



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

— **TELECOM** ESCUELA
TÉCNICA **VLC** SUPERIOR
DE INGENIERÍA DE
TELECOMUNICACIÓN

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Escuela Técnica Superior de Ingeniería de
Telecomunicación

Modelado del canal radio a partir de técnicas de
clusterización

Trabajo Fin de Grado

Grado en Ingeniería de Tecnologías y Servicios de
Telecomunicación

AUTOR/A: Abad Abad, José Luis

Tutor/a: Rubio Arjona, Lorenzo

CURSO ACADÉMICO: 2022/2023



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

— **TELECOM** ESCUELA
TÉCNICA **VLC** SUPERIOR
DE INGENIERÍA DE
TELECOMUNICACIÓN

PORTADA GENERADA AUTOMÁTICAMENTE EN EBRON

Escuela Técnica Superior de Ingeniería de Telecomunicación
Universitat Politècnica de València
Edificio 4D. Camino de Vera, s/n, 46022 Valencia
Tel. +34 96 387 71 90, ext. 77190
www.etsit.upv.es

VLC/
CAMPUS
VALENCIA, INTERNATIONAL
CAMPUS OF EXCELLENCE





UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

— **TELECOM** ESCUELA
TÉCNICA **VLC** SUPERIOR
DE INGENIERÍA DE
TELECOMUNICACIÓN

Agradecimientos

A mi familia, por su constante apoyo y los ánimos recibidos a lo largo de estos 5 años, en especial a mis hermanos. Agradecer a mis amigos de residencia y a mis compañeros de clase, he aprendido de todos ellos un poco.

Por último, agradecer a mi profesor y tutor, Lorenzo Rubio Arjona por el tiempo y la dedicación que ha empleado en mí para conseguir finalizar este trabajo. Gracias por tu paciencia durante estos meses.



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

— **TELECOM** ESCUELA
TÉCNICA **VLC** SUPERIOR
DE INGENIERÍA DE
TELECOMUNICACIÓN

Resumen

En este Trabajo Final de Grado se ha caracterizado el canal radio haciendo uso de medidas experimentales de propagación de las ondas electromagnéticas en un entorno *indoor*. Para ello se determinan modelos estadísticos basados en el modelo de Saleh-Valenzuela. Con el empleo de herramientas sofisticadas como SAGE, se obtienen las direcciones de llegada de las contribuciones multicamino, en concreto: el retardo, la amplitud, el ángulo acimut y el ángulo de elevación. Por último, se comparan los resultados de clusterizar con el algoritmo lineal, utilizado en trabajos anteriores, con los resultados de emplear el modelo de clusterización basado en Machine Learning, dando buenos resultados cuando se emplea la información que ofrece el canal en conjunto.

Palabras clave: Modelado de canal; clusterización; propagación; ML; milimétricas

Resum

En aquest Treball Final de Grau s'ha caracteritzat el canal radie fent ús de mesures experimentals de propagació de les ones electromagnètiques en un entorn *indoor*. Per a això es determinen models estadístics basats en el model de Saleh-Valenzuela. Amb l'ús d'eines sofisticades com SAGE, s'obtenen les direccions d'arribada de les contribucions multicamí, en concret: el retard, l'amplitud, l'angle azimut i l'angle d'elevació. Finalment, es comparen els resultats de clusteritzar amb l'algorisme lineal, utilitzat en treballs anteriors, amb els resultats d'emprar el model de clusterización basat en Machine Learning, donant bons resultats quan s'empra la informació que ofereix el canal en conjunt.

Paraules clau: Modelatge de canal; clusterizació; propagació; ML; mil·limètriques



Abstract

In this Final Degree Project, the radio channel has been characterized using experimental measurements of electromagnetic wave propagation in an indoor environment. To achieve this, statistical models based on the Saleh-Valenzuela model are determined. By employing sophisticated tools such as SAGE, the arrival directions of multi-path contributions are obtained, specifically: delay, amplitude, azimuth angle, and elevation angle. Finally, the results of clustering using the linear algorithm, used in previous works, are compared with the results obtained by employing the Machine Learning-based clustering model. Good results are achieved when utilizing the information provided by the channel as a whole.

