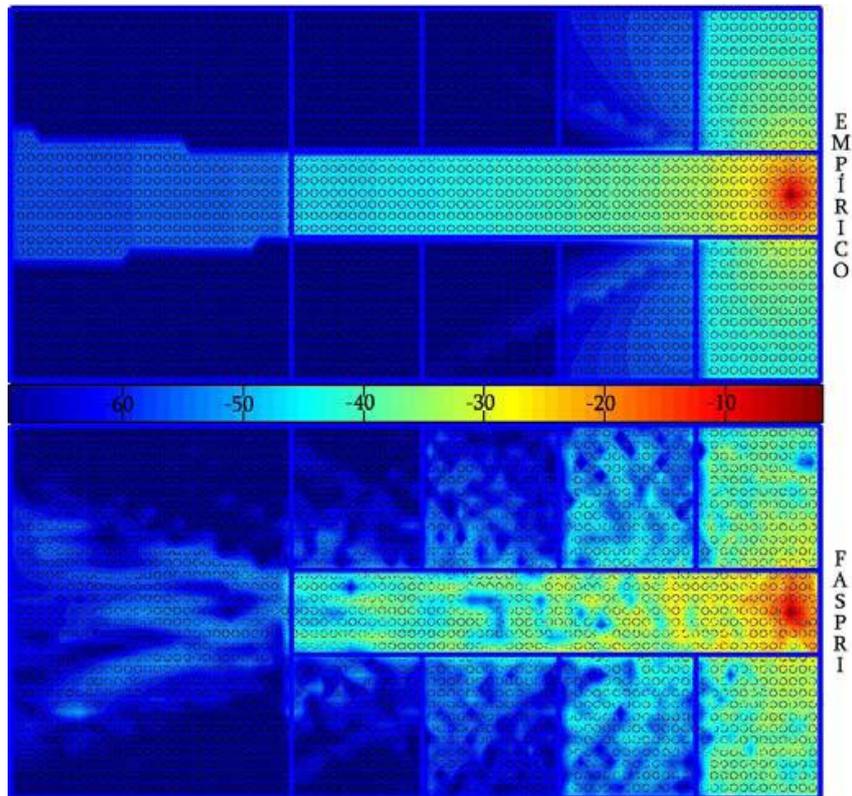


Figura 3.5. Resultados del método empírico y del FASPRI para el ala N22 de la Escuela Politécnica de la Universidad de Alcalá

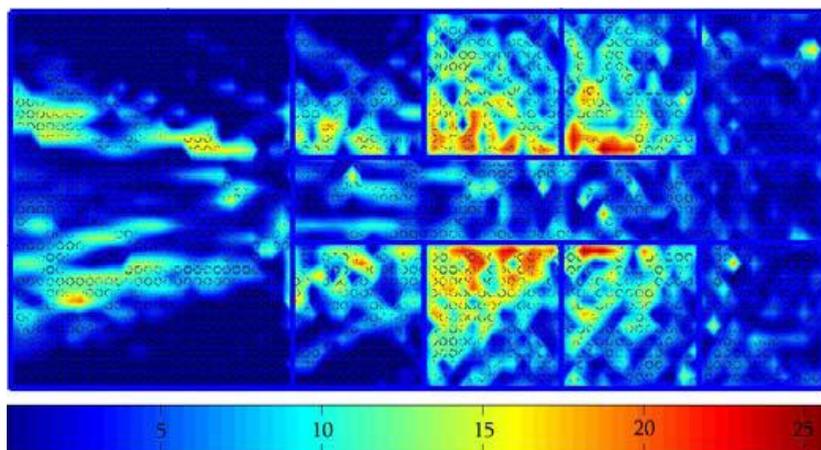


Figura 3.6. Diferencia entre el método empírico y el FASPRI para el ala N22 de la Escuela Politécnica de la Universidad de Alcalá

Se puede apreciar, al igual que en el caso del pasillo, que los resultados del ala N22 presentan máximos y mínimos locales debidos al efecto multitrayecto, que el método empírico no puede predecir. También se puede apreciar que las mayores diferencias entre ambos métodos se tienen en aquellas partes de los despachos donde la transmisión se produce de forma casi rasante. Es decir cuando el ángulo θ_p de la ecuación (3.8) es cerca de 90° y por tanto $\cos(\theta_p)$ cercano a 0. Esto es porque la aproximación tomada no es muy buena para estos casos. Fuera de estos casos, se puede apreciar que la diferencia entre ambos programas está entre 0 y 10db, siendo la media del error de 5.23dB con una desviación típica de 1.18dB.

Comparando ambos métodos de predicción de propagación, un método empírico basado en un modelo de pérdidas en el espacio libre en el que se tiene en cuenta las pérdidas por transmisión, y otro determinista basado en el trazado de rayos, se puede observar que ambos presentan resultados similares, no sólo para escenarios sencillos, como un pasillo, sino para escenarios más complejos compuestos por varias oficinas. En este último caso la diferencia media entre los dos métodos es de unos 5dB's.

3.3. MEDIDAS EN ESCENARIOS REALES.

En el apartado anterior se ha visto como caracterizar la señal mediante un modelo matemático, viendo varios de estos modelos. Estas herramientas son útiles a la hora de diseñar un sistema de localización. Permite determinar el número de puntos de acceso necesarios, su posicionamiento, etc.

Sin embargo, como ya se comentó anteriormente estas herramientas no son capaces de predecir efectos como la presencia humana. En esta tesis se ha realizado un estudio de la variabilidad de la señal en varios entornos. Estas medidas se han realizado en varios escenarios cambiando las condiciones ambientales, principalmente el cambio de tránsito humano en dichas zonas

Con el fin de realizar estas medidas se diseñó una pequeña red inalámbrica compuesta por dos puntos de acceso y un dispositivo móvil. Como puntos de acceso se utilizaron *routers*, standard 802.11. Como dispositivo móvil se utilizó un ordenador portátil equipado con una tarjeta WiFi, también estándar 802.11. Se diseñó un software basado en el lenguaje de programación Java, capaz de leer los parámetros de dicha tarjeta, en concreto la potencia recibida en dBm. Dicho programa fue el encargado de recoger los valores RSS y almacenar los valores, bien para crear el radio-map necesario para el algoritmo del *fingerprinting*, bien para analizar los valores y estudiar la variabilidad de la señal en función de las variaciones ambientales del escenario.

3.3.1. CARACTERIZACIÓN DE LOS ESCENARIOS DE MEDIDAS

Los resultados que se han obtenido para esta tesis corresponden a experimentos realizados en pasillos, despachos y aulas de la Escuela Politécnica la Universidad de Alcalá y en zonas de consultas en el Hospital Príncipe de Asturias, en situaciones de tráfico variable de personas, durante varios días en cada zona y con diversos emplazamientos de la antena receptora previamente definidos generando así un mapeado en cada escenario, que se mantuvo fijo durante cada conjunto de mediciones. Los escenarios seleccionados para realizar las mediciones son los que encontramos comúnmente en edificios. Estos escenarios incluyen recintos cerrados como despachos, aula de clases, laboratorios de computación, consultas clínicas, salas de espera y pasillos.

En todas las mediciones se han mantenido constantes la posición de los puntos de acceso, la potencia de emisión del *router* y la periodicidad de los paquetes enviados. Igualmente, para todas las pruebas realizadas, se han tomado medidas de la distancia entre las huellas y el *router*, y los parámetros de la potencia recibida o RSS en unidades dBm. El número de paquetes

analizados para cada escenario es de 100, de forma que los valores de RSS mostrados corresponden a un promedio respecto a este número.

3.3.2 ELEMENTOS BÁSICOS DEL SISTEMA

El sistema tiene como objetivo dar información sobre la localización de un terminal móvil situado en el interior de un edificio, utilizando información extraída del campo electromagnético propagado entre el terminal móvil y una infraestructura de radiocomunicaciones compuesta por un conjunto de terminales fijos.

El terminal móvil será un equipo portátil con un interfaz de comunicaciones que debe cumplir con el estándar IEEE 802.11 ya que la tecnología utilizada es WiFi. A lo largo de esta Tesis nos referiremos a dicho interfaz como adaptador, tarjeta o dispositivo inalámbrico. Además, los terminales fijos, en este caso, son los puntos de acceso (AP) de la red WiFi situados a lo largo de la zona de cobertura del sistema, usaremos *routers* WiFi. También se precisa de un servidor de localización que tenga almacenados los datos extraídos en la fase de entrenamiento y que realice los cálculos necesarios para proporcionar la localización en la fase de test. Por otra parte, el canal de comunicaciones y la propia estructura del edificio completan los elementos del sistema.

El procedimiento completo llevado a cabo por el sistema basado en fingerprinting con tecnología WiFi, comienza con la toma de medidas en la fase de entrenamiento. Para ello, se utiliza un dispositivo WiFi que se posiciona en un determinado número de puntos que cubren la zona de cobertura del servicio, esta rejilla de puntos, cómo ya hemos comentado en el capítulo anterior es comúnmente conocida como *radio-map*. En cada punto, el adaptador inalámbrico del terminal móvil mide el indicador de potencia de señal recibida (RSSI, de Received Signal Strength Indicator) descrito en el

estándar 802.11. Este indicador es leído por el controlador del dispositivo inalámbrico, de forma que el software de localización puede recopilar un conjunto de valores de RSS denominados valores instantáneos, interrogando a dicho controlador cada cierto tiempo, denominado periodo de muestreo. Posteriormente, el conjunto de medidas proveniente de los diferentes AP es preprocesado para formar el espacio de entrada. Una vez realizado este procedimiento sobre todos los puntos, el resultado se almacena en una base de datos situada en el servidor, que entrena el algoritmo de localización con los datos introducidos.

Terminada esta fase, el sistema ya dispone de suficiente información para proceder a la localización del terminal móvil en la fase de test, en la que el adaptador inalámbrico realiza el mismo procedimiento de medida descrito en la fase anterior. En cada punto, envía la información a través de la red WiFi al servidor, que la introduce en el algoritmo de localización entrenado y estima la posición del terminal móvil. La representación de la información se puede realizar en el servidor o en el terminal móvil, en función de la aplicación para la que se utiliza el sistema.

En la tabla que sigue se proporcionan los rangos de transmisión y recepción de señal WiFi..

Nivel dBm	Potencia	Notas
-70 dBm	100 pW	Típico rango de potencia de recepción de señal de wireless (-60 a -80 dBm)
15 dBm	32 mW	Típica potencia de transmisión WiFi en laptops
20 dBm	100 mW	Bluetooth clase1,y típica potencia de transmisión de <i>router</i> WiFi

Tabla I. Rangos de transmisión y recepción señal WiFi

El estándar IEEE 802.11b/g que cumplen los *routers* WiFi, define las siguientes características de transmisión :

- 802.11/b : tiene una velocidad máxima de transmisión de 11 Mbits/segundo funcionando en la banda 2.4 GHz
- 802.11/g : opera en la misma banda que el anterior pero consiguiendo una velocidad teórica de 54Mbits/segundo.

3.3.3 CONDICIONAMIENTO AMBIENTAL

Vamos a exponer cuáles son las características que afectan a la transmisión WiFi. Factores como reflexión, refracción, difracción, absorción por estructuras, cuerpos humanos, existencia de elementos de metal, etc influyen directamente en la propagación. Así pues, una señal incluso podría alcanzar al receptor por diferentes caminos, cada uno con diferente amplitud y fase. En nuestro caso, se intentó mantener siempre las mismas condiciones físicas (evitando materiales metálicos típicos del mobiliario del laboratorio, consiguiendo controlar la posición de los elementos – puertas, ventanas...–) y , como se ha comentado, se hicieron varias sesiones en diferentes condiciones ambientales. El tiempo (temperatura, humedad ...) y el momento del día (mañana o tarde) son también determinantes. Además, en la banda de 2,4GHz emiten dispositivos bluetooth, algunos teléfonos inalámbricos, Microondas (Hornos), y otros dispositivos menos comunes. Con todo, es la misma frecuencia de resonancia del agua. Recopilando, los factores se podrían resumir en :

- Ambientales : Temperatura, Humedad, Presión, Físicos, Disposición de elementos, Movimiento del dispositivo receptor, Metales, Puertas, Personas
- Influencia de otras frecuencias (otros dispositivos trabajando en 2,4GHz)

La variación puede ser tal, que en un mismo punto, la señal recibida puede variar en más de 50 dBm por el condicionamiento externo. Para ilustrar esto de manera preliminar en este punto se van a presentar varios casos de medidas realizadas.

Todas las mediciones se han realizado en un escenario interior formado por un pasillo central de 30 m de longitud y una serie de despachos laterales, con las estructuras y materiales que se detallan en la figura 3.7.

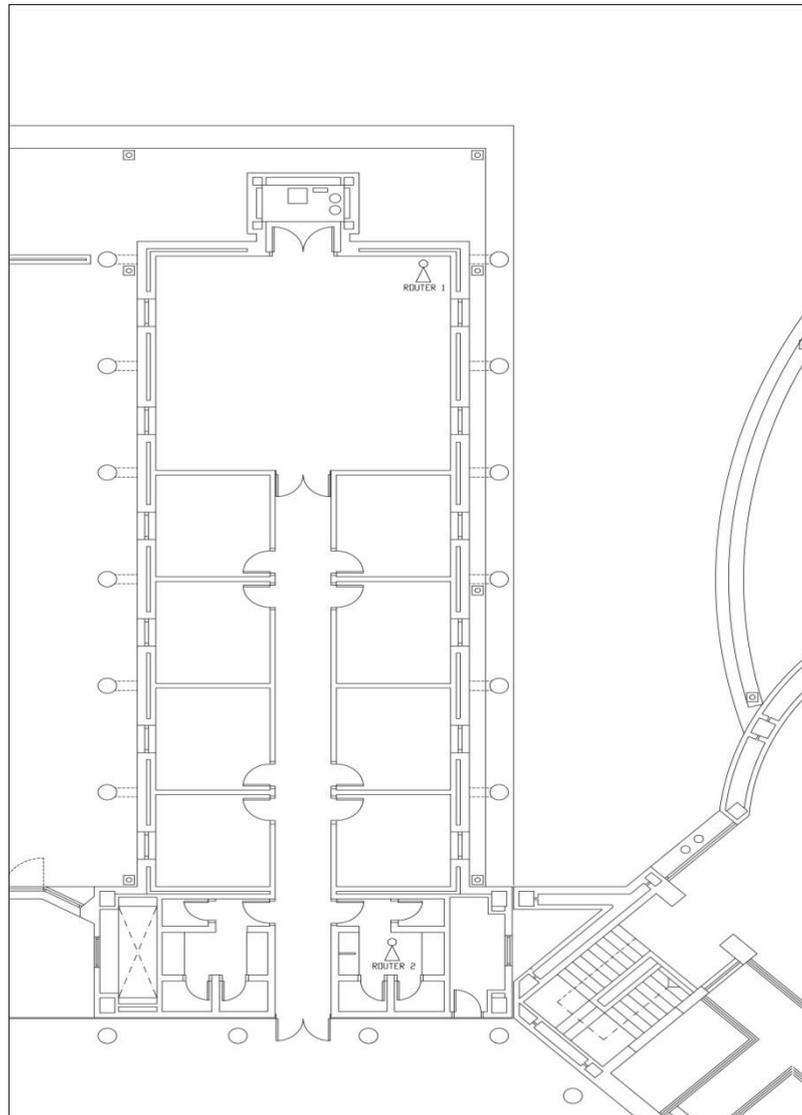


Figura 3.7. Pasillo de despachos, zona N22 del edificio Politécnico de la Universidad de Alcalá

Cómo primer experimento se han tomado medidas en tres condiciones diferentes. En el primer y segundo caso se han tomado medidas manteniendo la posición del *router* y del móvil en la misma posición en condiciones de bajo tránsito de personas y alto tránsito de personas. En el tercer caso la posición del móvil se ha situado cerca del *router* y evitando el tránsito humano.

a) Bajo tránsito humano.

En este caso se ha realizado una toma durante dos horas y media en una estancia de 18m² con poco trasiego de personas. En la gráfica 3.8 se muestra la variación de la potencia recibida con el tiempo.



Figura 3.8. Evolución de señal en trasiego medio-bajo.

Se observan en este caso, las siguientes características :

- Señal relativamente constante a lo largo del tiempo.
- Pico de potencia en negativo producida por movimiento del receptor (PC)

b) Variación por el tránsito de personas.

En este caso se han tomado medidas manteniendo en la misma posición el *router* y el móvil y variando la afluencia humana en la estancia de medida. En la figura 3.9. se observa la variación de la potencia recibida en este caso.

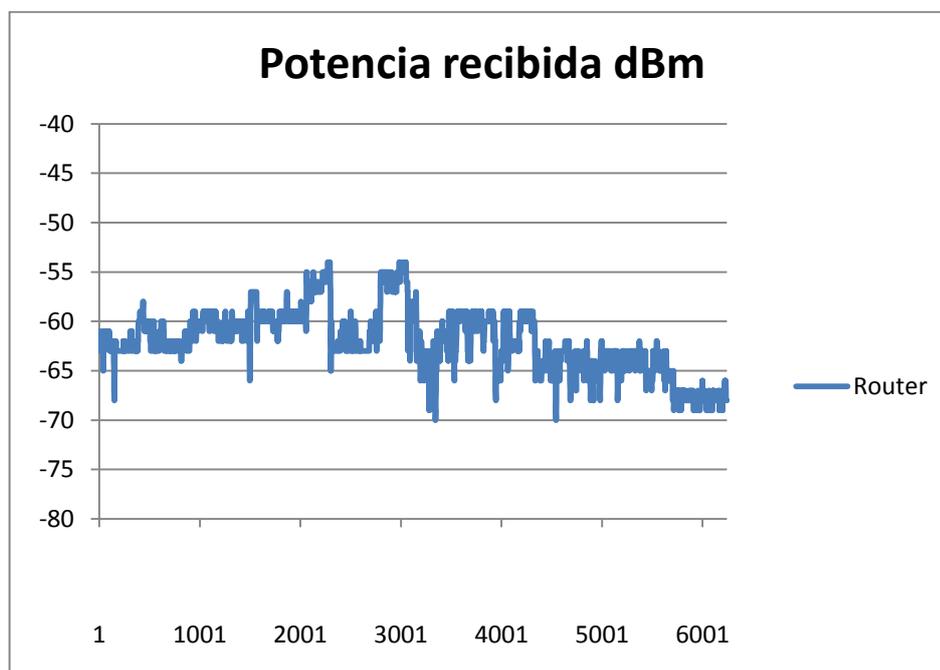


Figura 3.9. Evolución de señal con modificación del trasiego de personas.

De la gráfica anterior se observan los siguientes resultados :

- La señal se mantiene en torno a -60dBm durante aproximadamente el 70% del tiempo.
- Hay múltiples picos de potencia : la mayoría son leves disminuciones de la misma, debido a la intercepción por parte del propio trasiego. En dos momentos, se observa que aumenta bruscamente debido a desplazamientos realizados intencionadamente, y devueltos de nuevo a su posición original.
- Hay una disminución en torno a la primera parte del último tercio (entre la muestra 4251 y 5500) debido a que se ha llenado la estancia deliberadamente.

- La última parte de la gráfica, corresponde a una bajada brusca y constante de la señal que se mantiene hasta el final. Esto es debido al cierre intencionado e la puerta de la habitación donde estaba situado el Pc de medida.

c) Tercer caso: Sin tránsito humano.

En este segundo caso se va a presentar el resultado de una sesión realizada en un punto relativamente cerca del punto de acceso, o *router*, en zona de muy bajo tránsito de personas y durante cuatro horas y media, dos capturas por minuto:

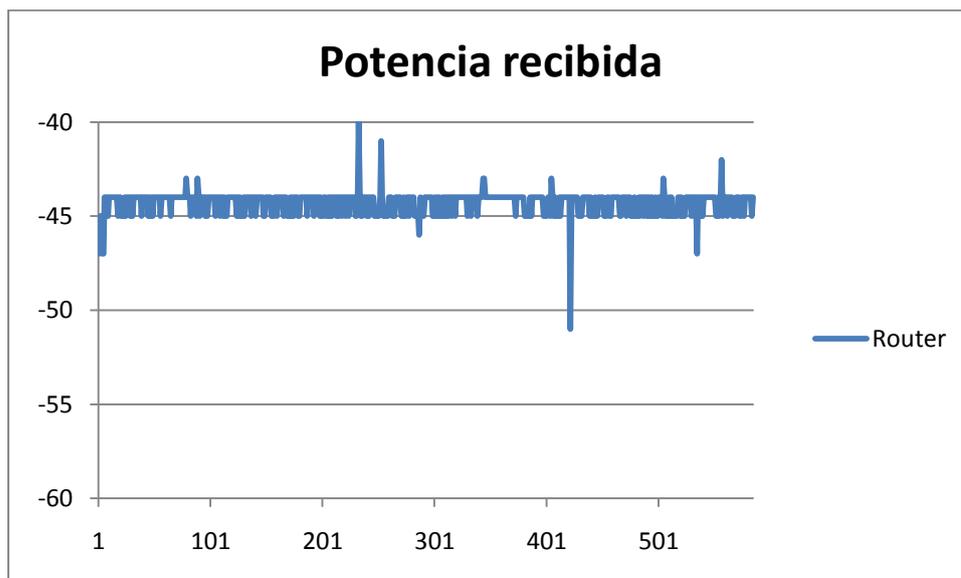


Figura 3.10. Evolución de medida en zona de muy bajo tránsito

A la vista de los resultados, hay poco que decir. La señal se mantiene constante pero se van a comentar dos características de la misma :

- En la última parte de la mitad del gráfico, entre las tomas 361 y 391, se alcanza una evolución perfecta, sin esa variación base que tienen todas las señales realizadas

- Hay pocos picos, pero éstos serán seguramente debido al tránsito muy leve de personas.