Key words: Channel modeling; clustering; multipath contributions; propagation; ML; mmWave



Índice General

Capítulo 1. INTRODUCCIÓN Y OBJETIVOS	1
1.1 Introducción	1
1.2 Objetivos	2
1.3 Organización de la memoria	3
Capítulo 2. METODOLOGÍA	5
2.1 Gestión del proyecto.....	5
2.2 Identificación de tareas.....	6
2.3 Diagrama de tiempo	7
Capítulo 3. FUNDAMENTOS DEL CANAL RADIO.....	9
Capítulo 4. CAMPAÑA DE MEDIDAS	15
4.1 Sonda de medidas.....	15
4.2 Entorno y bandas de frecuencia	16
Capítulo 5. TÉCNICAS DE CLUSTERIZACIÓN	19
5.1 Clusterización basada en amplitud y retardo.....	19
5.2 Clusterización por separado: Ángulos y retardo.	31
5.3 Algoritmo de identificación de solapes	36
Capítulo 6. APLICACIÓN DE MACHINE LEARNING A LA CLUSTERIZACIÓN..	40
6.1 Algoritmo Kmeans	40
6.1.1 Inicialización	41
6.1.2 Cálculo de distancias	42
6.1.3 Iteraciones y repeticiones	42
6.1.4 Proceso de validación.....	44
6.1.5 Implementación de Kmeans en Matlab	47
Capítulo 7. RESULTADOS.....	49
7.1 Clusterización con Kmeans	49
7.2 Modelizado del canal	54
7.3 Clusterización a partir del ángulo acimut y retardo.....	58
Capítulo 8. CONCLUSIONES Y LÍNEAS FUTURAS	63
8.1 Conclusiones	63
8.2 Líneas futuras	65
REFERENCIAS	66



Índice de figuras

Figura 1. Diagrama de Gantt con las tareas y sus periodos de realización.	7
Figura 2. Mapa de la propagación de las ondas EM en un entorno indoor.	9
Figura 3. Filtro FIR por el que se representa el efecto dispersivo del canal.	10
Figura 4. Representación de Interferencia Entre Símbolos.	11
Figura 5. Representación de la función de transferencia de las 2000 primeras contribuciones discretizadas del canal.	12
Figura 6. Identificación de los parámetros principales del modelo de propagación estadístico de Saleh-Valenzuela.	13
Figura 7. Plano de la sala de becarios del GRE, donde se realizarán las medidas en milimétricas.	16
Figura 8. Perfil de potencia sintético a partir de parámetros ideales de Saleh-Valenzuela.	20
Figura 9. Nube de puntos PDP ideal.	20
Figura 10. Clusterización del PDP ideal con algoritmo de Corrigan.	21
Figura 11. Representación del APDP y la línea de threshold (-25 dB).	22
Figura 12. Clusterización a partir de Corrigan del APDP para primera posición LOS en B1 del URA.	22
Figura 13. Clusterización con Corrigan del APDP para primera posición LOS en B2 del URA.	23
Figura 14. Clusterización con Corrigan del APDP para primera posición LOS en B3 del URA.	23
Figura 15. Clusterización con Corrigan del APDP para primera posición LOS en B4 del URA.	24
Figura 16. Superposición del APDP del URA (azul) con los datos obtenidos de SAGE (verde).	24
Figura 17. Clusterización con Corrigan a partir de los datos de SAGE para primera posición LOS en B1.	25
Figura 18. Clusterización con Corrigan a partir de los datos de SAGE para primera posición LOS en B2.	25
Figura 19. Clusterización con Corrigan a partir de los datos de SAGE para primera posición LOS en B3.	26
Figura 20. Clusterización con Corrigan a partir de los datos de SAGE para primera posición LOS en B4.	26
Figura 21. Clusterización con Corrigan de los datos de ángulo de elevación de SAGE para la primera posición LOS en B1.	31
Figura 22. Clusterización con Corrigan de los datos de ángulo de elevación de SAGE para la primera posición LOS en B2.	32
Figura 23. Clusterización con Corrigan de los datos de ángulo de elevación de SAGE para la primera posición LOS en B3.	32



Figura 24. Clusterización con Corrigan de los datos de ángulo de elevación de SAGE para la primera posición LOS en B4.	33
Figura 25. Clusterización con Corrigan de los datos de ángulo de acimut de SAGE para la primera posición LOS en B1.	33
Figura 26. Clusterización con Corrigan de los datos de ángulo de acimut de SAGE para la primera posición LOS en B2.	34
Figura 27. Clusterización con Corrigan de los datos de ángulo de acimut de SAGE para la primera posición LOS en B3.	34
Figura 28. Clusterización con Corrigan de los datos de ángulo de acimut de SAGE para la primera posición LOS en B4.	35
Figura 29. Clusterización con Corrigan del PDP en LOS1 B1.	36
Figura 30. Ejemplo de solape 1.	37
Figura 31. Ejemplo de solape 2.	37
Figura 32. Ejemplo de solape 3.	38
Figura 33. Ejemplo de proceso de iteraciones y repeticiones. Captura de Matlab.	43
Figura 34. Clusterización con el algoritmo Kmeans del PDP sintético ideal.	45
Figura 35. Proceso de validación con el método de Calinski-Harabasz cuyo resultado no es convergente.	45
Figura 36. Proceso de validación con el método de Calinski-Harabasz cuyo resultado sí es convergente.	46
Figura 37. Nube de puntos representados en 3D con el conjunto de variables de SAGE.	49
Figura 38. Proceso de validación con el método de Calinski-Harabasz, donde la solución converge en 4 clústeres.	50
Figura 39. Resultados de la clusterización con Kmeans empleando el conjunto de los datos de SAGE para la primera posición LOS en B1.	50
Figura 40. Proyección del ángulo de elevación y el retardo de la figura 38.	51
Figura 41. Proyección de la amplitud normalizada y el retardo de la figura 38.	51
Figura 42. Proyección del ángulo acimut y retardo de la figura 38.	52
Figura 43. Resultados de la clusterización con Kmeans empleando el conjunto de los datos de SAGE para la primera posición NLOS en B1.	53
Figura 44. Proyección del ángulo acimut y retardo de la figura 42.	53
Figura 45. Diagrama en polares del ángulo acimut y el retardo de llegada para la primera posición LOS en B1, clusterización Kmeans con el conjunto de datos.	58
Figura 46. Representación en polares del ángulo de llegada acimut y el retardo, para la primera posición LOS en B1.	59
Figura 47. Superposición de la figura 43 en el plano de la sala de becarios, indicando la dirección de llegada de las contribuciones LOS.	60
Figura 48. Representación en polares del ángulo de llegada acimut y el retardo, para la primera posición NLOS en B1.	61



Figura 49. Superposición de la figura 45 en el plano de la sala de becarios, indicando la dirección de llegada de las contribuciones NLOS. 61