A continuación se muestra un análisis estadístico de los tres casos. En primer lugar, en las figuras 3.11 a 3.13 se muestran las gráficas de frecuencias de las medidas de los tres casos:

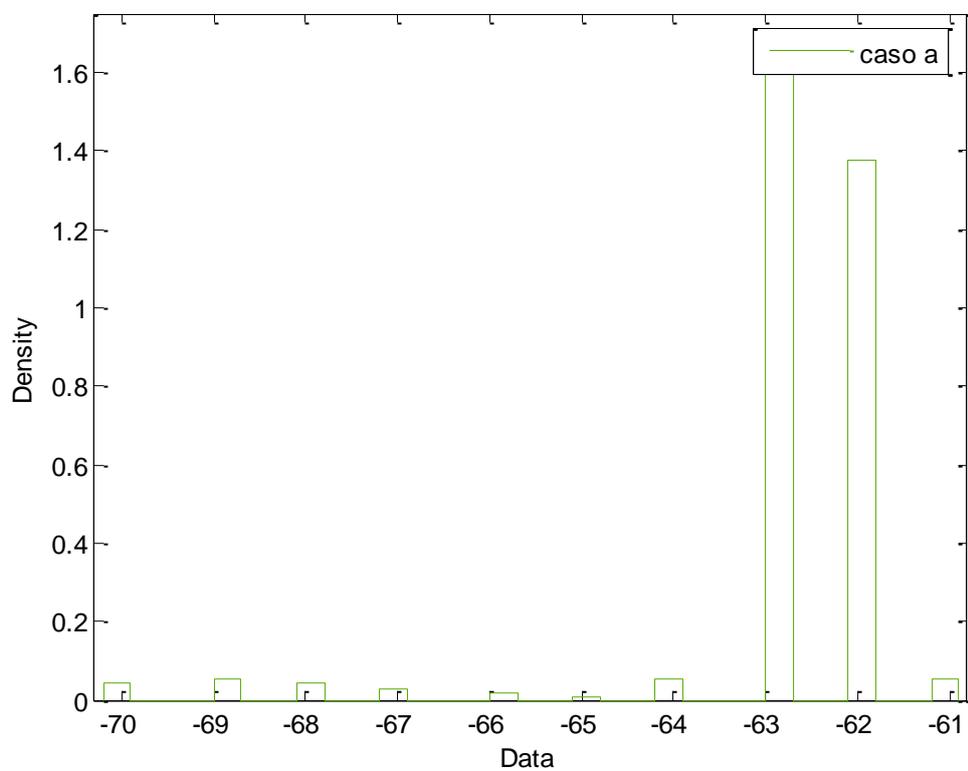


Figura 3.11. Evolución de señal en tránsito medio-bajo.

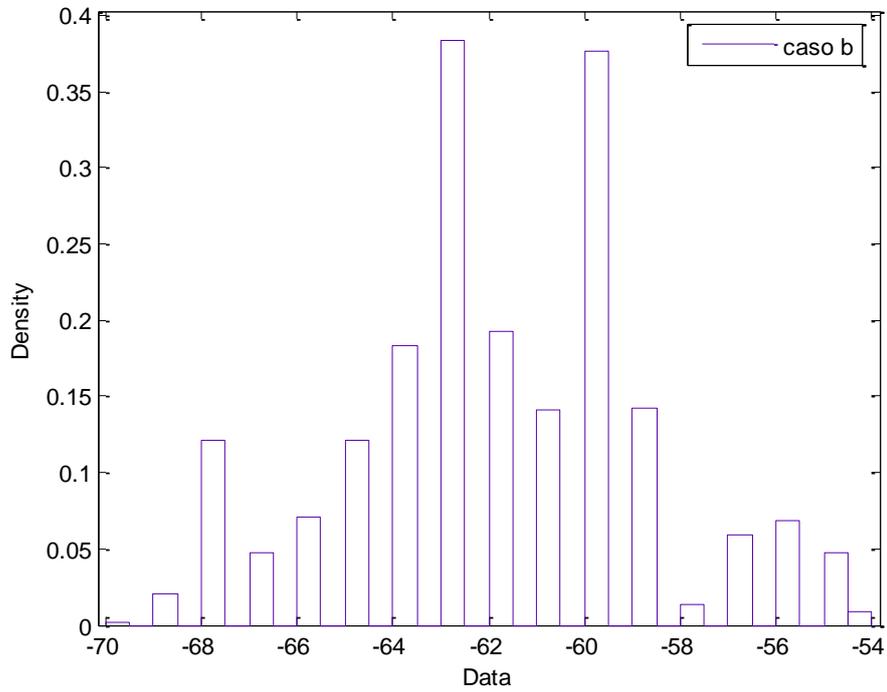


Figura 3.12. Evolución de señal con modificación del tránsito de personas.

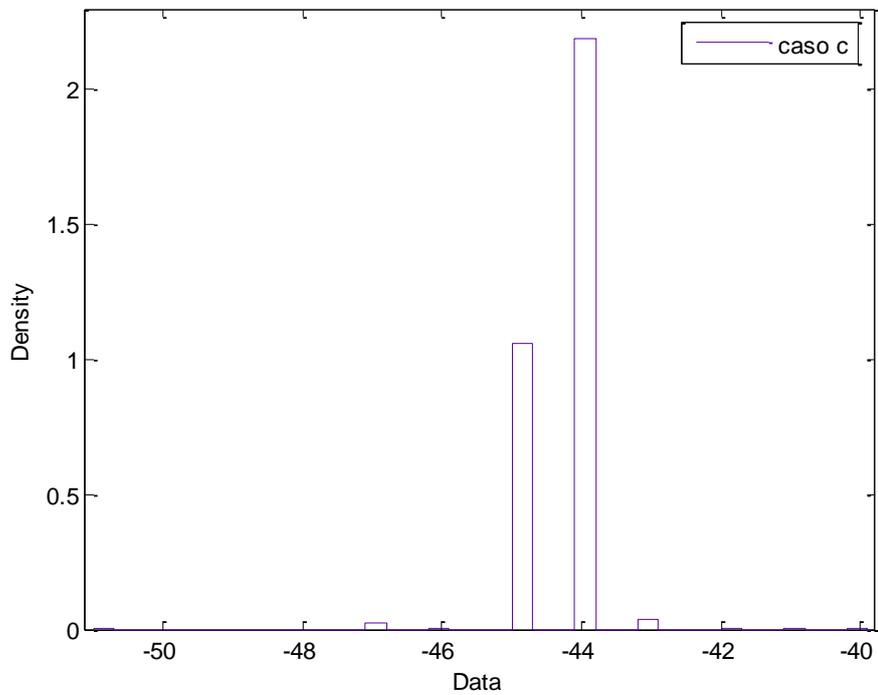


Figura 3.13. Evolución de medida en zona de muy bajo trasiego

Se puede observar que las distribuciones de frecuencia en los casos a y b se mantiene casi sin apenas dispersión, mientras, que en el caso b, por las causas comentadas anteriormente, los valores son muy dispersos. En este último caso aparecen dos modas causado por la variación de la presencia humana en el entorno. Dado que en los casos a y b se mantienen las mismas posiciones de transmisor y receptor, comparamos ambos casos ajustando las medidas a una distribución normal en el primer caso y a una normal con dos modas en el segundo. El resultado se muestra en la figura 3.14.

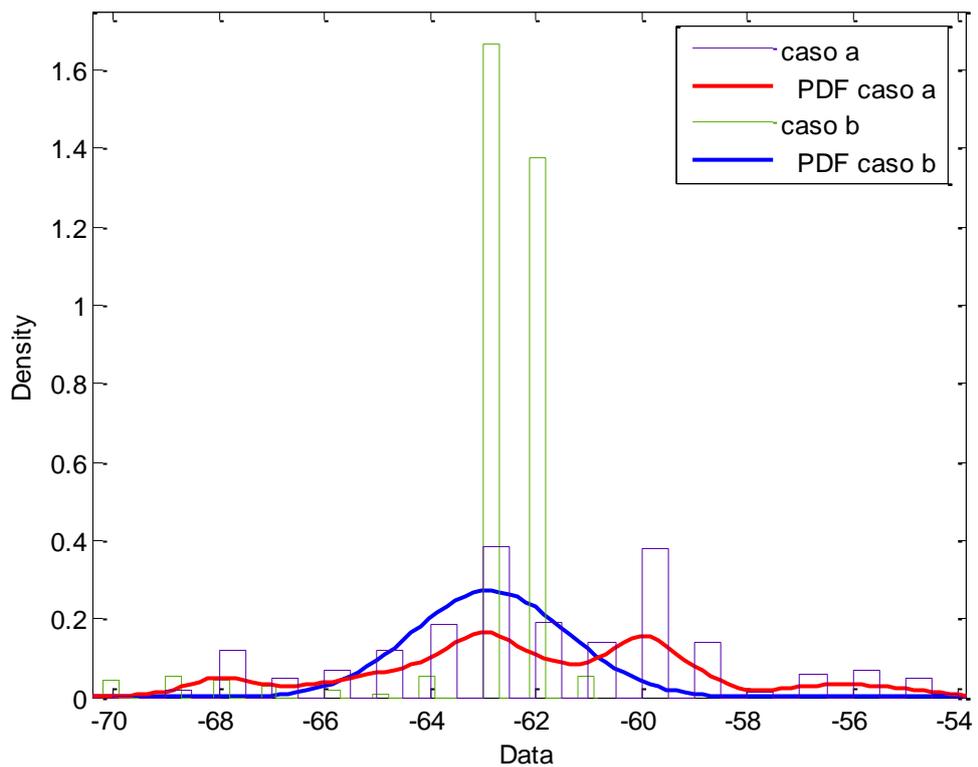


Figura 3.14. Comparación casos a y b

Se puede observar que la media en el primer caso coincide con una de las medias de la distribución multimodal del segundo caso. Esa desviación de la media se debe al incremento del tránsito humano variando en ese caso la señal. Este efecto puede influir sobre la precisión de la localización. Aparte de

desviarse la media de la señal recibida por la variación del tránsito humano, también se puede apreciar un incremento de la desviación típica, en el caso a este valor es de 2.16 mientras que en el caso b se incrementa hasta un valor de 9,9

Una de las conclusiones que se pueden obtener de este primer experimento es que la influencia humana es uno de los aspectos más importantes sobre el nivel de la señal recibida y por tanto de la precisión del algoritmo de localización usando la técnica del *fingerprinting*.

3.3.4. DESCRIPCIÓN DE LAS SESIONES REALIZADAS

Las sesiones realizadas de obtención de *fingerprints* han sido generalmente bajo las siguientes condiciones espaciales y temporales:

- Medidas realizadas en pasillo de departamento (Edificio Politécnico , Universidad de Alcalá de Henares, Zona Norte, Planta dos Pasillo dos,)
 - Fechas : 17 –Abril, 20 –Abril, 13 –Mayo, 28 –Mayo , 2 –Junio , 5 –Junio, 8 –Junio, 24 –Junio
 - Condiciones : Series de medidas en horas similares y alternando mañanas y tardes para comprobar la variación en distintas horas del día y con diferente afluencia de personas. Realizada con dos *routers*

En las figuras 3.15 y 3.16 se muestra un plano general de la zona de medida con los puntos en los que han tomado valores de potencia y la posición de los puntos de acceso utilizados :

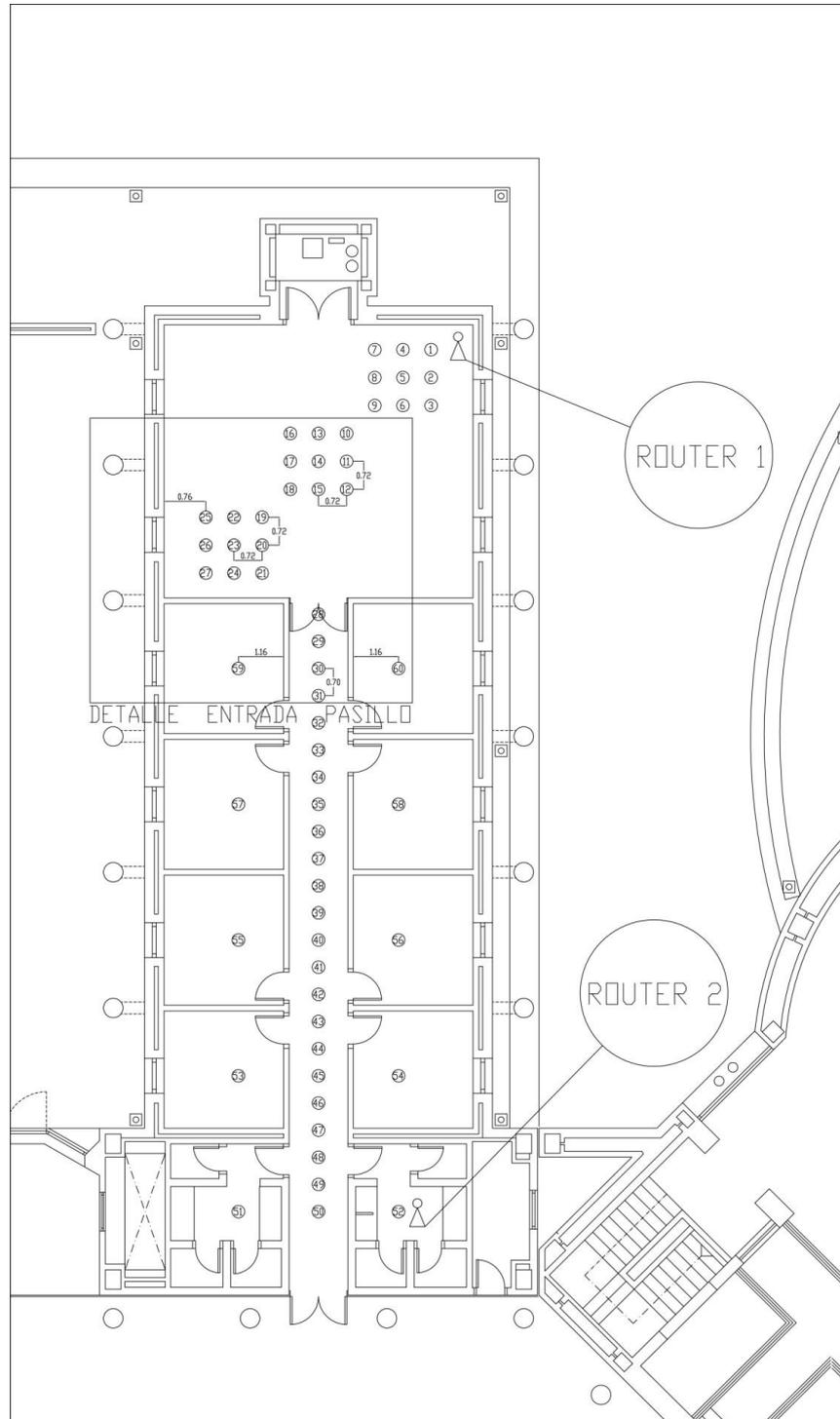


Figura 3.15. Pasillo de despachos Zona Norte, primera planta del edificio Politécnico de la Universidad de Alcalá.

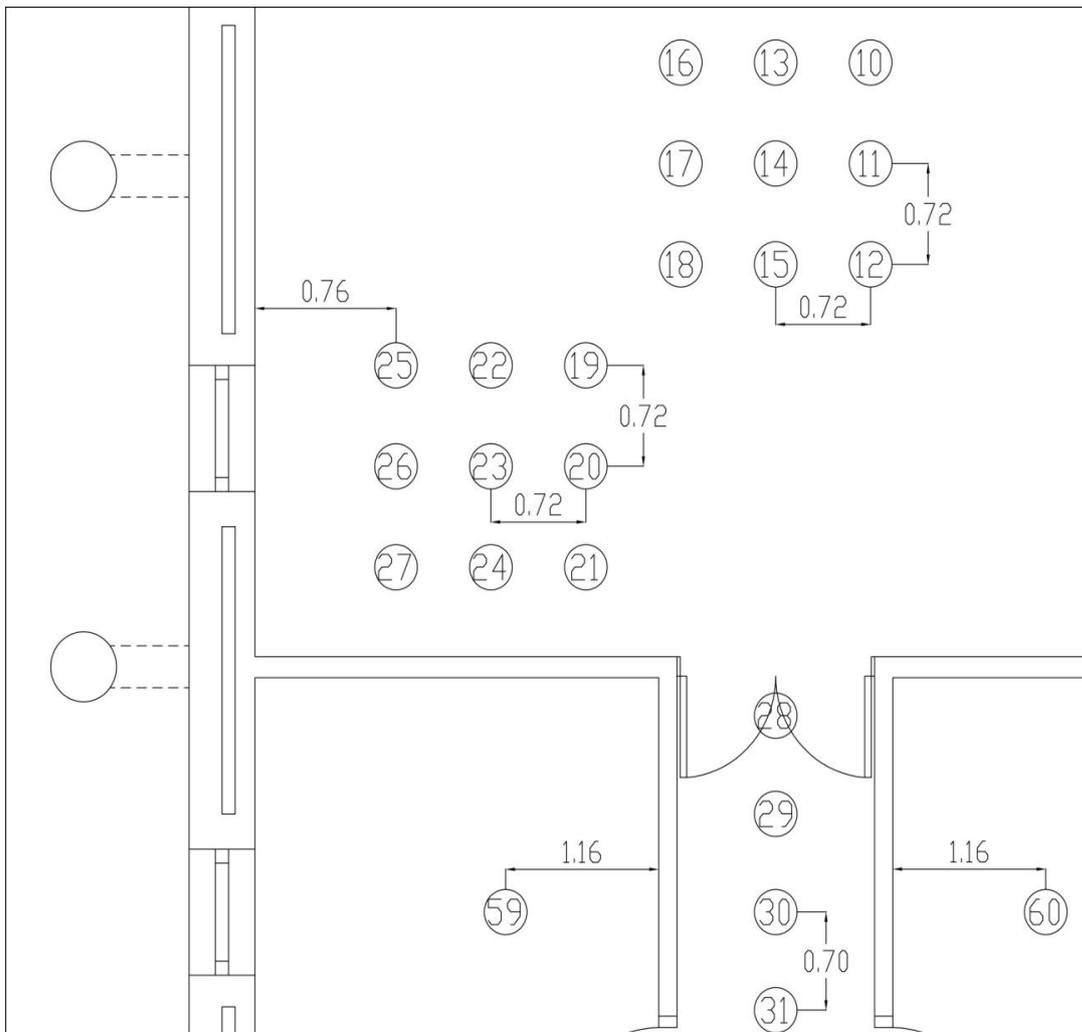


Figura 3.16. Zona con detalle de *fingerprintings* donde se muestra las distancias entre huellas.

- Medidas realizadas en planta baja de la universidad, zona Este (Edificio Politécnico , Universidad de Alcalá) :
 - Fechas : 9 – 10 de Junio
 - Condiciones : Huellas tomadas con distinta afluencia de personas.

En la figura 3.17. se aprecia una vista de la planta baja del edificio Politécnico. En la figura 3.18. se puede ver con más detalla la zona Este de dicho edificio y en las figuras 3.19 y 3.20 se muestran en detalle las zonas de esta parte del edificio donde se situaron los puntos de medida:

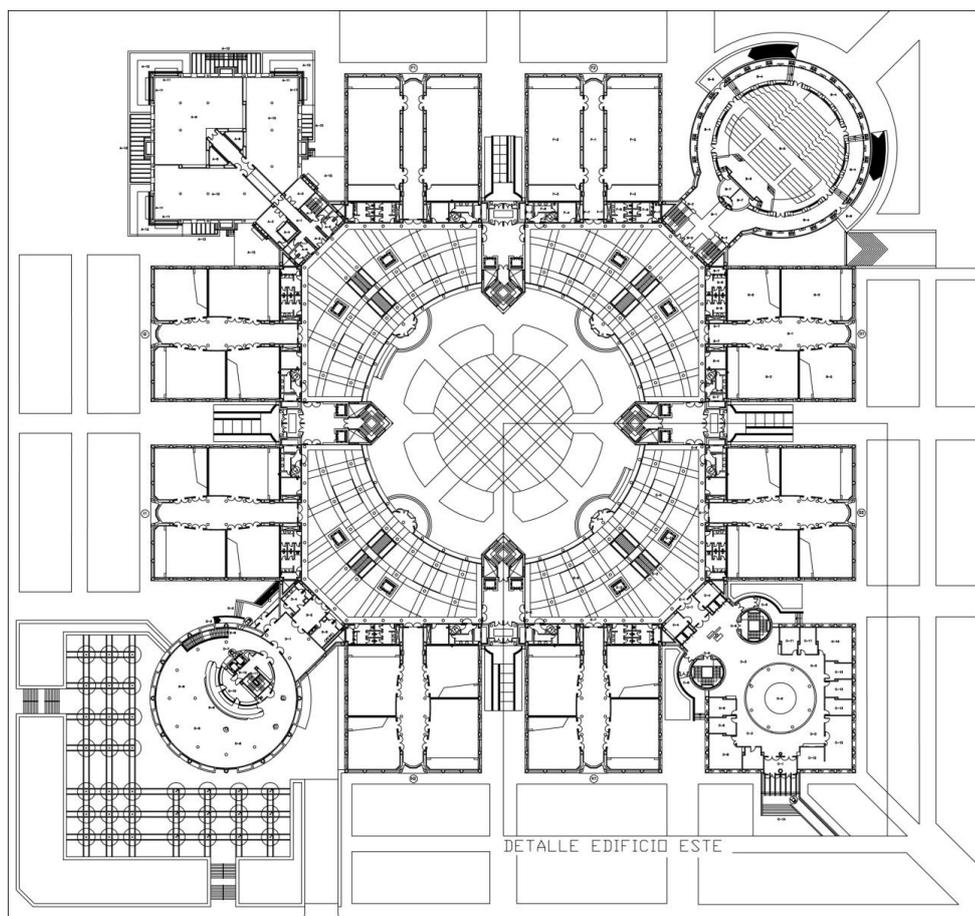


Figura 3.17. Planta baja del edificio Politécnico de la Universidad de Alcalá

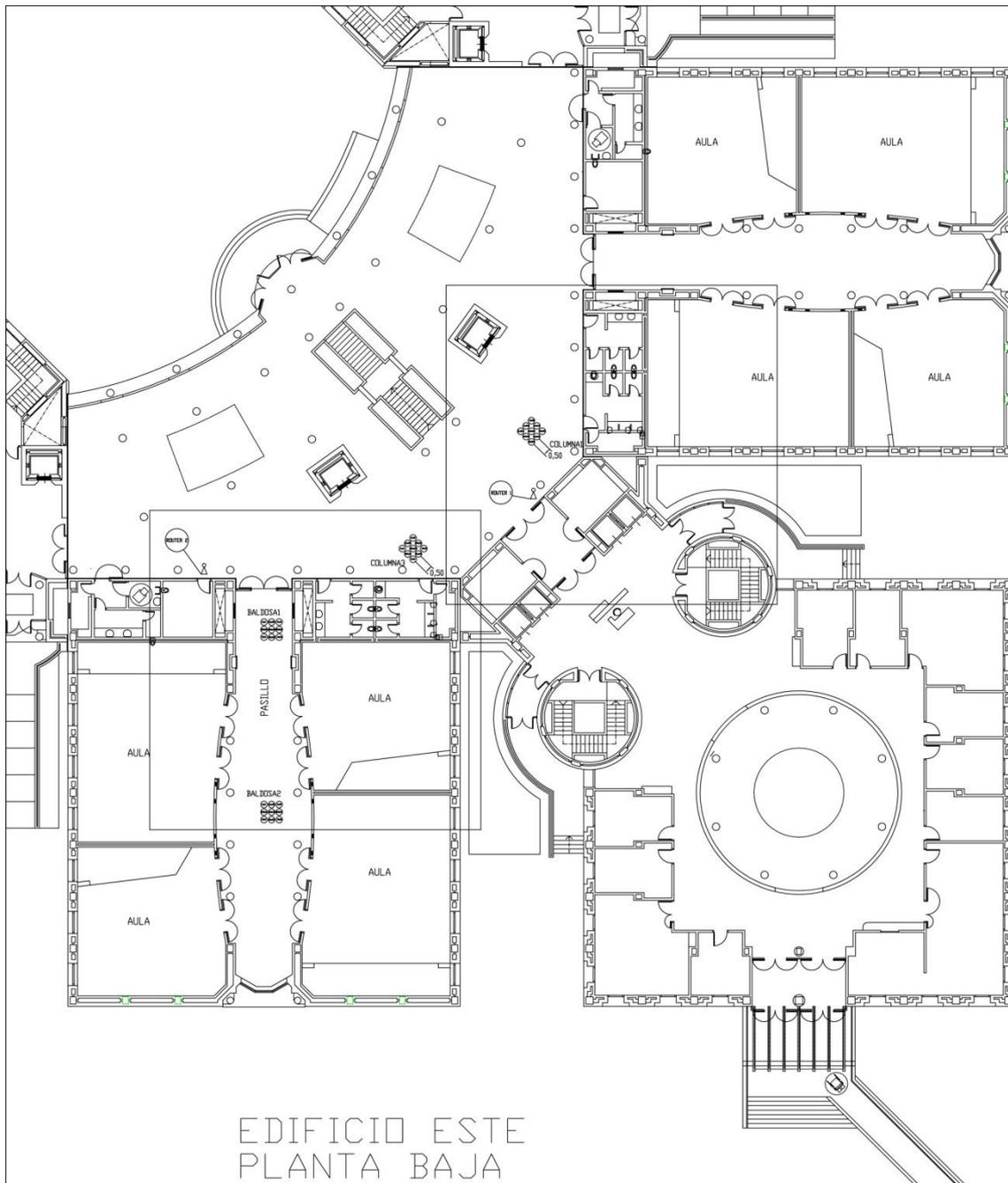


Figura 3.18. Zona Este. Edificio politécnico. En blanco, marcados los puntos de captura

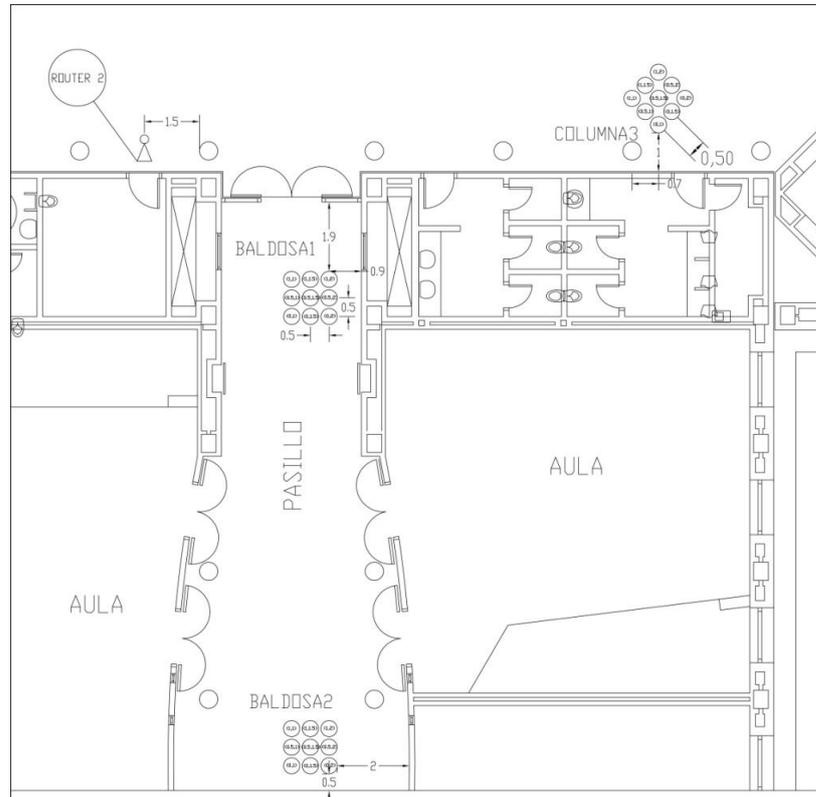


Figura 3.19. Zona Este de pasillo y columna 3.

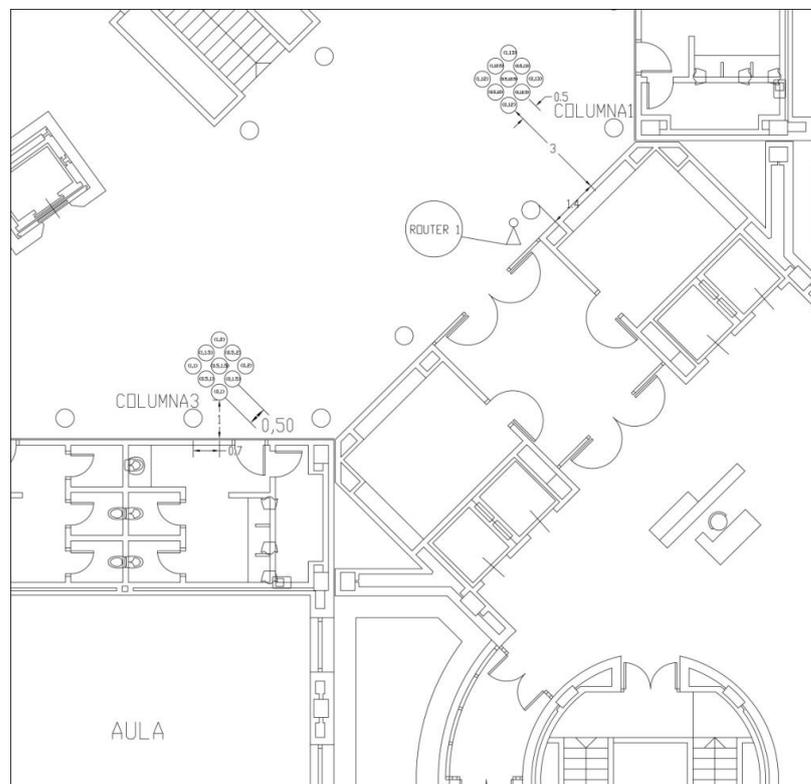


Figura 3.20. Zona Este de columna 1 y columna 3.

- Medidas realizadas en el Hospital del Henares, sala de extracciones:
 - Fechas : 11 – 12 de Junio
 - Condiciones : Huellas tomadas con distinta afluencia de personas.

En la figura 3.21. se aprecia el plano de la zona del hospital donde se realizaron las medidas. Se puede apreciar el detalle de las posiciones de medida y la colocación de los puntos de acceso.

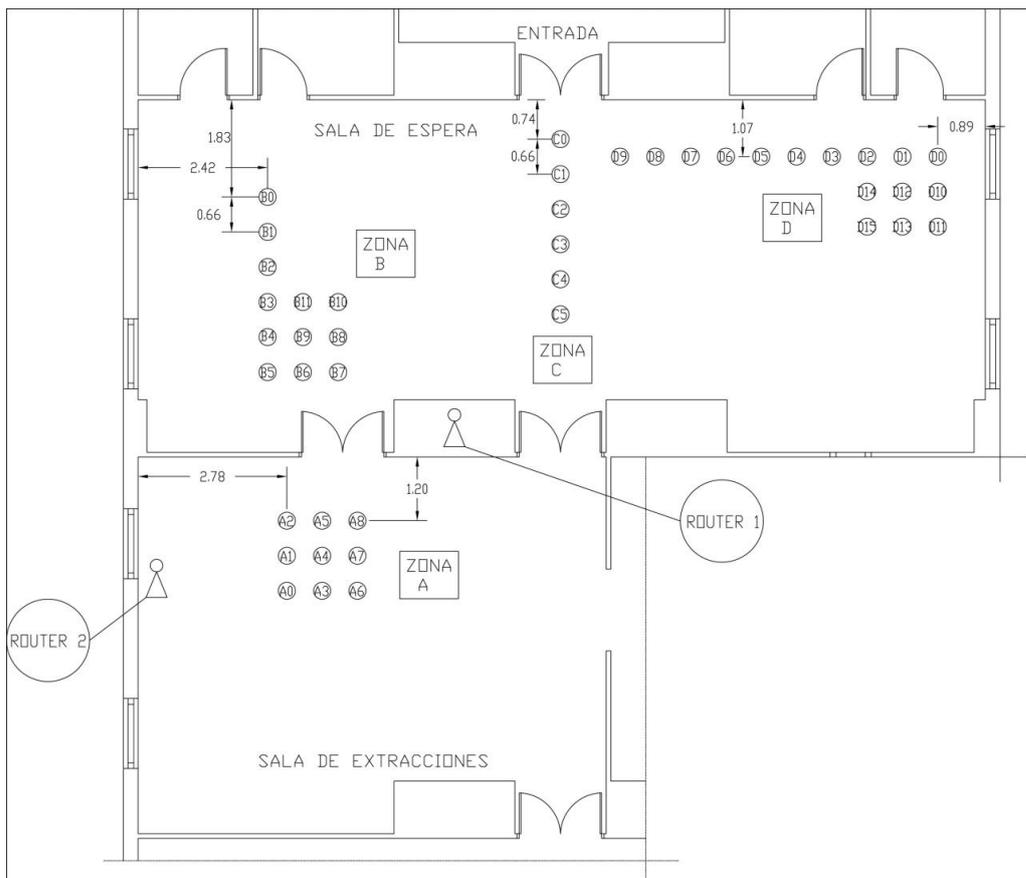


Figura 3.21. Zona del hospital, sala de extracciones con detalle puntos de captura y routers.

- Medidas realizadas en el Hospital del Henares, sala de extracciones :
Las tomas de huellas correspondientes a estos planos serán los correspondientes a las fechas: 11-6-09 (2) y 17-6-09.
 - Fechas : 11 – 17 de Junio
 - Condiciones : Huellas tomadas con distinta afluencia de personas, entre las 10:30 y las 13:00 horas

En la figura 3.22. se muestra la zona del hospital correspondiente a la sala de oftalmología donde se realizaron las medidas. Se detalla la posición de los puntos donde se tomaron las medidas la posición de los puntos de acceso.

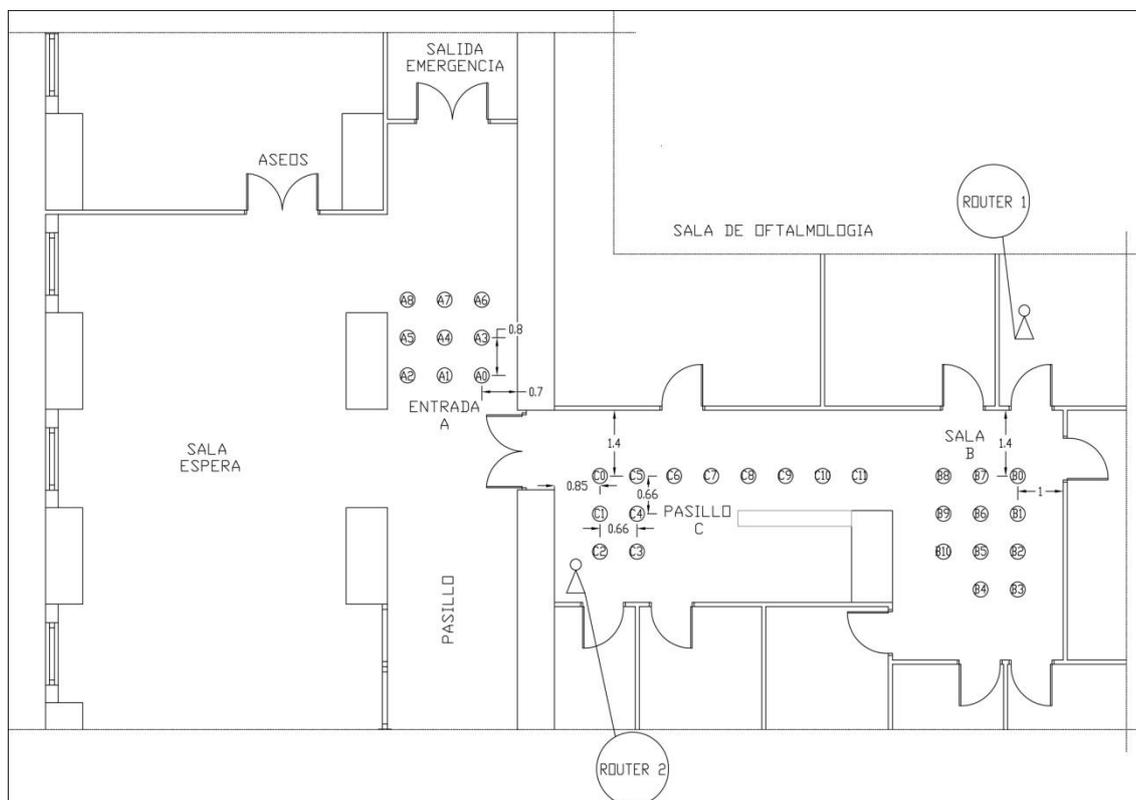


Figura 3.22. Zona del hospital, sala de oftalmología con detalle puntos de captura de huellas.

Condiciones de distribución de los puntos de acceso :

Se ha procurado que no se sitúen en posiciones complementarias pues, de esa forma, reportarán las mismas medidas en puntos opuestos cada uno. Así , se ha intentado situarlos en la misma línea.

Condiciones de los *Router* :

Los puntos de acceso, IEEE 802.11g están configurados para trabajar en radio frecuencias en un banda libre a 2.4GHz, emitiendo en el canal 3 .Para mantener igualdad de condiciones en las sesiones, ambos pc's estaban conectados (inalámbricamente) a uno de los *router* pues así sería en una situación normal : el sistema final obliga a tener una LAN WiFi.

Condiciones de recogida de datos :

Cada dispositivo de medida (de un total de dos) es un PC equipado con tarjeta Wireless 802.11 b/g que realiza su función mediante la librería JWiFiApi [18] y su posición será paralelo a la línea imaginaria que forman los dos *routers*, pues dicho dispositivo WiFi se sitúa en un lateral del Pc.

Tiempo de recogida :

Los datos se recogieron en tres series distintas :

- 10 capturas separadas en 2 segundos cada una
- 100 capturas segundo a segundo
- 8 capturas cada 5 segundos

Las tres configuraciones se aplicaron en el pasillo del departamento ; en el

hospital y la planta baja del edificio politécnico se utilizó la última opción.

Distribución de los puntos de medida :

Para poder obtener resultados correctos, se ha tenido que distribuir la superficie en puntos equidistantes, manteniendo una lógica que sirviera para poder establecer un tanto un estudio de la variación de señal como una posible tabla de referencia que nos pudiera servir en el propio sistema final

3.3.5. MEDIDAS REALIZADAS

A continuación se detallan las medidas realizadas en cada una de las sesiones descritas anteriormente. Las medidas se muestran en la tablas que se detallan a continuación. Cada una de las tablas tiene 4 columnas. Las columnas 1 y 3 corresponden al nivel de potencia captado por la tarjeta WiFi del ordenador portátil en unidades de dBm. La columna 1 corresponde con los valores medios de potencia recibidos del *router* 1 denominado WLAN y la columna 3 a los valores medios de potencia recibidos del *router* 2 denominado JAZZ. Las columnas 2 y 4 corresponden a las varianzas de los valores de potencia recibidos de los *routers* WLAN y JAZZ respectivamente

En primer lugar en las tablas II y III se muestran los valores de las medidas de potencia, en dBm en el pasillo N22 del Edificio Politécnico realizadas el 2 de Junio. En la tabla II las medidas corresponden a la zona de pasillos y la tabla III corresponde con los 9 puntos de la esquina superior derecha de la figura 3.15, denominados zona 3. Las tablas IV y V corresponde con las medidas del día 9 de Junio y las tablas VI y VII con las realizadas el 28 de mayo.

MEDIA_WLAN	VARIANZA_WLAN	MEDIA_JAZZ	VARIANZA_JAZZ
-51,700	0,458	-38,000	0,000
-54,000	0,000	-37,000	0,000
-56,500	0,592	-48,000	0,000
-57,800	0,748	-48,000	0,000
-54,900	0,300	-49,000	0,000
-60,150	0,963	-51,400	2,939
-54,250	0,994	-41,000	0,000
-55,400	0,860	-52,000	0,000
-50,350	0,476	-51,400	1,200
-52,150	0,909	-49,000	0,000
-52,400	1,020	-54,000	0,000
-51,450	0,804	-54,000	0,000
-41,800	0,400	-57,000	0,000
-50,850	0,476	-62,000	0,000
-49,200	0,980	-61,700	0,458
-47,600	0,490	-61,000	0,000
-41,100	0,436	-63,000	0,000
-43,400	0,583	-58,500	4,975
-36,050	1,161	-52,250	1,299
-31,500	0,592	-50,000	0,000
-33,000	0,000	-59,600	4,800
-39,200	0,678	-62,000	0,000

Tabla II. Medidas de potencia en el pasillo de despachos correspondiente al 2 de junio

MEDIA_WLAN	VARIANZA_WLAN	MEDIA_JAZZ	VARIANZA_JAZZ
-71,550	0,498	-34,000	2,000
-63,550	0,805	-38,000	0,000
-59,700	0,714	-39,000	0,000
-71,100	0,889	-39,000	0,000
-69,450	0,589	-35,000	0,000
-62,400	0,490	-32,000	0,000
-70,000	1,049	-35,000	0,000
-66,550	0,590	-33,000	0,000
-71,700	1,382	-35,000	0,000

Tabla III. Medidas de potencia en la zona 3 de despachos correspondiente al 2 de junio.

Capítulo 3: Características de la señal de propagación en interiores

MEDIA_WLAN	VARIANZA_WLAN	MEDIA_JAZZ	VARIANZA_JAZZ
-54,650	0,476	-40,000	0,000
-59,050	0,498	-46,000	0,000
-57,900	0,700	-49,600	1,200
-55,750	0,622	-48,000	2,449
-53,100	0,300	-45,000	0,000
-53,800	0,980	-46,000	0,000
-58,300	0,781	-67,000	21,000
-58,500	0,500	-88,000	0,000
-54,150	0,358	-49,000	0,000
-48,650	0,654	-48,000	0,000
-47,100	0,436	-51,000	0,000
-50,650	0,477	-58,200	3,600
-48,200	0,400	-60,000	0,000
-48,050	0,219	-59,200	0,980
-44,950	0,217	-58,800	0,980
-47,400	0,490	-60,000	0,000
-42,050	0,219	-60,350	0,476
-42,300	0,640	-61,000	0,000
-35,850	0,476	-61,600	0,490
-35,950	0,497	-62,000	0,000
-35,150	0,358	-65,000	0,000
-38,850	0,356	-63,000	0,000

Tabla IV. Medidas de potencia en el pasillo de despachos correspondiente al 9 de junio.