Índice de tablas

Tabla 1. Resumen campaña de medidas en milimétricas.....	15
Tabla 2. Resumen de las bandas medidas en milimétricas.....	17
Tabla 3. Número de clústeres obtenidos con Corrigan a datos de canal y SAGE en posición LOS.	27
Tabla 4. Número de clústeres obtenidos con Corrigan a datos de canal y SAGE en posición NLOS.	27
Tabla 5. Parámetros de canal del modelo SV en cada banda de las posiciones LOS obtenidos con algoritmo de Corrigan.	29
Tabla 6. Parámetros de canal del modelo SV en cada banda de las posiciones NLOS obtenidos con algoritmo de Corrigan.....	30
Tabla 7. Tasa de solape en las bandas de la primera posición LOS.....	38
Tabla 8. Tasa de solape en las bandas de la primera posición NLOS.....	38
Tabla 9. Número total de clústeres en las bandas de la primera posición LOS.	39
Tabla 10. Número total de clústeres en las bandas de la primera posición NLOS.....	39
Tabla 11. Tabla resumen de los métodos de inicialización del algoritmo Kmeans.....	41
Tabla 12. Tabla resumen con los tipos de cálculo de distancia en el algoritmo Kmeans.....	42
Tabla 13. Identificación de los criterios internos de la función Kmeans de Matlab.	48
Tabla 14. Tabla resumen con los resultados de emplear todos los datos de SAGE.	52
Tabla 15. Parámetros de canal del modelo SV en cada banda de las posiciones LOS obtenidos con algoritmo Kmeans.....	55
Tabla 16. Parámetros de canal del modelo SV en cada banda de las posiciones NLOS obtenidos con algoritmo Kmeans.....	56
Tabla 17. Resumen media y desviación típica de los parámetros de las posiciones LOS obtenidos a partir del algoritmo de Corrigan.....	57
Tabla 18. Resumen media y desviación típica de los parámetros de las posiciones NLOS obtenidos a partir del algoritmo de Corrigan.....	57
Tabla 19. Resumen media y desviación típica de los parámetros de las posiciones LOS obtenidos a partir del algoritmo Kmeans.	57
Tabla 20. Resumen media y desviación típica de los parámetros de las posiciones NLOS obtenidos a partir del algoritmo Kmeans.	57



Listado de acrónimos

CIR	Channel Impulse Response
TDL	Time Delay Line
PDP	Power Delay Profile
SNR	Signal to-Noise Ratio
LOS	Line Of Sight
NLOS	Non Line Of Sight
SV	Saleh-Valenzuela
URA	Uniform Rectangular Array
VNA	Vector Analyzer Adapter
UPV	Universidad Politécnica de Valencia
GRE	Grupo de Radiación Electromagnética
iTEAM	Instituto de Telecomunicaciones y Aplicaciones Multimedia
APDP	Average Power Delay Profile
CH	Calinski-Harabasz
BCSM	Between Cluster Scatter Matrix
WCSM	Whitin Cluster Scatter Matrix
ML	Machine Learning
SAGE	Space-Alternating Generalized Expectation-maximization
MPC	Multipath Contributions
FIR	Finite Impulse Response
Knn	K Nearest Neighbor

Capítulo 1. INTRODUCCIÓN Y OBJETIVOS

1.1 Introducción

El estudio del canal radio tiene como objetivo describir la propagación de las ondas EM (*Electromagnéticas*) en un entorno específico, con el fin de obtener modelos estadísticos que permitan a los sistemas optimizar la señal recibida. Por ello surge interés en analizar los múltiples efectos que distorsionan la propagación de las ondas, como por ejemplo la propagación multicamino o el efecto Doppler. Estos efectos distorsionan las comunicaciones, introduciendo ruido, atenuación, dispersión temporal, interferencia entre símbolos, fading, etc. Contrarrestar estos efectos de canal es especialmente útil en el campo de las comunicaciones inalámbricas.

Conseguir un modelo estadístico que caracterice el canal radio es una tarea compleja, ya que éste es variante en el tiempo, haciendo que exista más de un posible camino de propagación de la señal.

El canal puede ser estudiado a partir de modelos basados en la respuesta impulsional (CIR, *Channel Impulse Response*). Estos modelos representan la respuesta del canal en función del instante y el retardo de llegada de las contribuciones, $h(t, \tau)$. Para describir la respuesta impulsional se emplean modelos basados en línea de retardo (TDL, *Time Delay Line*), que utilizan la potencia de cada contribución percibida en el receptor con su retardo de llegada y ponderadas por la atenuación sufrida en el camino. Un ejemplo de este tipo de representaciones es el perfil de retardo de potencia (PDP, *Power Delay Profile*). Aplicando herramientas más sofisticadas de análisis de canal SAGE (*Space-Alternating Generalized Expectation-maximization*) se puede reconstruir el frente de onda y extraer información sobre los ángulos de llegada de las contribuciones: acimut y elevación.

En este trabajo se prestará especial atención al efecto multicamino, ya que influye en gran medida en la propagación de las ondas EM en entornos *indoor*. Este efecto se define como el resultado de la superposición de las contribuciones ó réplicas originadas por los diferentes caminos que ha tomado el frente de onda transmitido. Está provocado por aquellos elementos que interfieren en la propagación y las consecuencias directas en la comunicación son la distorsión temporal y la selectividad en frecuencia.

La distorsión temporal impide al receptor diferenciar dos pulsos consecutivos de uno ensanchado. Esto hace que distinguir las contribuciones multicamino (MPC, *Multipath Contributions*) tenga

un efecto positivo, permitiendo separar las contribuciones y combinarlas, reduciendo el nivel de señal a ruido (SNR, *Signal to-Noise Ratio*).

Para conseguir distinguir las contribuciones multicamino se emplean herramientas de clusterización, que, a partir de la información obtenida del canal, permiten formar grupos con características comunes.

Se pueden emplear diversos criterios, clásicamente se emplea la clusterización sobre datos de amplitud y retardo, aplicándose directamente sobre el PDP. Ejemplos de estos métodos pueden ser el algoritmo de Corrigan [1] que emplea líneas de tendencia de mínimos cuadrados, la clusterización en función del retardo de grupo, etc. Todas ellas basadas en el retardo y en la amplitud, dando lugar a métodos lineales de clusterización. Por otro lado, acompañado del auge del desarrollo de modelos basados en ML (*Machine Learning*), se propone utilizar un método basado en ML no supervisado, en concreto, el algoritmo Kmeans.

Estos últimos modelos son considerados más sofisticados y eficientes, ya que se basan en el cálculo de distancias entre los puntos y entre los clústeres, siendo posible emplear más variables a parte de la amplitud y el retardo, por ejemplo: ángulos de elevación y acimut. Se propone demostrar que los modelos multivariante de ML son significativamente más eficientes cuando utilizan información adicional de ángulos de llegada.

1.2 Objetivos

El objetivo principal es parametrizar el canal radio con el fin de obtener modelos de canal estadísticos basados en el modelo de Saleh-Valenzuela. Para ello se hará uso de información de amplitud y retardo en combinación con información de las direcciones de llegada de las contribuciones, como son los ángulos de acimut y elevación.

El primer objetivo será determinar que la información de las direcciones de llegada es especialmente útil en la clusterización de las contribuciones multicamino, y por tanto en la modelización del canal.

El segundo objetivo será proponer un método de clusterización basado en Machine Learning, para ello se hará uso de la Toolbox de Matlab de Estadística y Machine Learning, que contiene la función Kmeans. Se compararán los resultados obtenidos con el algoritmo Kmeans con los resultados obtenidos a partir del modelo de clusterización lineal de trabajos anteriores [2].



1.3 Organización de la memoria

La memoria está estructurada en 7 capítulos:

- En el Capítulo 2 se describe la metodología, la búsqueda de tareas y la distribución temporal.
- En el Capítulo 3 se definen los fundamentos teóricos básicos del canal radio que influyen en la caracterización del canal.
- El Capítulo 4 se detalla la campaña de medidas en ondas milimétricas realizada que servirá para el desarrollo del TFG (*Trabajo Final de Grado*).
- En el Capítulo 5 se emplean los métodos de clusterización lineales empleados en trabajos anteriores con el fin de obtener los parámetros SV del canal.
- El Capítulo 6 describe el método de clusterización basado en ML y se utilizan los resultados para obtener los parámetros del modelo SV.
- Finalmente, en el Capítulo 7, se recogen las conclusiones y las líneas futuras alcanzadas en el trabajo



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

— **TELECOM** ESCUELA
TÉCNICA **VLC** SUPERIOR
DE INGENIERÍA DE
TELECOMUNICACIÓN

Capítulo 2. METODOLOGÍA

En este capítulo se tratan los procesos llevados a cabo durante el desarrollo del proyecto, desde un planteamiento inicial que surge del trabajo realizado previamente por otros autores, hasta las conclusiones alcanzadas en el capítulo sexto. Además, se describen las tareas que se han realizado a lo largo del proceso de elaboración del trabajo, desde búsqueda bibliográfica hasta programación y redacción. Todo ello quedará plasmado en un diagrama de Gantt que identifica temporalmente la duración de cada tarea.