MEDIA_WLAN	VARIANZA_WLAN	MEDIA_JAZZ	VARIANZA_JAZZ
-64,950	1,071	-36,000	0,000
-63,200	0,872	-35,600	0,800
-65,450	0,497	-34,000	0,000
-66,150	1,352	-36,000	0,000
-63,000	0,000	-35,000	0,000
-69,500	1,245	-35,000	0,000
-70,450	0,497	-37,000	0,000
-71,000	0,000	-32,500	1,500
-71,650	0,476	-32,800	1,600

Tabla V. Medidas de potencia en la zona 3 de despachos correspondiente al 9 de junio.

MEDIA_WLAN	VARIANZA_WLAN	MEDIA_JAZZ	VARIANZA_JAZZ
-56,400	0,490	-39,000	0,000
-54,050	1,244	-50,000	0,000
-53,900	0,539	-20,000	0,000
-54,700	2,170	-52,000	0,000
-58,850	0,653	-49,000	0,000
-59,100	1,136	-50,800	1,470
-54,650	0,937	-40,200	14,885
-57,250	0,767	-52,000	0,000
-54,650	1,038	-50,350	2,445
-51,350	1,590	-53,500	0,500
-56,050	2,784	-53,000	0,000
-53,150	1,062	-50,000	0,000
-51,250	0,434	-60,000	0,000
-51,300	0,781	-59,000	1,000
-50,900	0,831	-58,000	0,000
-51,800	1,600	-61,000	0,000
-50,250	1,337	-60,950	0,217
-46,500	1,323	-60,000	0,000
-41,950	0,804	-67,000	0,000
-47,400	1,020	-59,500	7,500
-34,950	0,740	-52,000	0,000
-30,550	0,497	-65,000	0,000

Tabla VI. Medidas de potencia en el pasillo de despachos correspondiente al 28 de mayo.

MEDIA_WLAN	VARIANZA_WLAN	MEDIA_JAZZ	VARIANZA_JAZZ
-69,400	0,490	-37,000	0,000
-66,650	0,476	-37,000	0,000
-63,200	0,980	-33,000	0,000
-60,950	0,864	-39,000	0,000
-74,650	0,653	-38,000	0,000
-64,750	0,434	-40,000	0,000
-71,800	0,927	-39,000	0,000
-71,100	0,300	-34,000	0,000
-72,200	0,400	-34,900	2,142

Tabla VII. Medidas de potencia en la zona 3 de despachos correspondiente al 28 de mayo.

Si comparamos los valores de todas las tablas., vemos que la media de potencias varía según el día que se hicieron las medidas. En primer lugar calculamos las diferencias en dBm de la potencia media recibida en cada huella, y determinados la media de diferencias entre todas las huellas y tenemos los siguientes resultados:

Zona 3:

- Diferencias medias entre las medidas del día 28 de mayo y el 2 Junio: 3.70 dBm para el router WLAN y 2.90 dBm para el router JAZZ
- Diferencias medias entre las medidas del día 28 de Mayo y las medidas del 9 de Junio: 3.75 dBm para el router WLAN y 2.22 dBm para el router JAZZ
- Diferencias medias entre las medidas del día 2 Junio y las medidas del 9 de Junio: 4.02 dBm para el router WLAN y 2.23 dBm para el router JAZZ

Pasillo:

- Diferencias medias entre las medidas del día 28 de mayo y el 2 Junio: 4.03 dBm para el router WLAN y 4.82 dBm para el router JAZZ
- Diferencias medias entre las medidas del día 28 de Mayo y las medidas del 9 de Junio: 4.43 dBm para el router WLAN y 6.99 dBm para el router JAZZ
- Diferencias medias entre las medidas del día 2 Junio y las medidas del 9 de Junio: 2.86 dBm para el router WLAN y 6.24 dBm para el router JAZZ

Las tablas VIII a XI muestran los valores medios de potencia medidos el día 9 de Junio en el edificio Politécnico de la Universidad de Alcalá, en la zona Este lugar de celebración de los exámenes de selectividad durante esas fechas. Las tablas VIII y IX corresponden a las zonas de medida denominadas columna 1 y columna 3 en la figura 3.20. y las tablas X y XI con las zonas denominadas baldosa 1 y baldosa 2 en la figura 3.19. Las tablas XII a XV corresponde con las medidas realizadas ese mismo día en las mismas zonas. Las tablas XVI a XVIII muestran las medidas en las mismas zonas realizadas el día posterior a los exámenes de selectividad

MEDIA_WLAN	VARIANZA_WLAN	MEDIA_JAZZ	VARIANZA_JAZZ
-48,000	0,000	-59,875	1,763
-43,000	0,000	-65,250	1,299
-46,000	0,000	-62,875	1,166
-47,875	1,452	-60,500	0,866
-44,625	1,653	-60,625	1,653
-45,250	0,969	-59,875	1,268
-44,500	1,936	-63,875	1,363
-51,000	0,000	-60,250	1,299
-53,000	0,000	-56,750	2,278

Tabla VIII. Medidas de potencia en zona este columna 1 de planta baja correspondiente al 9 de junio.

MEDIA_WLAN	VARIANZA_WLAN	MEDIA_JAZZ	VARIANZA_JAZZ
-78,000	0,000	-52,500	0,707
-63,000	0,000	-50,000	0,707
-57,500	0,866	-48,625	0,992
-43,500	7,794	-49,375	0,484
-50,875	9,198	-49,250	1,199
-58,000	0,000	-47,875	1,053
-55,500	1,936	-50,875	0,780
-69,500	15,500	-48,375	0,857
-74,500	13,555	-49,875	0,780

Tabla IX. Medidas de potencia en zona este columna 3 de planta baja correspondiente al 9 de junio.

MEDIA_WLAN	VARIANZA_WLAN	MEDIA_JAZZ	VARIANZA_JAZZ
-78,500	14,722	-71,250	1,090
-90,500	1,323	-73,500	1,000
-88,500	1,936	-69,625	2,118
-87,000	0,000	-68,250	1,714
-87,000	0,000	-71,625	1,727
-86,000	0,000	-70,625	1,576
-80,000	0,000	-69,000	1,414
-75,000	0,000	-61,375	1,218
-76,000	0,000	-61,000	1,871

Tabla X. Medidas de potencia en zona este pasillo baldosa 1 de planta baja correspondiente al 9 de junio.

MEDIA_WLAN	VARIANZA_WLAN	MEDIA_JAZZ	VARIANZA_JAZZ
-86,000	0,000	-70,625	1,653
-70,250	5,953	-69,500	1,000
-75,000	0,000	-70,125	0,927
-76,500	0,866	-68,500	1,000
-75,500	1,500	-68,750	1,090
-74,000	0,000	-70,500	1,000
-75,500	0,866	-68,500	1,500
-78,000	2,000	-69,250	3,382
-77,000	3,873	-69,875	

Tabla XI. Medidas de potencia en zona este pasillo baldosa 2 de planta baja correspondiente al 9 de junio.

MEDIA_WLAN	VARIANZA_WLAN	MEDIA_JAZZ	VARIANZA_JAZZ
-44,000	2,646	-37,875	2,976
-52,750	3,031	-30,000	0,000
-71,125	10,167	-22,500	5,809
-48,375	23,722	-18,000	0,000
-55,000	0,000	-51,000	0,000
-45,750	9,093	-24,750	15,155
-80,000	0,000	-82,000	0,000
-65,750	8,227	-82,000	0,000
-60,000	1,000	-64,500	17,500

Tabla XII. Medidas de potencia en zona este columna 1 de planta baja correspondiente al 9 de junio, durante el intermedio del examen.

MEDIA_WLAN	VARIANZA_WLAN	MEDIA_JAZZ	VARIANZA_JAZZ
-43,000	0,000	-55,000	0,000
-57,500	18,187	-50,750	4,763
-83,625	14,221	-55,875	8,268
-46,000	0,000	-34,000	0,000
-42,875	1,452	-53,625	3,389
-41,625	1,653	-57,500	1,323
-46,000	0,000	-54,000	0,000
-46,000	0,000	-48,000	0,000
-50,000	0,000	-40,000	0,000

Tabla XIII. Medidas de potencia en zona este columna 3 de planta baja correspondiente al 9 de junio, durante el intermedio del examen.

MEDIA_WLAN	VARIANZA_WLAN	MEDIA_JAZZ	VARIANZA_JAZZ
-52,000	0,000	-50,000	0,000
-28,375	8,929	-59,625	3,638
-22,750	1,299	-82,000	12,124
-60,000	0,000	-65,000	0,000
-18,000	0,000	-67,000	0,000
-50,000	0,000	-61,000	0,000
-52,625	8,992	-72,938	4,115
-45,250	7,276	-71,875	5,622
-46,125	7,607	-64,875	2,976

Tabla XIV. Medidas de potencia en zona este pasillo baldosa 1 de planta baja correspondiente al 9 de junio, durante el intermedio del examen.

MEDIA_WLAN	VARIANZA_WLAN	MEDIA_JAZZ	VARIANZA_JAZZ
-73,000	0,000	-62,000	0,000
-73,000	0,000	-62,000	0,000
-72,750	0,902	-59,688	0,982
-76,500	2,500	-60,000	0,000
-78,000	1,732	-59,500	0,866
-74,250	1,984	-61,875	10,252
-69,000	0,000	-89,000	0,000
-72,125	2,421	-60,250	0,969

Tabla XV. Medidas de potencia en zona este pasillo baldosa 2 de planta baja correspondiente al 9 de junio, durante el intermedio del examen.

MEDIA_WLAN	VARIANZA_WLAN	MEDIA_JAZZ	VARIANZA_JAZZ
-49,000	0,000	-50,875	0,330
-56,000	0,000	-48,500	0,866
-48,250	0,662	-47,375	0,484
-48,000	0,000	-50,000	0,707
-51,625	0,992	-51,750	1,199
-48,750	2,165	-45,750	0,662
-51,000	0,000	-47,250	0,662
-49,000	0,000	-51,750	0,434
-45,000	0,000	-51,250	1,199

Tabla XVI. Medidas de potencia en zona este columna 3 de planta baja correspondiente al 10 de junio.

MEDIA_WLAN	VARIANZA_WLAN	MEDIA_JAZZ	VARIANZA_JAZZ
-41,500	21,301	-63,125	0,780
-69,125	0,330	-62,000	0,866
-70,000	0,000	-57,625	1,409
-74,000	1,732	-61,875	0,927
-77,750	0,434	-61,125	0,927
-71,000	7,000	-59,875	0,599
-66,750	4,763	-60,875	1,452
-75,000	0,000	-58,875	1,833
-75,000	0,000	-58,750	0,829

Tabla XVII. Medidas de potencia en zona este pasillo baldosa 1 de planta baja correspondiente al 10 de junio.

MEDIA_WLAN	VARIANZA_WLAN	MEDIA_JAZZ	VARIANZA_JAZZ
-81,625	3,638	-65,000	0,500
-72,000	0,000	-67,375	0,696
-72,000	0,000	-70,000	0,500
-75,750	2,165	-67,000	0,866
-80,000	3,000	-67,250	1,090
-80,750	2,905	-72,375	0,992
-77,250	0,662	-73,125	1,166
-79,000	0,000	-65,500	1,658
-82,000	0,000	-71,000	0,707

Tabla XVIII. Medidas de potencia en zona este pasillo baldosa 2 de planta baja correspondiente al 10 de junio.

Haciendo el mismo análisis que el realizado en el pasillo N22 del edificio politécnico, tenemos las siguientes diferencias medias entre las sesiones de medidas:

Columna 1:

- Diferencias medias entre las medidas del día 9 de Junio durante el examen de selectividad y ese mismo día durante el intermedio del examen: 11.94 dB para el router WLAN y 25.83 dBm para el router JAZZ

Columna 3:

- Diferencias medias entre las medidas del día 9 de Junio durante el examen de selectividad y ese mismo día durante el intermedio del examen: 16.78 dB para el router WLAN y 5.92 dBm para el router JAZZ
- Diferencias medias entre las medidas del día 9 de Junio durante el examen de selectividad y el 10 de Junio: 12.69 dBm para el router WLAN y 2 dBm para el router JAZZ
- Diferencias medias entre las medidas del día 9 Junio en el descanso del examen de selectividad y las medidas del 10 de Junio: 8.19 dBm para el router WLAN y 7.36 dBm para el router JAZZ

Baldosa pasillo:

- Diferencias medias entre las medidas del día 9 de Junio durante el examen de selectividad y ese mismo día durante el intermedio del

examen: 10.48 dB para el router WLAN y 10.12 dBm para el router JAZZ

- Diferencias medias entre las medidas del día 9 de Junio durante el examen de selectividad y el 10 de Junio: 14.20 dBm para el router WLAN y 8.01 dBm para el router JAZZ
- Diferencias medias entre las medidas del día 9 Junio en el descanso del examen de selectividad y las medidas del 10 de Junio: 6.16 dBm para el router WLAN y 4.77 dBm para el router JAZZ

Las tablas XIX a XX muestran las medias de las potencias medidas en el hospital, sala de extracciones en las zonas marcadas como ZONA A y ZONA C en la figura 3.21. Estas medidas fueron tomadas en un día normal con gente en la sala de espera, el día 12 de junio. Las tablas XXI a XXII muestran los valores medidos en las mismas zonas en un día sin gente en la sala de espera, el día 11 de junio.

MEDIA_WLAN	VARIANZA_WLAN	MEDIA_JAZZ	VARIANZA_JAZZ
-36,000	0,000	-51,000	0,000
-35,125	0,330	-58,000	2,646
-36,000	0,000	-51,000	0,000
-38,375	4,357	-44,250	4,841
-35,000	0,000	-48,000	0,000
-33,500	0,866	-56,000	1,732
-42,625	3,638	-40,125	5,622
-35,000	0,000	-48,875	0,330
-33,000	0,000	-55,000	0,000

Tabla XIX. Medidas de potencia en zona A de sala de extracciones del hospital con gente.

MEDIA_WLAN	VARIANZA_WLAN	MEDIA_JAZZ	VARIANZA_JAZZ
-54,625	3,389	-40,875	1,452
-59,000	0,000	-47,000	8,000
-59,000	0,000	-55,000	0,000
-50,250	3,307	-40,125	5,622
-56,500	4,330	-46,250	4,763
-59,625	1,653	-47,000	5,292

Tabla XX. Medidas de potencia en zona C de sala de extracciones del hospital con gente.

MEDIA_WLAN	VARIANZA_WLAN	MEDIA_JAZZ	VARIANZA_JAZZ
-35,000	0,000	-49,000	0,000
-39,500	4,500	-52,500	3,500
-38,750	3,031	-52,250	2,165
-39,250	5,953	-51,375	0,992
-55,000	0,000	-54,000	0,000
-33,000	0,000	-46,000	0,000
-33,000	0,000	-46,000	0,000
-42,000	0,000	-46,000	0,000

Tabla XXI. Medidas de potencia en zona A de sala de extracciones del hospital sin gente.

MEDIA_WLAN	VARIANZA_WLAN	MEDIA_JAZZ	VARIANZA_JAZZ
-54,000	0,000	-47,000	0,000
-51,750	1,299	-41,000	3,464
-49,750	0,969	-37,750	0,969
-49,000	0,000	-37,000	0,000
-49,000	0,000	-42,625	7,262
-49,000	0,000	-49,625	6,284

Tabla XXII. Medidas de potencia en zona C de sala de extracciones del hospital sin gente.

A continuación se detallan las diferencias medias entre las diferentes medidas:

Zona A:

- Diferencias medias entre medidas realizadas con gente y medidas realizadas sin gente: 5.77 dBm router WLAN y 5.08 dBm router JAZZ

Zona C:

- Diferencias medias entre medidas realizadas con gente y medidas realizadas sin gente: 6.08 dBm router WLAN y 6.46 dBm router JAZZ

Las tablas XXIII y XXIV muestran los valores medidos en el hospital en la sala de oftalmología el día 7 de Julio. Las zonas de medida corresponde con las zonas marcadas como entrada y sala b en la figura 3.22. Las tablas XXV y XXVI corresponden con las medidas realizadas en esas mismas zonas en el día 11 de Julio

MEDIA_LIVEBOX	VARIANZA_LIVEBOX	MEDIA_JAZZ	VARIANZA_JAZZ
-66,000	0,000	-43,000	0,000
-67,000	0,000	-51,000	0,000
-71,000	0,000	-48,000	0,000
-66,000	0,000	-43,000	0,000
-66,500	1,323	-51,125	0,330
-71,000	0,000	-48,000	0,000
-68,000	0,000	-46,000	0,000
-63,000	0,000	-52,000	0,000
-66,000	0,000	-44,000	0,000

Tabla XXIII. Medidas de potencia en zona de entrada de oftalmología del hospital el 7 de Julio.

MEDIA_WLAN	VARIANZA_WLAN	MEDIA_JAZZ	VARIANZA_JAZZ
-52,000	0,000	-46,000	0,000
-58,125	2,315	-37,250	3,307
-51,000	0,000	-46,000	0,000
-51,000	0,000	-46,000	0,000
-52,500	3,500	-38,500	0,500
-56,000	0,000	-39,000	0,000
-47,875	1,452	-39,500	1,936
-43,000	0,000	-55,000	0,000
-39,000	0,000	-44,000	0,000
-39,000	0,000	-44,000	0,000
-56,000	0,000	-41,000	0,000

Tabla XXIV. Medidas de potencia en sala b de oftalmología del hospital el 7 de Julio.

MEDIA_LIVEBOX	VARIANZA_LIVEBOX	MEDIA_JAZZ	VARIANZA_JAZZ
-57,250	4,763	-48,250	6,495
-61,500	0,866	-48,250	2,165
-61,125	0,330	-39,875	2,315
-61,750	0,434	-50,750	3,897
-62,875	0,330	-43,375	3,638
-65,000	0,000	-39,000	0,000
-58,000	0,000	-49,500	0,500
-59,250	2,165	-48,500	0,866
-62,750	0,662	-47,375	0,992

Tabla XXV. Medidas de potencia en zona de entrada de oftalmología del hospital el 11 de julio.

MEDIA_WLAN	VARIANZA_WLAN	MEDIA_JAZZ	VARIANZA_JAZZ
-44,000	6,000	-40,500	0,500
-58,750	3,307	-35,625	1,653
-49,000	0,000	-35,000	0,000
-49,000	0,000	-39,000	0,000
-46,000	0,000	-46,000	0,000
-52,250	4,841	-38,125	2,421
-46,625	0,484	-37,500	1,936
-46,000	0,000	-38,000	0,000
-46,000	0,000	-47,000	0,000
-50,000	0,000	-39,000	0,000
-34,000	0,000	-37,000	0,000

Tabla XXVI. Medidas de potencia en sala b de oftalmología del hospital el 11 de julio .

A continuación se detallan las diferencias medias entre las diferentes medidas:

Entrada:

- Diferencias medias entre medidas realizadas el día 7 de Julio y medidas realizadas el día 11 de Julio: 6.11 dBm router WLAN y 5.67 dBm router JAZZ

Sala b:

- Diferencias medias entre medidas realizadas con gente y medidas realizadas sin gente: 6.10 dBm router WLAN y 5.86 dBm router JAZZ

De todas las medidas realizadas podemos sacar las siguientes conclusiones:

1. Las diferencias de las medidas realizadas en el hospital en diferentes días son similares. Esto puede ser debido a que estamos en un entorno similar en cuanto a mobiliario y condiciones ambientales
2. Las diferencias entre las medidas del hospital y las medidas realizadas en el pasillo N22 de la Escuela Politécnica son también similares. Siendo las medidas en este último entorno más reducidas.
3. Las diferencias en diferentes días y horas realizadas en la planta baja del edificio Este de la Escuela Politécnica son muy grandes. Esto puede ser debido a varias causas, la primera y más importante es que durante esas fechas, en ese lugar se celebraban los exámenes de selectividad y había gran movimiento de personas por esa zona. A resultas de lo cual se produjo una gran variabilidad de la potencia recibida en las huellas
4. Aunque en el hospital también se han realizado medidas en presencia humana, al ser realizadas en sala de esperas, el movimiento humano es menor y por tanto las diferencias son menores.
5. De lo comentado en los puntos anteriores se puede deducir que el factor humano es una variable muy importante en la precisión de los algoritmos de localización en interiores. Y que a mayor movilidad mayor error

3.4 REFERENCIAS

- [1] K. Pahlavan, L. Xinrong, and J. P. Makela, "Indoor geolocation science and technology," *IEEE Communications Magazine*, vol. 40, no. 2, pp. 112–118, Feb. 2002
- [2] Aleksandar Neskovic, Natascha Neskovic y George Paunovic, "Modern approaches in modelling of mobile radio systems propagation environment", *IEEE Communications Surveys*. Third Quarter 2000.
- [3] Naval Air Warfare Center, "Electronic Warfare and Radar Systems Engineering Handbook", Tech. report. TP 8347, Apr. 1997.
- [4] T. S. Rappaport, "Wireless Communications: Principles and Practice" Upper Saddle River, Prentice Hall, 1996
- [5] Perez, C., M. L. Mediavilla, and M. C. Diez, "Efectos del tráfico de personas sobre la atenuación en el canal de propagación en interiores," *URSI*, 1997.
- [6] J. Gomez, A. Tayebi, F. Saez de Adana, and O. Gutierrez, "Localization approach based on ray-tracing including the effect of human shadowing," *Progress In Electromagnetics Research Letters*, Vol. 15, 1-11, 2010.
- [7] S.Y. Seidel and T.S. Rappaport, "914 MHz path loss prediction models for indoor wireless communications in multifloored buildings", *IEEE Trans. on Antennas & Propagation*, vol. 40, no. 2, pp. 207-217, 1992
- [8] T.S. Rappaport, S.Y. Seidel, and K. R. Schaubach, "Site-specific propagation prediction for PCS system design" *Wireless Personal Communications*, M.J. Feuerstein and T.S. Rappaport, Eds. Norwell, MA: Kluwer, 1993, pp. 281-315.

- [9] AWE Communications, Germany, Software tool "WinProp for the planning of mobile communication networks" (incl. demo-version), www.awecomcommunications.com, Marzo 2003.
- [10] Kwok-Wai Cheung, Jonathan H.-M. Sau, and R.D. Murch, "A newwmpirical model for indoor propagation prediction" IEEE Trans. On Vehicular Technology, vol. 47, no. 3, pp. 996-1001, Aug. 1998.
- [11] Lähteenmäki, J. (1999). Indoor Propagation Models. Sección del Capítulo 4 "PropagationPrediction Models", en el Informe Final de la Acción COST 231, eds. E. Damosso, L.Correia, disponible en http://www.lx.it.pt/cost231/final_report.htm.
- [12] Akl, R., Tummala, D., Li, X. (2006). "Indoor Propagation Modeling at 2.4GHz for IEEE 802.11 Networks". The 6th International Multi-Conference on Wireless and Optical Communications. Banff, Canadá: IASTED.
- [13] W. Honcharenko, H.L. Bertoni, and J.L. Dailing, "Mechanism governing propagation between floors in buildings" IEEE Trans. On Antennas and propag., vol. 41, no. 6, pp. 787-790, 1993.
- [14] A. Falsafi, K. Pahlavan, and G. Yang, "Transmision techniques for radio LAN's - A comparative performance evaluation using ray tracing," IEEE J. Selected Areas in Communications, vol. 14, no. 3, pp. 477-491, Apr. 1996.
- [15] M.F. Catedra, J. Perez, M.F. Saez de Adana, O. Gutierrez, "Efficient Ray-tracing Techniques for Three-Dimensional Analyses of Propagation in Mobile Communications: Application to Picocell and Microcell Scenarios" IEEE Antennas & Propag. Mag, vol. 40, no. 2, pp. 15-28, Apr. 1998.

[16] M.F. Saez de Adana, O. Gutierrez, I. Gonzalez, J. Perez, M.F. Catedra, “Propagation Model Based on Ray Tracing for the Design of Personal Communication Systems in Indoor Environments” IEEE Trans. On Vehicular Technol., vol. 40, no. 6, pp. 2105-20112, Nov. 2000.

[17] J. Pérez and M. F. Cátedra, “Cell-Planing Personal Communications”, Reading, MA: Artech House, 1999.

[18] [http://msdn.microsoft.com/en-us/library/ms705969\(VS.85\).aspx](http://msdn.microsoft.com/en-us/library/ms705969(VS.85).aspx). The Native Wifi API that contains functions, structures, and enumerations that support wireless network.

4. Algoritmo de *clustering* para la selección de huellas.

4.1 INTRODUCCIÓN.

Como ya se ha comentado en el capítulo 2, la técnica del *fingerprinting* [1,2,3] utiliza la relación única entre niveles de potencias recibidas por los puntos de acceso de la red inalámbrica. Por tanto, la localización se determina calculando la distancia entre los niveles de potencia de las señales de radio frecuencia de la señal recibida y los niveles de radio frecuencia almacenados en la base de datos de calibración o radio-map. Para ello se minimiza la distancia calculada según la expresión (2.5)

Hay que tener en cuenta que el nivel de potencia en una huella, o punto del radio-map, no va ser fijo. En el capítulo 3 vimos como este nivel puede variar por diferentes razones, una de las más importantes es la presencia humana que no es estática. Una forma de determinar el nivel de potencia recibida por una huella, es tomar el valor medio de los valores de potencia medidos en dicha huella.

En este capítulo se va a estudiar la viabilidad de las técnicas de *clustering* a la hora de identificar huellas en el método del *fingerprinting* por grupos o *clusters* de valores de potencia. En [4] se utiliza la separación de huellas en *clusters* con el fin de estudiar las propiedades de la señal recibida. De esta forma si los clusters resultantes quedan perfectamente identificados con las correspondientes huellas, la cantidad de puntos de acceso y la

separación física entre huellas es óptima. En caso contrario sería necesario modificar algún parámetro, es decir incrementar el número de estaciones base o separar la distancia entre huellas.

En esta tesis se pretende estudiar diferentes técnicas de clusterización aplicadas a la identificación de huellas. Se aplicaran a grupos más amplios de huellas que las utilizadas en [4]. Para ello en primer lugar se han realizado una serie de medidas de la potencia recibida en una serie de puntos en el interior de un laboratorio de la Escuela Politécnica de la Universidad de Alcalá. El conjunto de valores de potencia posteriormente se ha agrupado en grupos o clusters en base a dos algoritmos, k-medidas y rek-medias. Se ha comparado el resultado obtenido con ambos algoritmos.

Posteriormente los resultados obtenidos pueden ser utilizados para mejorar el diseño de la red de localización, tal como se sugiere en [4]. También se pueden utilizar a la hora de localizar el móvil, estudiando la pertenencia a un *cluster* en lugar de calcular la distancia como en las técnicas actuales.

4.2. MÉTODOS DE CLUSTERIZACIÓN

El proceso de agrupamiento o clustering, consiste en la división de los datos en grupos de objetos similares, en buscar grupos en un conjunto de observaciones de forma tal que aquellos que pertenecen a un mismo grupo se parecen, mientras que aquellas que pertenecen a grupos distintos son disímiles, según algún criterio de distancia o de similitud.

4.2.1. TIPOS DE AGRUPACIÓN O *CLUSTERS*

Las formas de agrupación de objetos, tales como asignar clases predeterminadas a cada elemento ó agruparlos en forma significativa, son susceptibles de dividirse según el esquema de la figura 4.1 [5]:

- No exclusivas: Un mismo objeto puede pertenecer a varias categorías o grupos.
- Exclusivas: Cada objeto pertenece solamente a una categoría, clase o grupo.
 - a) Supervisadas: Las clases a las que pertenecen los objetos están predefinidas, y se conocen ejemplos de cada una, ó algunos de los objetos ya están agrupados y son utilizados por el algoritmo para aprender a clasificar a los demás.
 - b) No supervisadas: La agrupación se realiza en base a las características propias de los objetos, sin conocimiento previo sobre las clases a las que pertenecerán.
 - i. Jerárquicas: Los métodos jerárquicos consiguen la agrupación final mediante la separación (métodos divisivos) o la unión (métodos aglomerativos) de grupos. Así, estos métodos generan una estructura en forma de árbol en la que cada nivel representa una posible agrupación.
 - ii. Particionales (No jerárquicas): Los métodos particionales, o de optimización llegan a una única agrupación que optimiza un criterio predefinido o función objetivo, sin producir una serie de anidaciones de los grupos.

En la figura 4.1 mostramos una clasificación de las técnicas existentes para agrupar cualquier tipo de objeto:

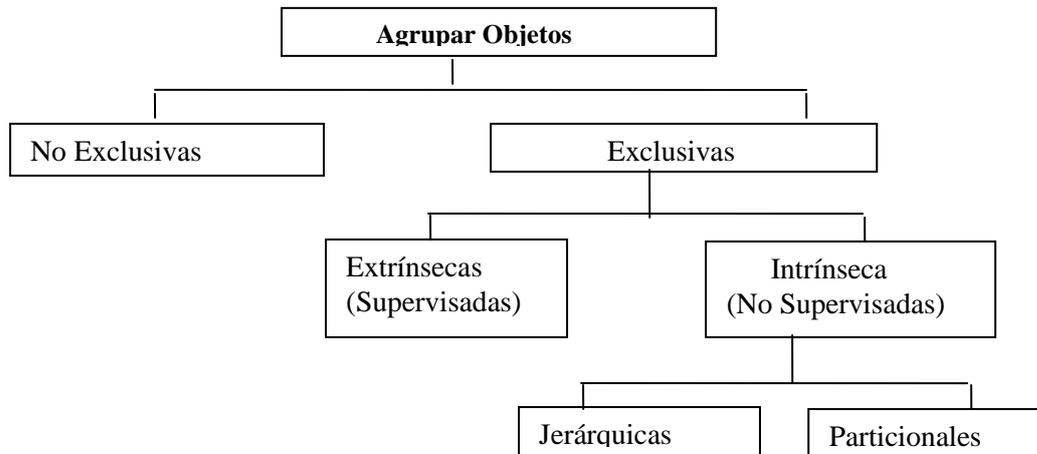


Figura 4.1: Formas de agrupar cualquier tipo de objeto

Debido a que para resolver el problema que se plantea en la presente tesis se utilizan técnicas exclusivas vamos a comentar a continuación este tipo de técnicas.

4.2.2. TÉCNICAS DE AGRUPAMIENTO EXCLUSIVAS

Una *técnica de agrupamiento* o clusterización exclusiva se puede definir como una técnica diseñada para realizar una agrupación asignando elementos a *grupos* de tal forma que cada grupo resulte más o menos homogéneo y distinto de los demás. Se suele emplear indistintamente los términos: grupo, agrupamiento, agrupación o el término inglés *cluster*, y también se suele hablar de *clustering* como sinónimo de técnica de agrupamiento. Intuitivamente, un conjunto de elementos se consideran un *grupo* si forman un conjunto homogéneo y diferenciado en algún sentido constatable.

El criterio de homogeneidad más simple está basado en la *distancia*, se espera que la distancia entre los elementos de un mismo agrupamiento sea *significativamente* menor que la distancia entre elementos de agrupamientos diferentes. En definitiva, una técnica de agrupamiento debe encontrar y caracterizar, en primer lugar, los agrupamientos, para posteriormente etiquetar los patrones en base a éstos.

Algunos algoritmos de agrupación requieren conocer previamente el número de agrupamientos a realizar (o un máximo) y en general, todos requieren unos parámetros específicos del algoritmo. Para hacer un agrupamiento se considera un conjunto P de patrones o criterios, cada uno de los cuales es característico de los agrupamientos, es decir:

Sea M el número de elementos X_1, X_2, \dots, X_M de P . Un *proceso de agrupamiento* consiste en buscar K grupos (subconjuntos de P) S_1, S_2, \dots, S_K tales que todo $X_i, i = 1, 2, \dots, M$ pertenece a uno y solo uno de estos grupos:

$$S_1 \cup S_2 \cup \dots \cup S_K = \{X_1, X_2, \dots, X_M\} = P \quad (4.1.)$$
$$S_i \cap S_j = \emptyset \quad \forall i \neq j$$

es decir, $\{ S_1, S_2, \dots, S_k \}$ constituye una partición de P .

Para la agrupación se dispone de dos posibles técnicas diferentes, supervisada y no supervisada. hay bastantes diferencias entre ambas técnicas, las dos más importantes son: el objetivo de la agrupación y la información de la que se dispone para llevarla a cabo.

El objetivo de la agrupación no supervisada consiste en intentar establecer la estructura de subconjuntos subyacente para un conjunto de datos bien para simplificarlos o para hacer un análisis, predicción o extracción de conocimiento de los mismos. Por otra parte, la agrupación supervisada pretende construir un sistema que asigne los objetos actuales del conjunto y objetos venideros a grupos o clases cuyo número viene prefijado de antemano.

En cuanto a la información que posee cada tipo de agrupación, los no supervisados únicamente disponen de los datos. Por el contrario los supervisados disponen del número de grupos y también de un ejemplo (ó muestras) para cada grupo.

Nos centraremos en los no supervisados en la que los objetos vienen descritos por valores numéricos. En particular presentaremos los métodos que producen particiones, esto es:

- Algoritmos de partición, método de dividir el conjunto de observaciones en k grupos o *clusters*, en el que k lo define inicialmente el usuario.
- Algoritmos jerárquicos, método que entrega una jerarquía de divisiones del conjunto de elementos en grupos.

La mayoría de los métodos de clusterización trabajan en un sólo nivel, en el que se optimiza un agrupamiento. Estos métodos asumen que el valor de k (la cantidad de grupos) está definida de antemano [5,6,7].

La estructura general de estos métodos se compone de los siguientes pasos [8]:

- 1) Seleccionar k puntos representantes (cada punto representa un grupo de la solución).
- 2) Asignar cada elemento al grupo del representante más cercano.
- 3) Actualizar los k puntos representantes de acuerdo a la composición de cada grupo.
- 4) Volver al punto 2)

Este ciclo se repite hasta que no sea posible mejorar el criterio de optimización. A continuación comentaremos las características de los principales algoritmos de este tipo.

4.2.3. ALGORITMO K-MEDIAS

Se trata de un algoritmo clasificado como Método de Particionado y Recolocación. El método de las k-medias [9,10], es hasta ahora el más utilizado en aplicaciones científicas e industriales. El nombre le viene porque representa cada uno de los clusters por la media (o media ponderada) de sus puntos, es decir, por su centroide. Este método únicamente se puede aplicar a atributos numéricos. Tiene como desventaja de que los puntos que no se encuentren cerca de ningún agrupamiento o cluster cercano afectar muy negativamente.

Sin embargo, la representación mediante centroides tiene la ventaja de que tiene un significado gráfico y estadístico inmediato. La suma de las discrepancias entre un punto y su centroide, expresado a través de la distancia apropiada, se usa como función objetivo. La función objetivo, suma de los cuadrados de los errores entre los puntos y sus centroides respectivos, es igual a la varianza total dentro del propio *cluster*.

Existen dos versiones del método de las k-medias. La primera se basa en dos pasos iterativos: primero reasigna todos los puntos a sus centroides más cercanos, y en segundo lugar recalcula los centroides de los nuevos grupos creados en el anterior. El proceso continúa hasta alcanzar un criterio de parada (por ejemplo que no se realizan nuevas reasignaciones). Esta versión se conoce como algoritmo de Forgy [11]. La segunda versión [12] reasigna los puntos basándose en un análisis más detallado de los efectos causados sobre la función objetivo al mover un punto de su *cluster* a otro nuevo. Si el traslado es positivo, se realiza, en caso contrario se queda como está.

A diferencia de otros algoritmos, como COBWEB y EM, k-medias necesita la previa especificación del número de *clusters* que se desean obtener. La implementación utilizada en el presente trabajo, de este algoritmo, es también la ofrecida por Weka. Al algoritmo de las k-medias hay que proporcionarle de antemano el número de *clusters* en los que se quiere segmentar la base de huellas.

Inicialmente toma k instancias (c_1, c_2, \dots, c_k) del conjunto de datos X, donde k es el número fijo de grupos requeridos; a estas instancias se les llama prototipos, y cada una de ellas se asume como el centroide inicial de cada uno de los k grupos.

El objetivo principal de k-medias es obtener k grupos disjuntos, donde los elementos de cada grupo, sean tan parecidos entre sí, y tan diferentes a los elementos de los otros grupos como sea posible. Para lograr esto, el algoritmo intenta minimizar el error dado por E:

$$E = \sum_{j=1}^k \sum_{x_i \in C_j} |x_i - c_j|^2 \quad (4.2.)$$

para $x_i \in X, \quad i = \{1, 2, \dots, n\}$

Donde C_j es el j-esimo grupo, c_j es su centroide y n es el número total de instancias x_i en el conjunto de datos.

A continuación, se examina cada instancia en el conjunto de datos y se asigna a un grupo dependiendo de la distancia mínima que presente a los centroides. La posición de los centroides se vuelve a calcular después de que se ha hecho el agrupamiento para todos los elementos, usando la expresión:

$$c_j = \sum_{x_i \in C_j} \frac{x_i}{|C_j|} \quad (4.3.)$$

Este proceso se repite hasta que el error E no cambie significativamente, o hasta que la pertenencia de los grupos se mantenga muy similar.

La elección de un k adecuado depende del problema, y repercute significativamente en los resultados que produce el algoritmo. Generalmente el usuario hace pruebas con diferentes valores, hasta que se obtienen los resultados deseados; aunque también existen maneras automáticas de encontrar un número óptimo de grupos, por ejemplo en el trabajo presentado por Davies y Bouldin [5], donde se promedian las medidas de similitud entre cada grupo y su grupo más cercano para obtener un conjunto de índices relacionados a los diferentes valores de k . En el problema propuesto en el presente trabajo el valor de k coincide con las huellas o localizaciones usadas en el método del fingerprinting.

En la figura 4.2. se muestra una representación gráfica de un paso del algoritmo

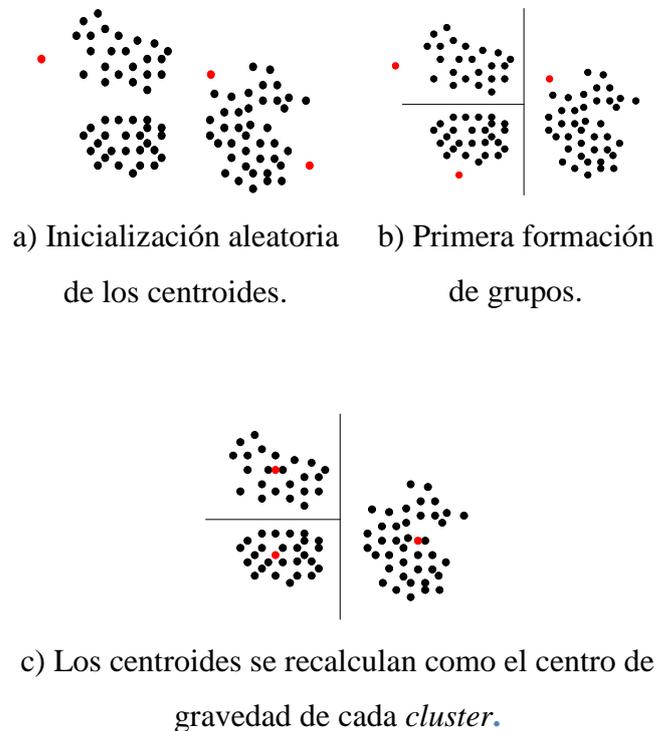


Figura 4.2 Representación gráfica del algoritmo de las k-medias

Este algoritmo es muy atractivo debido a su sencillez, pero presenta una fuerte dependencia con las condiciones de inicialización de los centroides. La veracidad de los resultados finales depende en gran medida de la generación aleatoria de las coordenadas iniciales de los centroides y del número inicial de estos que se seleccione. Con esta inicialización aleatoria, el algoritmo puede converger en ocasiones a soluciones erróneas. Por esta razón existen numerosas variantes del algoritmo k-medias [13, 14]. Una primera solución que se puede plantear es ejecutar varias veces el algoritmo k-medias y seleccionar la mejor solución. No obstante, en esta tesis se ha optado por implementar la variante rek-medias [15].

4.2.4. ALGORITMO REK-MEDIAS

Este algoritmo resuelve los problemas comentados para el algoritmo del k-medias. En un paso inicial se sobredimensionan los *clusters* envolviendo estos en rectángulos, para posteriormente ir mezclando estos últimos en *clusters* más grandes. El prefijo *re* del método viene de la palabra rectángulo. El algoritmo consta de cinco pasos que se detallan a continuación. Para una mejor comprensión del mismo vamos a considerar que los valores a los que se aplica el algoritmo tienen dos dimensiones que vamos a denominar X e Y. Sin embargo se puede aplicar a variables n dimensionales

En un primer paso se ejecuta el algoritmo k-medias con un valor de k igual a una cuarta parte de los datos totales. Este valor se toma como un compromiso entre velocidad de ejecución y número de *clusters* que se pueden asociar en el siguiente paso. Las pruebas realizadas en [15] demuestran que este valor da buenos resultados.

En un segundo paso se eliminan los *clusters* cuyo tamaño es igual o inferior a dos puntos. Esto permite que puntos aislados de los centroides, *outliers* según la terminología en inglés, se eliminen y no afectan al resultado final.

En un tercer paso se construye un rectángulo alrededor de los puntos de cada *cluster* de forma que uno de los vértices coincide con las coordenadas mínimas de X e Y y el otro con las coordenadas máximas. En la figura 4.3. a) se muestra una representación gráfica de la creación de este rectángulo

El siguiente paso consiste en asociar o mezclar *clusters*. El criterio que se toma para mezclar dos *clusters* es que la distancia entre los dos rectángulos que envuelven dichos *clusters* no superen una distancia determinada de antemano.

En la figura 4.3 b) se puede observar de forma gráfica este paso:

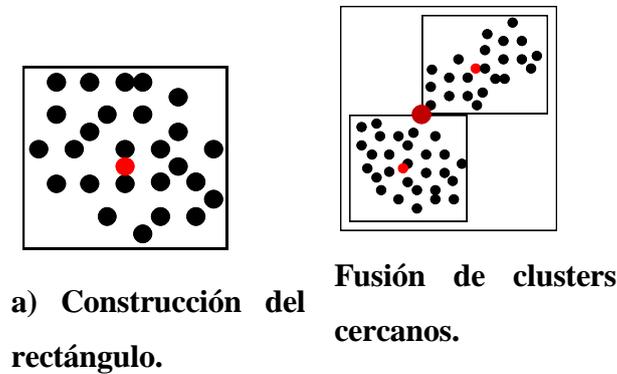


Figura 4.3. Representación gráfica del algoritmo rek-medias

Aunque la utilización de rectángulos parece indicar que el algoritmo sólo se puede aplicar a valores bidimensionales en realidad se puede aplicar a valores con cualquier número de dimensiones. Por ejemplo para puntos tridimensionales en lugar de un rectángulo se utilizaría un paralelogramo.

El último paso consiste en la validación. Mediante este paso se compara si los *clusters* están bien separados. Se utilizan métodos estadísticos y lo que se comprueba es si los valores pertenecientes a un *cluster* siguen una función densidad de probabilidad. Si suponemos que los datos siguen una función de densidad de probabilidad normal, el test estará basado en ver si los datos de un *cluster* siguen dicha función. El proceso se detalla en [15].