2.1 Gestión del proyecto

Este proyecto tendrá una estructura que permita seguir la línea argumentativa descrita en los objetivos y por tanto permita al lector entender las conclusiones y las necesidades de mejorar los métodos que aquí se van a emplear.

En primer lugar, se realizará una explicación resumida sobre los principios teóricos y físicos que ocurren en el canal radio y que son los principales actores que justifican la necesidad de modelizar el canal radio a partir de modelos TDL. En él se definirán los principales motivos por los que es necesario obtener modelos de propagación eficientes, desde el ruido que introduce la propagación de las ondas electromagnéticas, hasta los efectos dispersivos como son el multicamino y el Doppler.

La segunda parte del proyecto consistirá en una aplicación práctica de los algoritmos de clusterización con datos obtenidos de la campaña de medidas realizada en el laboratorio de becarios del iTEAM (*Instituto de Telecomunicaciones y Aplicaciones Multimedia*). Se tomará como punto de partida la aplicación de los algoritmos de clusterización lineales basados en retardo y amplitud para demostrar que los ángulos complementan los resultados aportando información útil, clave a la hora de obtener los parámetros de los modelos de canal típicos (SV).

Si la información de los ángulos es útil, o por lo menos proporciona otro punto de vista a la hora de agrupar las contribuciones, el siguiente paso será determinar la combinación de información que optimiza la clusterización. Para ello se procederá a utilizar el algoritmo de clusterización de Corrigan a los datos de amplitud y de ángulos por separado, de forma que se pueda comprobar que existe una combinación de información tal que mejore la clusterización del perfil de retardo de potencia. En función de los resultados obtenidos por la clusterización de la información por separado se procederá a la clusterización de los datos en conjunto, para ello será imprescindible el uso de nuevos métodos de clusterización, para lo cual se precisará algoritmos basados en ML



que sean capaces de tratar toda la información con el fin de agrupar las contribuciones de la forma más eficiente.

Por último, se obtendrán los parámetros que definen los modelos de canal de Saleh-Valenzuela atendiendo a los resultados de la clusterización con algoritmos lineales y con el algoritmo Kmeans, diferenciando entre medidas de posición con visión directa y medidas sin visión directa.

2.2 Identificación de tareas

Tarea 1 (T1): Documentación sobre el canal radio y los efectos dispersivos que influyen en la propagación de ondas electromagnéticas. Así como los aspectos teóricos útiles para interpretar las medidas en milimétricas obtenidas del URA.

Tarea 2 (T2): Documentación de los algoritmos de clusterización lineales como el de Corrigan y los algoritmos basados en ML.

Tarea 3 (T3): Programación en Matlab y obtención de la función de transferencia, PDP y APDP a partir de datos de canal. Así como el desarrollo del código que implementa el método de clusterización lineal.

Tarea 4 (T4): Programación del modelo de clusterización basado en ML, así como los procesos de validación de la función Kmeans.

Tarea 5 (T5): Parametrización de los modelos estadísticos de propagación para medidas SAGE LOS y NLOS.

Tarea 6 (T6): Realización de la memoria, así como borradores e informes.

Tarea 7 (T7): Realización de la presentación para la defensa.

Tarea 8 (T8): Reuniones de seguimiento de las entregas que se han realizado a lo largo del periodo de realización del TFG.

2.3 Diagrama de tiempo

El diagrama de Gantt está dividido en cuatro áreas de trabajo principales, como son la documentación, la programación, la elaboración de los informes y los sucesivos borradores de la memoria y por último las reuniones entre alumno y tutor. Dichas reuniones se han llevado a cabo de forma frecuente, tanto en forma presencial como a través de la aplicación Teams.