En la figura 4.4 . se muestra un diagrama de flujo del algoritmo:

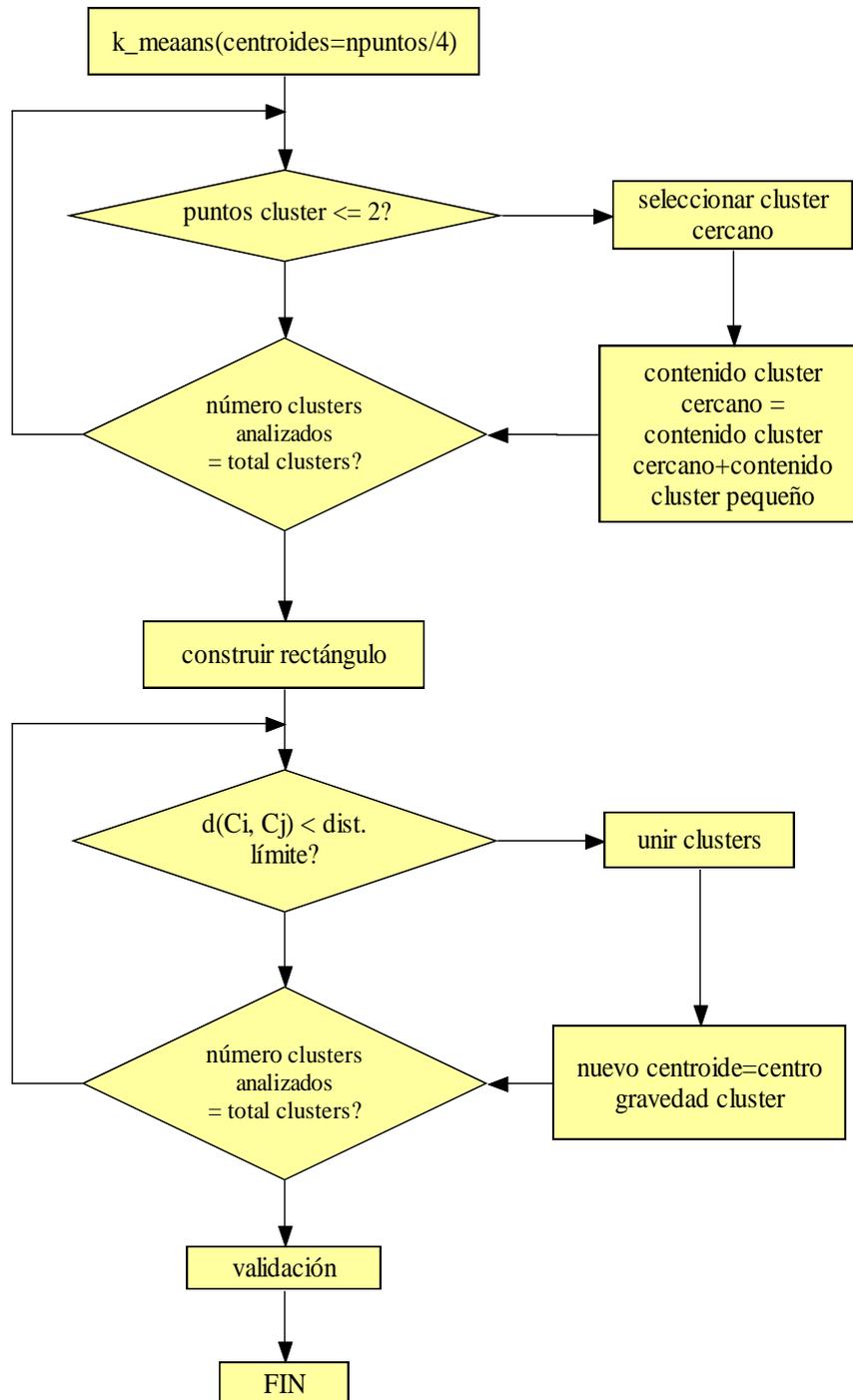


Figura 4.4. Diagrama de flujo del algoritmo rek-medias.

4.3. RESULTADOS EXPERIMENTALES

Con el fin de probar la validez de los algoritmos de clusterización detallados en el apartado anterior se realizaron una serie de medidas en el edificio politécnico de la Universidad de Alcalá. En el apartado 4.3.1. se detalla el proceso de medidas realizadas, mientras que en el apartado 4.3.2. se muestran los resultados tras haber aplicado, sobre el conjunto de medidas, los algoritmos del k-medias y el rek-medias.

4.3.1. CONFIGURACIÓN DE MEDIDAS

Las medidas se realizaron en un laboratorio de la Escuela Politécnica perteneciente a la Universidad de Alcalá. Se situaron dos transmisores en dicho laboratorio, cada uno de ellos transmitiendo un pulso de onda continua a 2.412GHz y 2.484GHz respectivamente. La elección de estas frecuencias se debe a que pertenecen a los canales 1 y 11 de un sistema WiFi. La potencia de salida fue de -10 dBm. En la figura 4.4. se puede ver una vista del generador de onda continua utilizado. Como antena transmisora se utilizaron antenas omnidireccionales, cuya frecuencia de operación es 2.4GHz con 2DBi de ganancia. En la figura 4.5. se puede observar una vista de una de esas antenas conectado al generador de señal utilizado como transmisor.



Figura 4.4. Generador de señal utilizado como transmisor



Figura 4.5. Conjunto transmisor, generador de señal y antena

En recepción, con el fin de medir el nivel de potencia recibido en diferentes posiciones dentro del laboratorio, se situó un monopolo vertical de 10Dbi de ganancia. Este se conectó a un analizador de espectros. Las medidas se realizaron en posiciones tales que hubiera condiciones de visión directa o LoS, para ello se tomaron los valores de potencia recibidos en 9 puntos. Estos están situados dentro del laboratorio formando un mallado de 3x3 puntos con una separación entre puntos o huellas de 30cm. En la figura 4.6. se puede observar la posición tanto de los transmisores como de los puntos de observación.

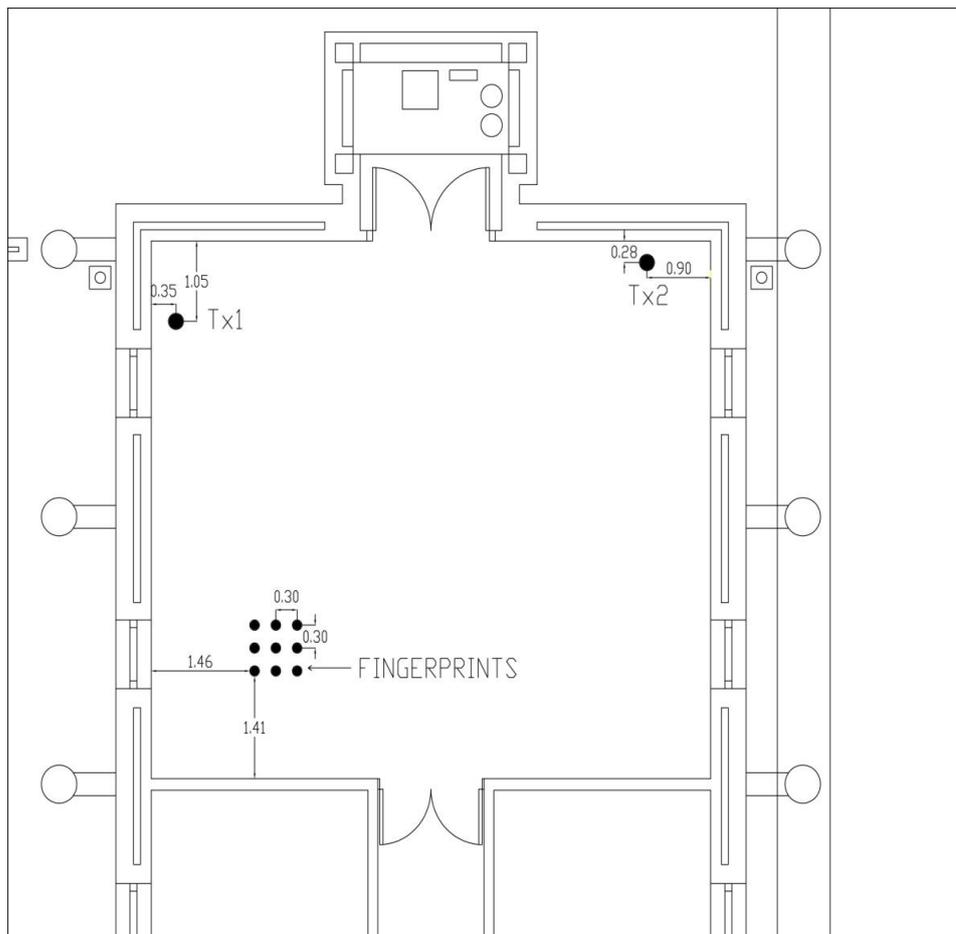


Figura 4.6. Zona de laboratorio, mallado y ubicación de trasmisores

En cada una de estas huellas o puntos de observación se recogieron 100 muestras de potencia, recibidas en cada uno de los canales por los dos transmisores, tomando un intervalo entre medidas de 1sg. Para las tomas de medidas se utilizó un software específico encargado de tomar las muestras y almacenarlas en diferentes ficheros. En la figura 4.7. se muestra una vista del equipo receptor. En la figura 4.8 se puede ver una muestra el escenario completo de medidas.

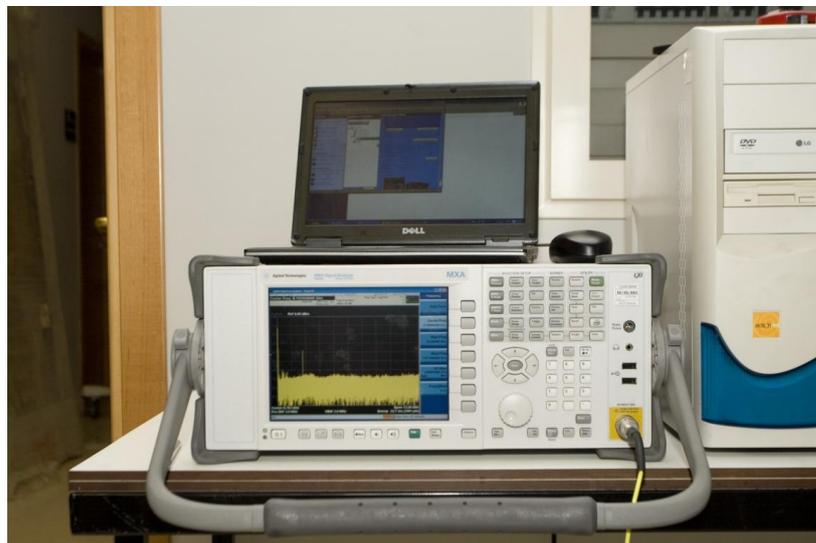


Figura 4.7. Equipo receptor

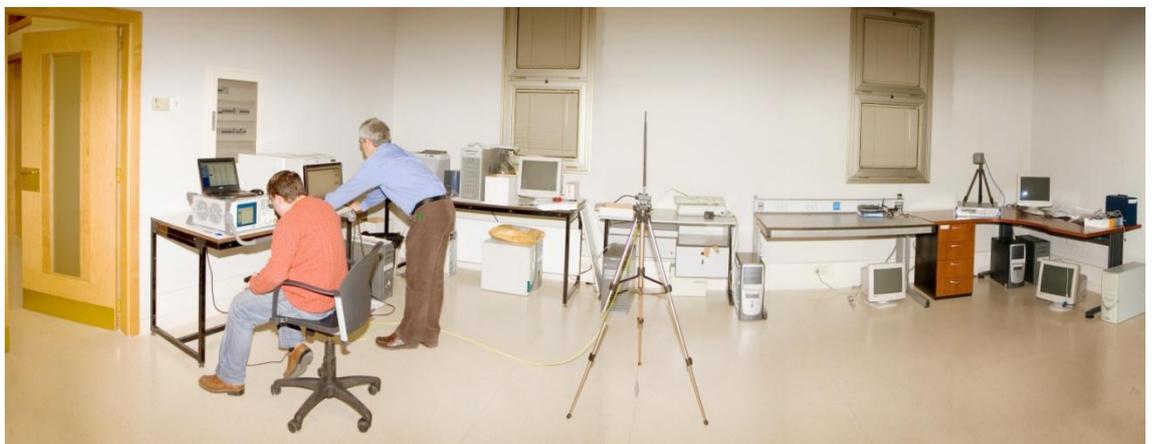


Figura 4.8. Escenario de medidas

4.3.2. RESULTADOS

Tras realizar las medidas de la potencia recibida en cada una de las huellas siguiendo el procedimiento descrito en el apartado 2, se procedió a aplicar los algoritmos de clusterización descritos en el apartado 4.1. Para una mejor visualización de los resultados, estos se muestran como valores de 2 dimensiones, siendo una de las dimensiones la potencia recibida del transmisor 1, y otra de las dimensiones la potencia recibida por el transmisor 2. En la figura 4.9 se muestra una representación gráfica del conjunto de potencias medidas en cada una de las posiciones o huellas en unidades dBm. Se presenta en diferentes colores los valores de potencia recibidos en cada huella.

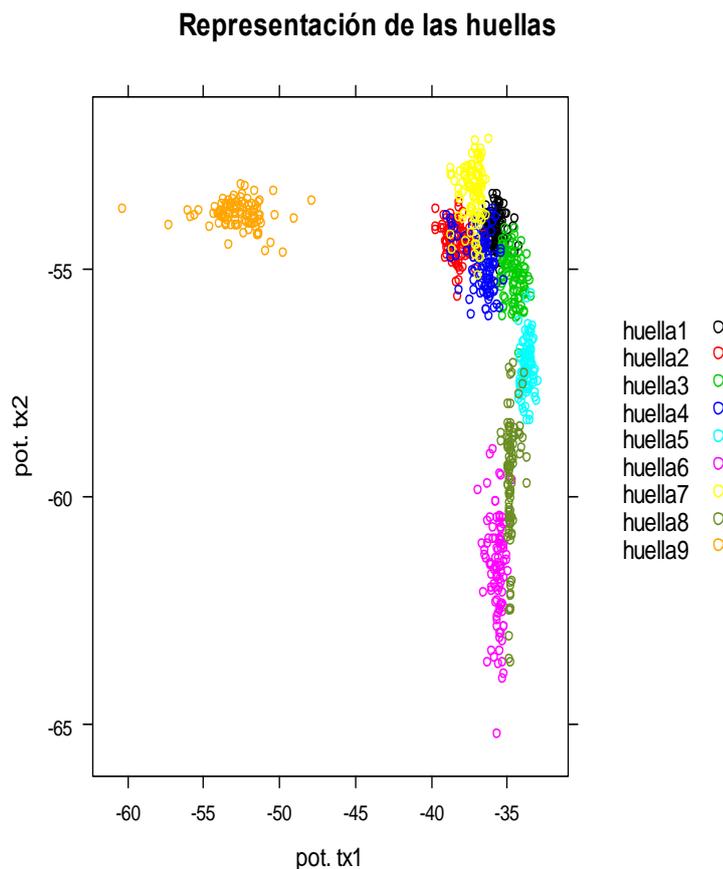


Figura 4.9. Disposición de las huellas tomadas en el laboratorio.

Sobre este conjunto de medidas se aplicó el método del k-medias. Como se comentó en el apartado anterior el resultado final depende de los valores iniciales de los *clusters*, los cuales se toman de forma aleatoria. En las figuras 4.10 y 4.11 se muestran dos resultados diferentes de dos simulaciones. En la figura 4.12 se muestran los clusters obtenidos al aplicar el algoritmo del rek-medias

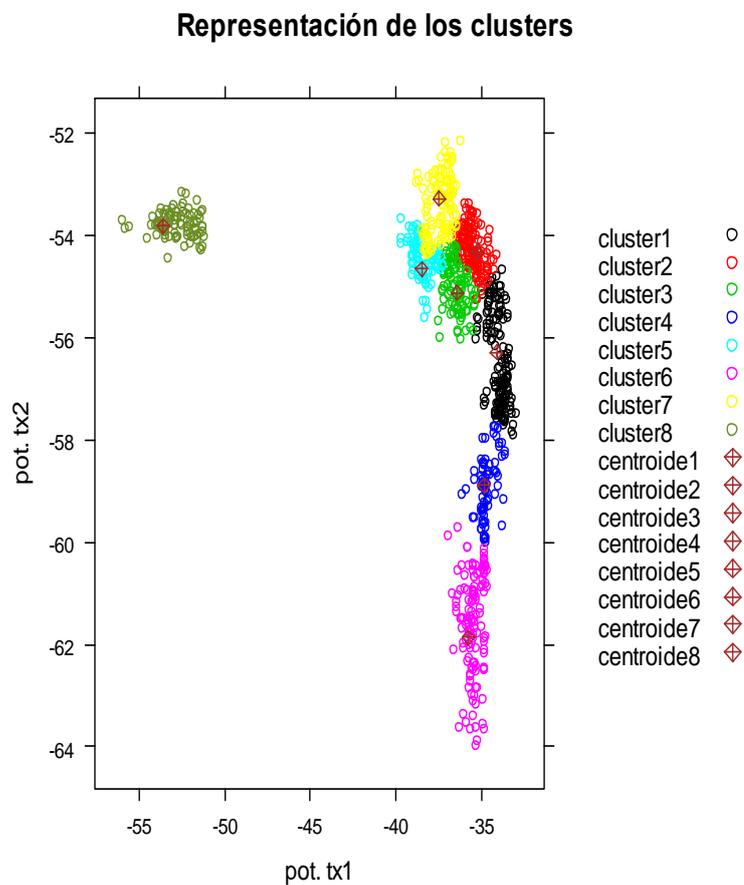


Figura 4.10. *Clusters* resultantes después de aplicar el algoritmo del k-medias con diferentes valores iniciales

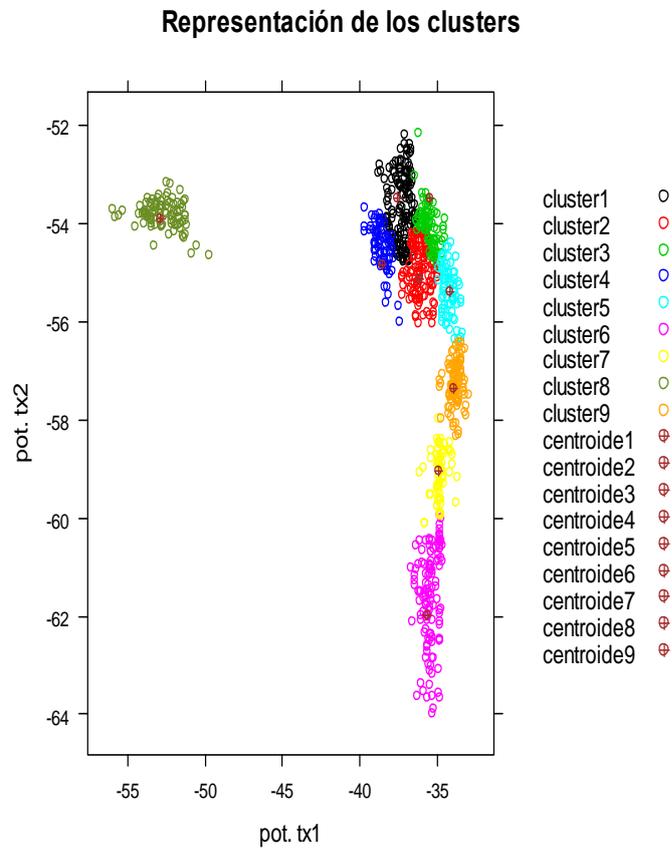


Figura 4.11. *Clusters* resultantes después de aplicar el algoritmo del k-medias con otros valores iniciales

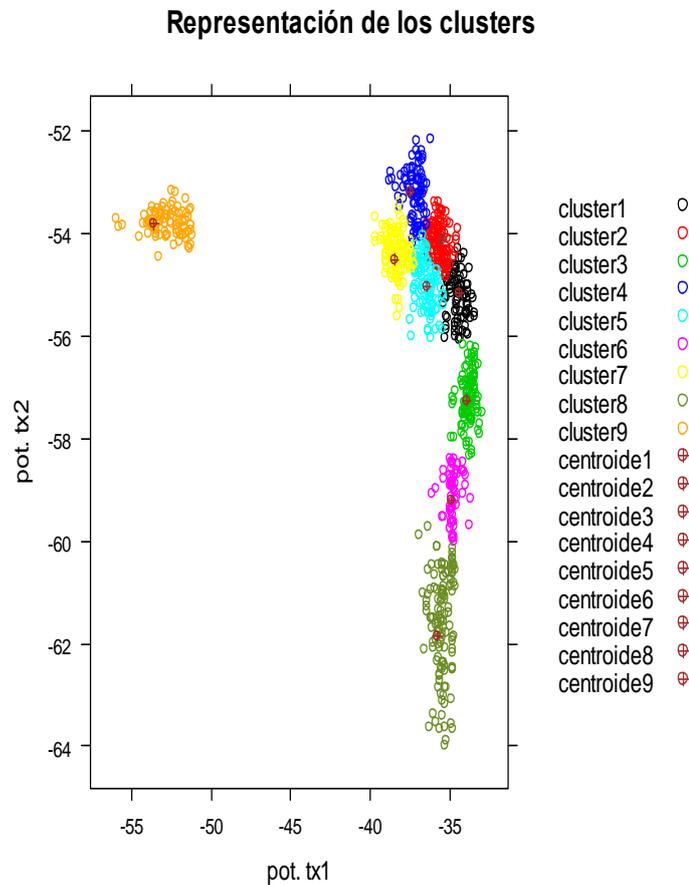


Figura 4.12. Clusters resultantes después de aplicar el algoritmo del rek-medias

En primer lugar se puede apreciar que el algoritmo k-medias da resultados diferentes dependiendo de los valores iniciales tomados aleatoriamente. En el primer caso, figura 4.10. el algoritmo detectó 8 clusters, mientras que en el segundo caso detectó 9. El algoritmo rek-medias siempre converge al mismo resultado final, mismo número de clusters y mismos centroides.

Hay que hacer notar que en este último algoritmo, los rectángulos se agrupan cuando la distancia entre ellos es inferior a una determinada distancia

máxima. Este valor es ajustable y, con el fin de llegar a resultados convergentes, es necesario ajustarlo previamente.

También se puede observar que los *clusters* detectados en el segundo caso coincide prácticamente con los valores medidos de las huellas. Sin embargo se pueden observar algunas características:

-Debido al carácter aleatorio del valor de la potencia recibida algunos valores medidos en una huella se mezclan con los valores medidos en otra huella. En otras palabras, los valores de potencia medidos en diferentes huellas son similares. Cuando se realiza la clasificación, debido a las características del algoritmo, los valores no se mezclan, creando zonas diferenciadas entre *clusters*.

-Para algunas huellas existen valores de potencia recibida alejados de los valores medios medido *outliers*. Al aplicar el algoritmo de clusterización estos valores son eliminados y no se tienen en cuenta a la hora de clasificar la huella.

Como resulta de esto, se puede observar que los *cluster* calculados se dividen en zonas bien diferenciadas. Estas zonas pueden ser aplicadas posteriormente al algoritmo de localización. Para ello en lugar de aplicar el criterio de la distancia mínima se puede aplicar el procedimiento de pertenencia a un *cluster*. Es decir, tras medir la potencia recibida en el móvil de cada una de las estaciones base se determina a que *cluster* pertenece. Un posible algoritmo de localización podría ser el siguiente:

- En primer lugar se toman los valores de potencia recibidos de cada AP en las diferentes huellas y se dividen en *clusters* utilizando el algoritmo descrito anteriormente

- Dividir los *clusters* en zonas geométricas. Por ejemplo en el caso de 2 dimensiones (2 estaciones base) zonas pueden ser rectángulo, rombos, o en general figuras en dos dimensiones.
- A la hora de asignar un móvil a una huella se puede determinar, a partir de los valores de potencia recibidos de los correspondientes transmisores, dentro de que zona geométrica se encuentra.
- La posición del móvil se asigna a la huella correspondiente al *cluster* en cuestión

Los resultados mostrados en las figuras 4.10 a 4.12 muestran los resultados de las 9 huellas. Para ver con algo de más detalle los resultados obtenidos, a continuación se presentan diferentes grupos de 4 huellas, correspondientes a diferentes cuadrantes del mallado original. En primer lugar, las figuras 4.13 y 4.14 se muestran las huellas 1, 2, 4 y 5 y los *clusters* resultantes; las figuras 4.15 y 4.16 corresponden con las huellas 2, 3, 5 y 6; las figuras 4.17 y 4.18 a las huellas 4, 5, 7 y 8 y las figuras 4.19 y 4.20 a las huellas 5, 6, 8 y 9.

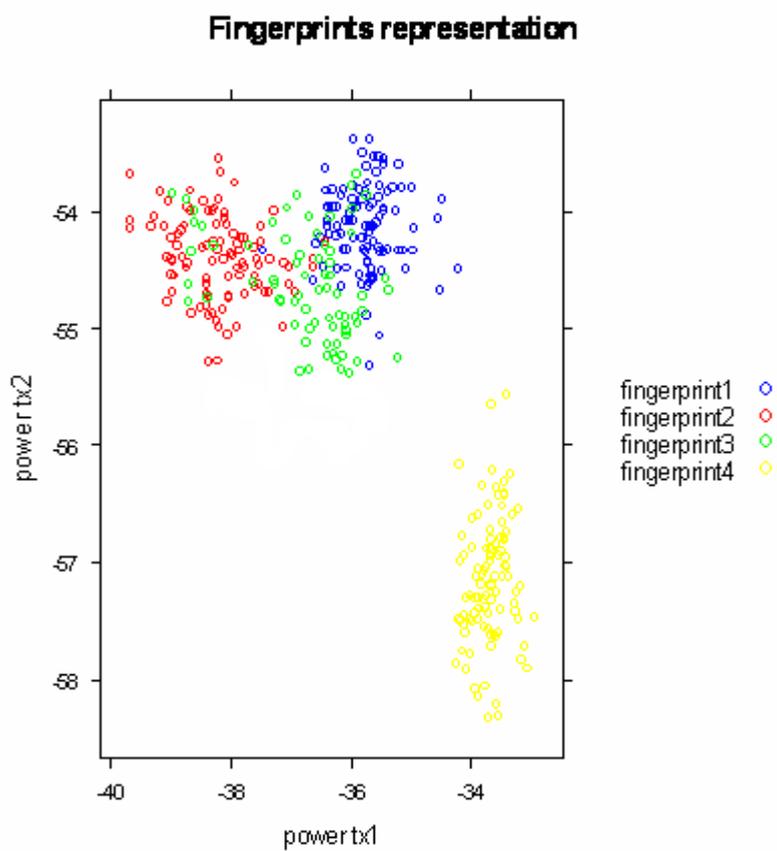


Figura 4.13. Valores de las huellas 1, 2, 4 y 5

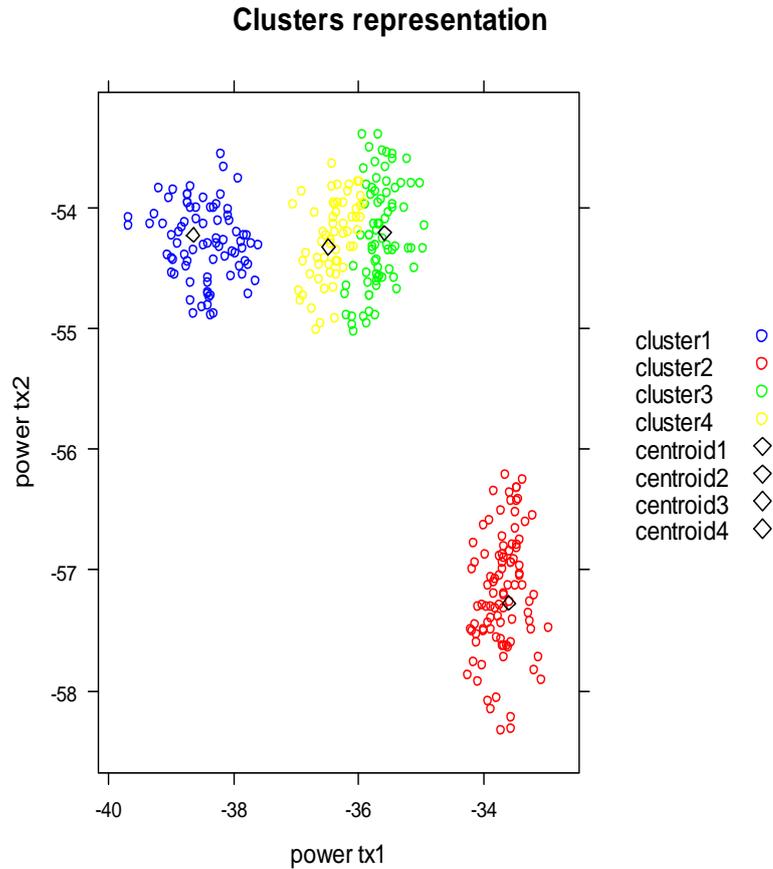


Figura 4.14. Valores de *cluster* correspondientes a las huellas 1, 2, 4 y 5

De la figura 4.13. se puede observar que, mientras que los valores de 3 de las huellas aparecen claramente definidas, las marcadas en rojo azul y amarillo. Sin embargo los valores del huella marcada en verde aparecen mezclados con los dos anteriores. Como resultado a la hora de aplicar el algoritmo de clusterización, como se puede ver en la figura 4.14, dos de los *clusters* los separa bien, mientras que los otros dos aparecen ligeramente mezclados.

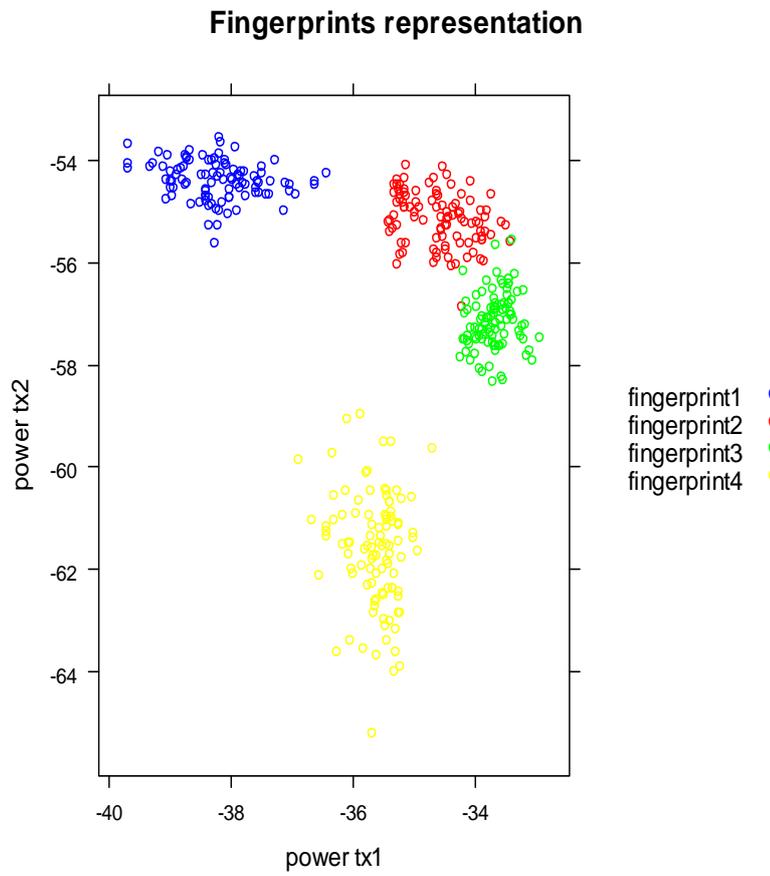


Figura 4.15. Valores de las huellas 2, 3, 5 y 6

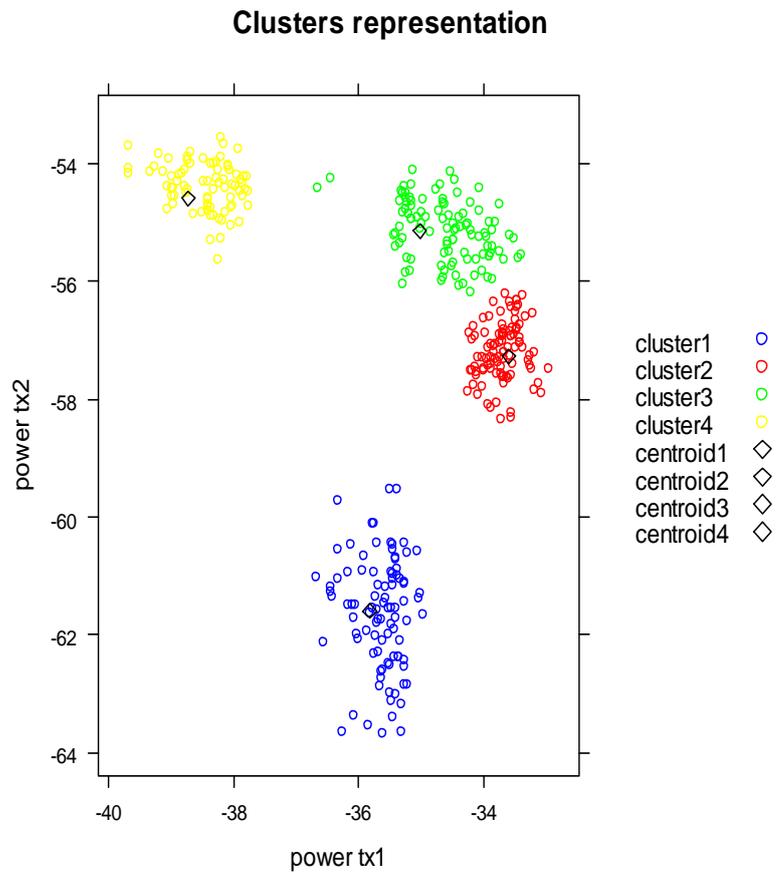


Figura 4.16. Valores de cluster correspondientes a las huellas 2, 3, 5 y 6

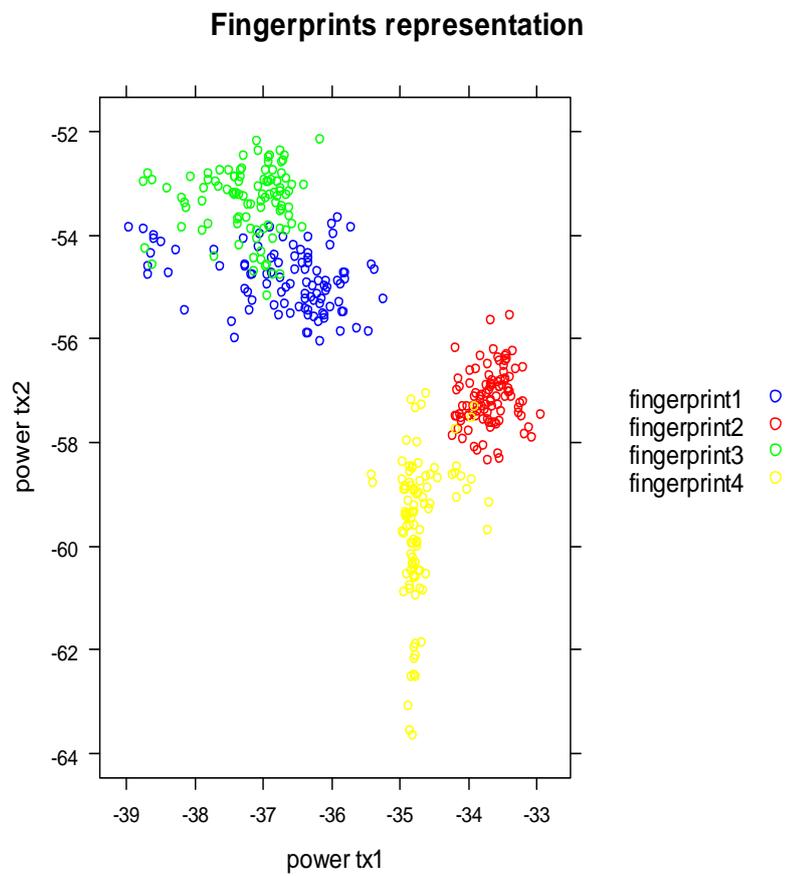


Figura 4.17. Valores de las huellas 4, 5, 7 y 8

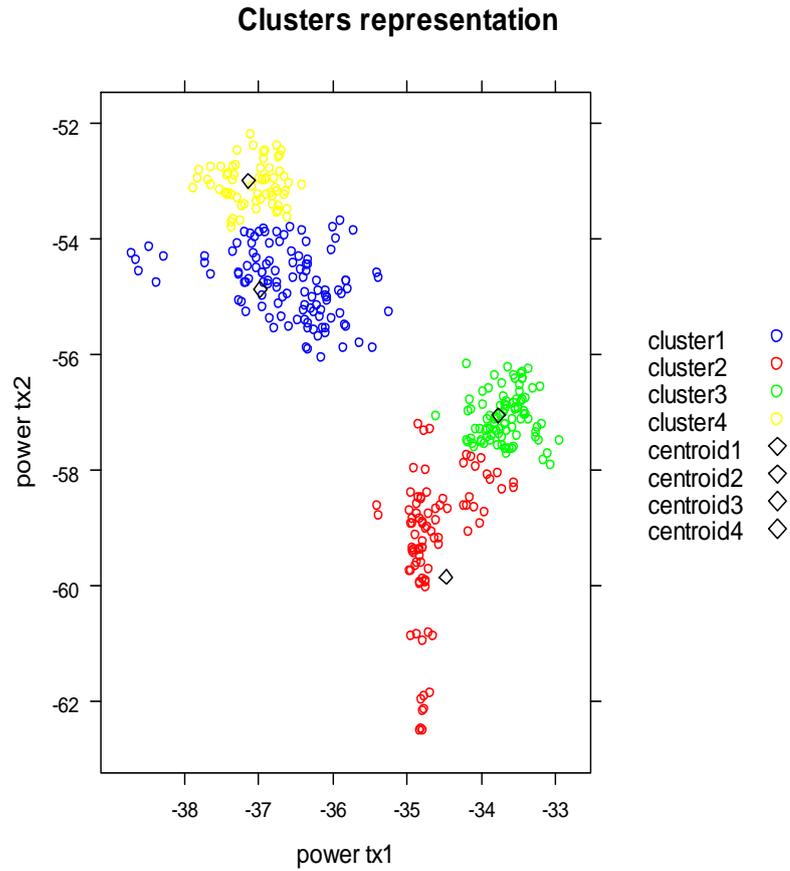


Figura 4.18. clusters correspondientes a las huellas 4, 5, 7 y 8

Para el grupo de huellas correspondientes a la figura 4.15. al no mezclarse los valores de las medidas los *clusters* aparecen bien definidos como puede verse en la figura 4.16. Lo mismo puede decirse respecto de las huellas y los *clusters* de las figuras 4.17 y 4.18.

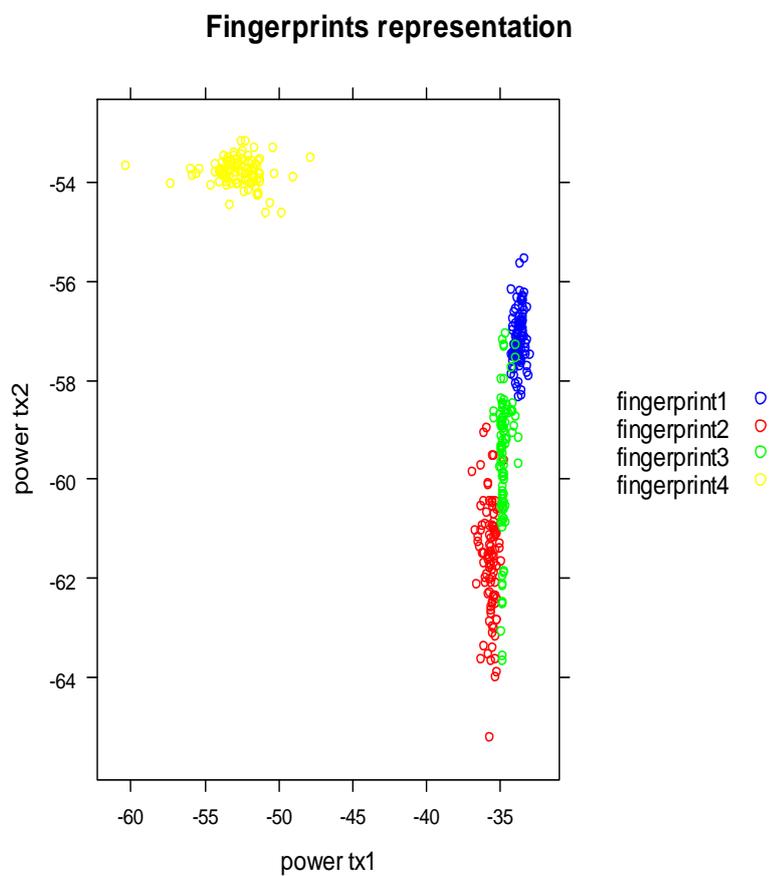


Figura 4.19. Valores de las huellas 5, 6, 8 y 9

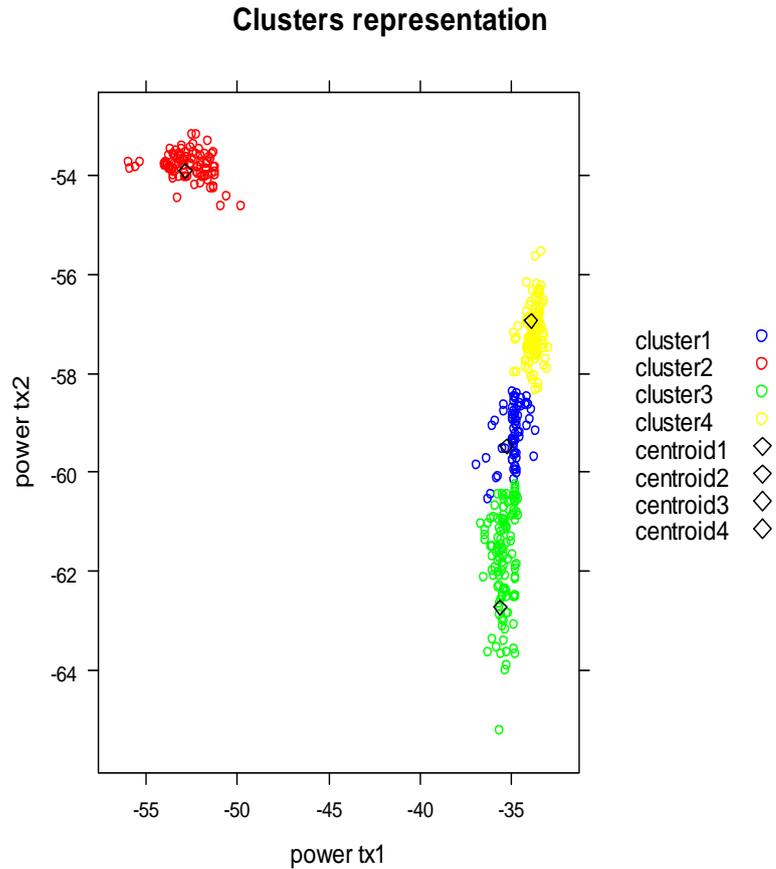


Figura 4.20. Cluster correspondientes a las huellas 5, 6, 8 y 9

En la figura 4.19. los valores de las medidas en las huellas representadas por los colores, azul, verde y rojo se entremezclan ligeramente. Por ejemplo, se puede observar que algunas potencias relativas al *fingerprint2* (en color rojo) caen en la zona del *fingerprint3* (en color verde), y viceversa. Al aplicar el algoritmo del *fingerprinting* estas tres huellas se definen claramente. Marcándose zonas de potencia correspondientes a cada cluster. Respecto a la huella de la figura 4.19 representada en amarillo, al estar sus valores bastante alejados del resto, no existe ningún problema al asignarla a su correspondiente *Cluster*

Capítulo 4: Algoritmo de *clustering* para la selección de huellas.

Una vez demostrada la validez del método de clusterización, a la hora de identificar grupos de valores de potencia medidos en cada huella, se procedió a realizar otro experimento. En este se aumentó la zona de medida, siempre en el mismo laboratorio de la Escuela Politécnica de la Universidad de Alcalá. En este experimento se utilizaron 25 posiciones o huellas formando una rejilla de 5x5 puntos. La posición de las huellas, así como la de los dos transmisores se muestra en la figura 4.21.

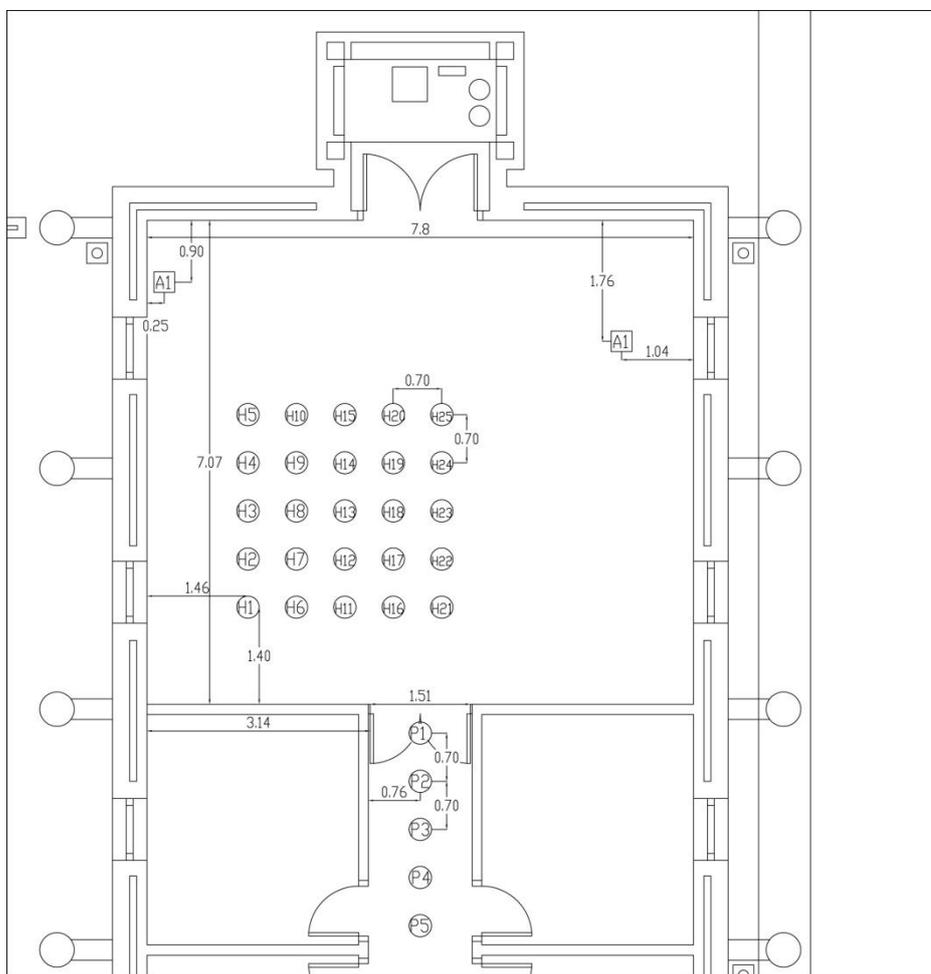


Figura 4.21. Zona de laboratorio, mallado y ubicación de transmisores

Se tomaron medidas en dos condiciones diferentes. En una primera sesión, se tomaron valores de potencia recibidos en cada una de las huellas.

Cien valores por huella con un intervalo temporal de 1sg entre medidas. En una segunda sesión, las medidas se tomaron con presencia humana dentro del laboratorio realizando sus labores habituales.

En la figuras 4.22. y 4.23. se muestran los valores de potencia medidos en las huellas y los *clusters* resultantes respectivamente para el caso de la segunda sesión. Las figuras 4.24 y 4.25. muestran las medidas tomadas en la segunda sesión y los *clusters* calculados a partir de estas respectivamente.

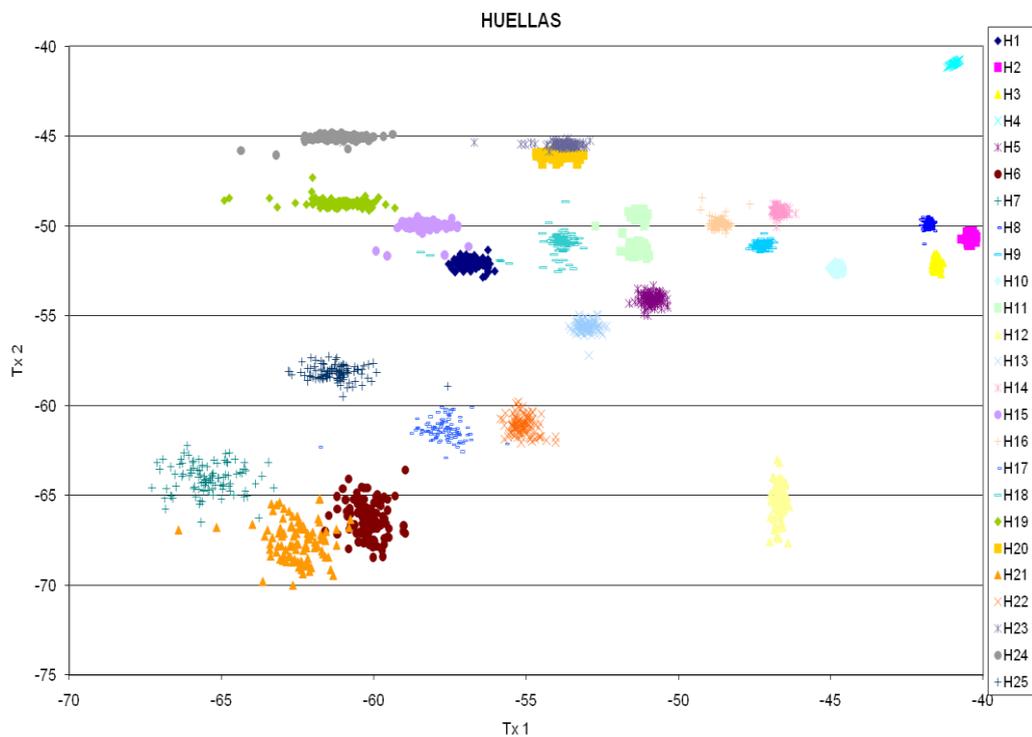


Figura 4.22. Representación de los valores de potencia medidos en las 25 huellas. Medidas tomadas sin presencia humana

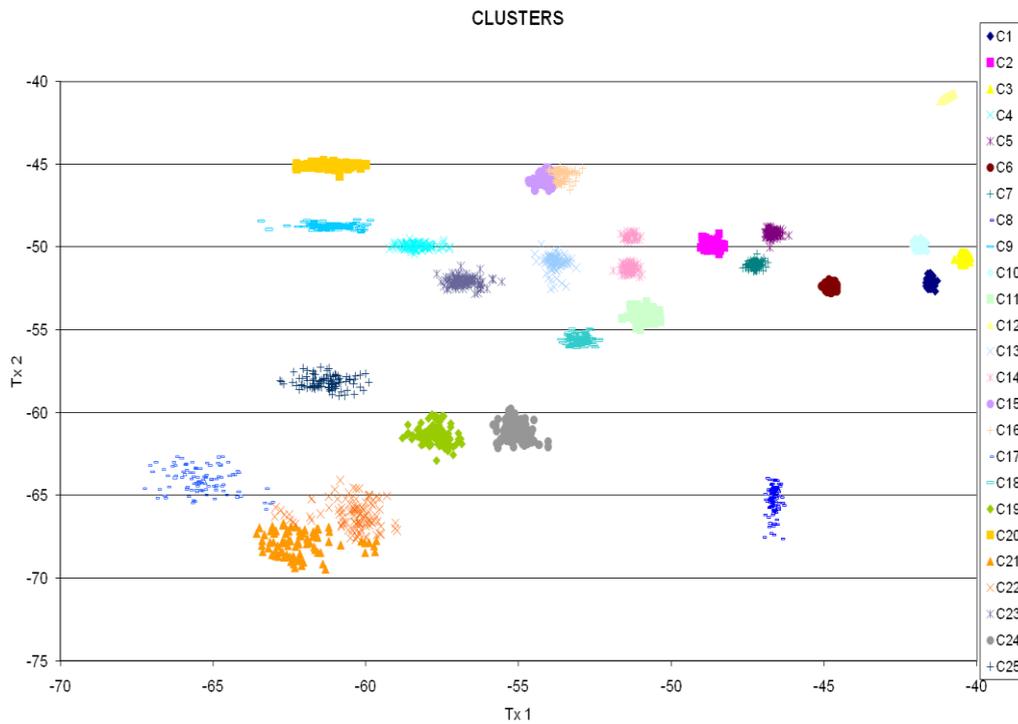


Figura 4.23. Clusters resultantes de las huellas de la figura 4.22

Se puede observar en la figura 4.22. que los valores de potencia medidos en cada huella están, en la mayoría de puntos, muy concentrados en torno a la media. Como se vio en el capítulo anterior, cuando no hay presencia humana la dispersión de las medidas es baja. Sin embargo en algunas huella, como por ejemplo las 21, 6, 7 y 25, los valores están más dispersos. Esto se debe a que esas huellas están más alejadas de los transmisores y las medidas se tornan más aleatorias. Aun con todo el algoritmo es capaz de calcular perfectamente los 25 *clusters* correspondientes a las 25 huellas.

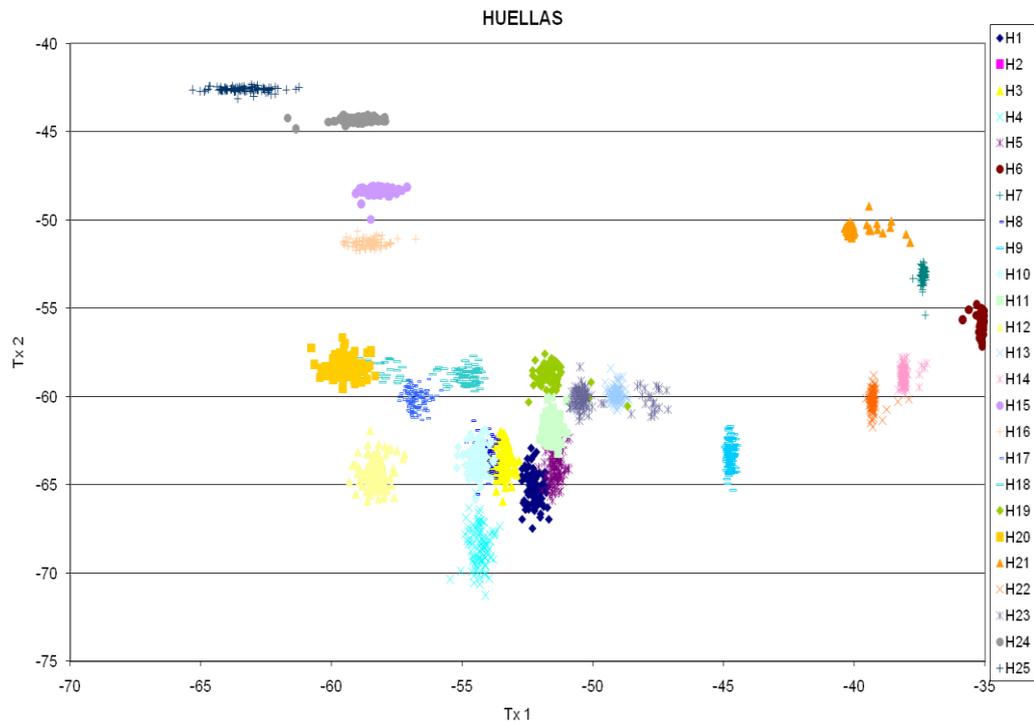


Figura 4.24. Representación de los valores de potencia medidos en las 25 huellas. Medidas tomadas con presencia humana

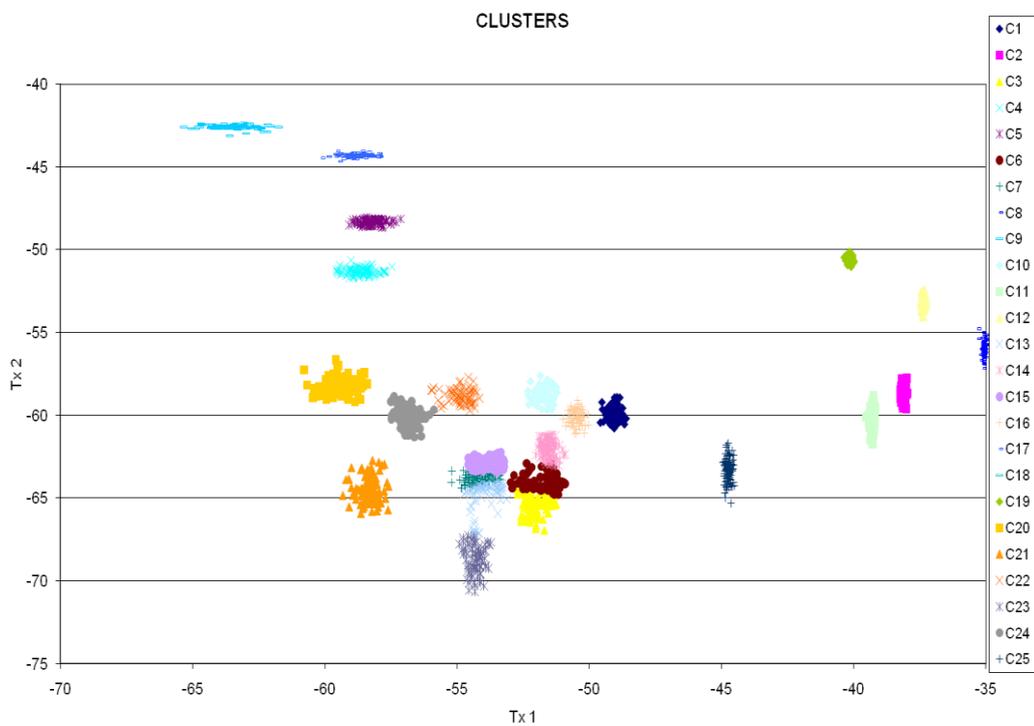


Figura 4.25. Clusters resultantes de las huellas de la figura 4.22

En el caso de las medidas con presencia humana se puede apreciar que existe más dispersión de las medidas. Esto es la causa de algunos errores a la hora de calcular los *clusters*. Por ejemplo los valores de potencia de las huellas 10, 8, 3, 1 y 5 corresponderían con los *clusters* 15, 7, 13, 3 y 6. Si nos fijamos en la forma de dichos *clusters* vemos que no coinciden con la forma de las huellas, lo que conlleva una identificación incorrecta. Una posible solución para evitar estos errores es aumentar el número de estaciones base.

Podemos extraer como conclusión parcial de estas dos sesiones de medidas que en escenarios con presencia humana el número necesario de estaciones base es mayor, que en aquellos escenarios con baja o nula presencia humana.

4.3.3. CONCLUSIONES

En este capítulo se ha visto como aplicar el algoritmo de clusterización a la hora de identificar huellas en la técnica de localización del *fingerprinting*. En los resultados experimentales obtenidos que el rek-means es más estable, ya que converge a los mismos resultados, independientemente de los valores iniciales. Esto es cierto siempre que se ajuste el parámetro de distancia máxima entre rectángulos

Los resultados obtenidos aquí pueden ser aplicados al algoritmo de localización. Para ello en lugar de aplicar criterios de distancias se pueden aplicar criterios de pertenencia a un determinado *cluster*.

4.4 REFERENCIAS

- [1] S. Saha, K. Chaudhuri, D. Sanghi, and P. Bhagwat, "Location determination of a mobile device using iee 802.11b access point signals," in Proc. IEEE Wireless Communications and Networking Conference (WCNC'03), New Orleans, LA, Mar. 2003, pp. 1987-1992.
- [2] P. Bahl and V. N. Padmanabhan, "RADAR: an in-building RF-based user location and tracking system," in Proc. IEEE Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM'00), Tel Aviv, Israel, Mar. 2000, pp. 775-784.
- [3] J. Hightower and G. Borriello, "Location systems for ubiquitous computing," IEEE Computer, vol. 34, no. 8, pp. 57-66, Aug. 2001.
- [4] K. Kaemarungsi, P. Krishnamurthy, "Properties of Indoor Received Signal Strength for WLAN Location Fingerprinting", Proc. First Annual International Conf. on Mobile and Ubiquitous Systems: Networking and Services, 2004. pp. 14-23, Aug. 2004
- [5] Qin He, 1996. "A review of clustering algorithms as applied in IR", UIUC LIS- 1996/ University of Illinois at Urbana- Champaign.
- [6] Witten, I. H. & Frank, E. Data Mining: Practical Machine Learning Tools and Techniques, second edition, Elsevier: San Francisco, 2005.
- [7] Christopher M. Bishop. Pattern Recognition and Machine Learning. Springer, 2006

- [8] Han, J, Kamber, M. y Tung, A. K. H. 2001. “Spatial clustering methods in data mining: A survey. Geographic Data Mining and Knowledge Discovery”, H. Miller and J. Han, editors, Taylor and Francis.
- [9] Hartigan, J., Clustering Algorithms, John Wiley & Sons, New York, 1975.
- [10] Hartigan, J. y Wong, M., “Algorithm AS139: A k-means clustering algorithm”, Applied Statistics, Vol. 28, 1979, pp. 100-108.
- [11] Forgy, E., “Cluster analysis of multivariate data: Efficiency versus interpretability of classification”, Biometrics, Vol. 21, 1965, pp. 768-780.
- [12] Duda, R. y Hart, P., Pattern Classification and Scene Analysis, Wiley & Sons, 1973.
- [13] D. Arthur, S. Vassilvitskii, “k-means++: the Advantages of Careful Seeding”.
- [14] D. Pelleg, A. Moore, “Accelerating Exact k-means Algorithms with Geometric Reasoning”.
- [15] D. D. Bloisi, L. Iocchi, “Rek-means, A k-means Based Clustering Algorithm”.

5. CONCLUSIONES Y FUTURAS LINEAS DE TRABAJO.

5.1. CONCLUSIONES

En esta tesis se han realizado una serie de medidas encaminadas a estudiar los diversos factores que pueden incidir en la precisión de los algoritmos de localización en interiores basados en la técnica del *fingerprinting*. Además se ha probado la técnica de clusterización con el fin de identificar las diferentes huellas del método del *fingerprinting* con los posibles niveles de potencia medidos en las posiciones de las mencionadas huellas.

Se han realizado diferentes sesiones de medidas en diferentes escenarios en los cuales se han variado las condiciones ambientales. Se han realizado medidas en la Escuela Politécnica de la Universidad de Alcalá y en el hospital Universitario Príncipe de Asturias. Las conclusiones a las que se ha llegado es que el parámetro que más influencia tiene sobre los valores medidos es la movilidad de las personas dentro del entorno de medidas.

También se ha probado que la técnica del rek-medias para agrupar valores de potencia recibidos con localizaciones es un método válido. Sin embargo si no se dispone de suficientes puntos de acceso esta técnica puede dar resultados erróneos. Para solventar este problema se puede aumentar el número de puntos de acceso o incrementar la separación entre huellas. El

problema de esta última solución es la pérdida de precisión ya que esta depende de la granularidad del mallado del *radiomap*.

5.2. FUTURAS LINEAS DE TRABAJO

Una futura línea de trabajo que se abre es ampliar las sesiones de medidas a más escenarios y más días de medidas en cada escenario. Además incrementar el tiempo de medida en cada escenario. Por ejemplo en un mismo sitio realizar medidas en día entero. Con este fin se dispondría de más datos para realizar estudios estadísticos más amplios que los presentados en esta tesis.

Otro tipo de estudio que se puede hacer es ver qué ocurre cuando se varía el número de puntos de acceso.

En esta tesis se ha diseñado un pequeño sistema de localización con el que se ha hecho las medidas. Se puede utilizar el mismo para aplicar la técnica del *fingerprinting* y estudiar la variación del error de localización en las diferentes situaciones.

La aplicación del método de clusterización a la clasificación de huellas, también abre otra importante vía de investigación. Para ello se podría diseñar un algoritmo basado en determinar, a partir de la potencia recibida en el móvil de los diferentes puntos de acceso, a que *cluster* pertenece el conjunto de esas potencias.