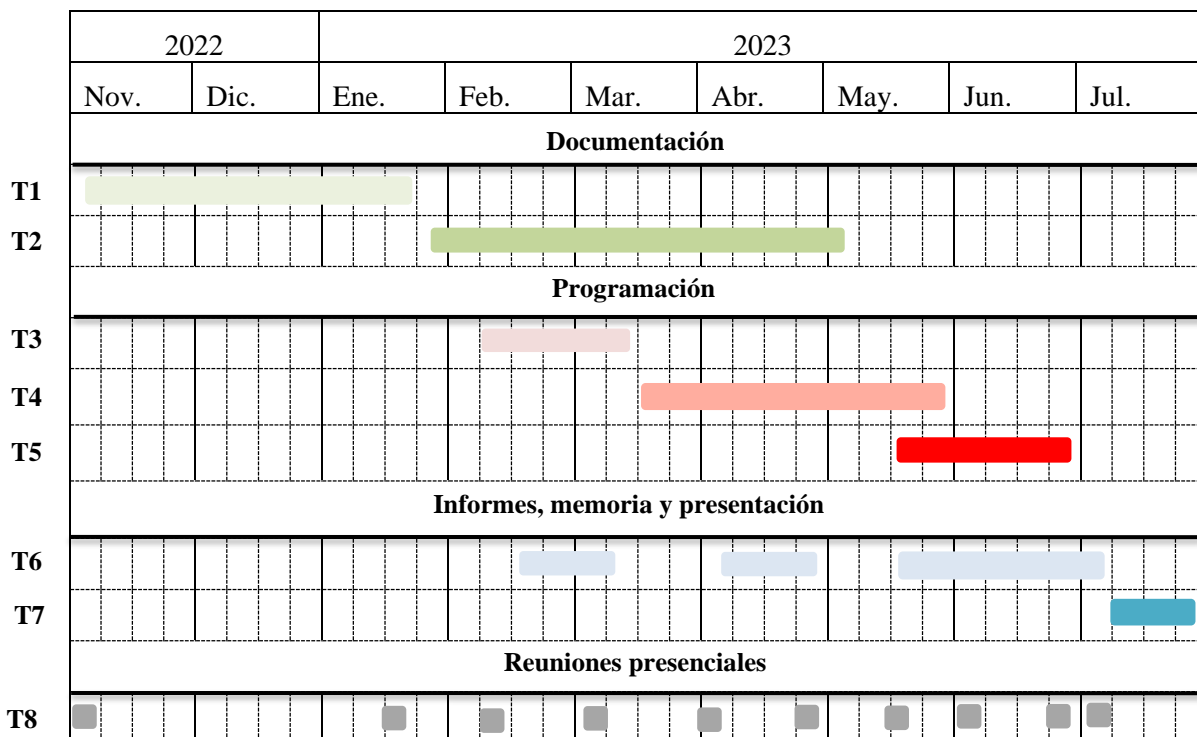


Figura 1. Diagrama de Gantt con las tareas y sus periodos de realización.



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

— **TELECOM** ESCUELA
TÉCNICA **VLC** SUPERIOR
DE INGENIERÍA DE
TELECOMUNICACIÓN

Capítulo 3. FUNDAMENTOS DEL CANAL RADIO.

El canal radio se define como el medio existente entre transmisor y el receptor por el que las ondas EM emitidas avanzan hasta alcanzar al receptor. Los sistemas que implementan al emisor y al receptor pueden ser controlados desde el punto de vista práctico, mediante un diseño o una programación previa, pero el canal radio es parte de la naturaleza, es decir, no es estático y por tanto lo convierte en un sistema cambiante e impredecible. Un medio siempre en constante cambio, obligando a aquellos que lo utilizan a comprenderlo, estudiarlo y parametrizarlo con la finalidad de entender y contrarrestar sus defectos en las comunicaciones. El canal radio *indoor*, como el representado en la Figura 2, se define en el dominio temporal como un sistema cronovariable con una respuesta impulsional $h(t, \tau)$.

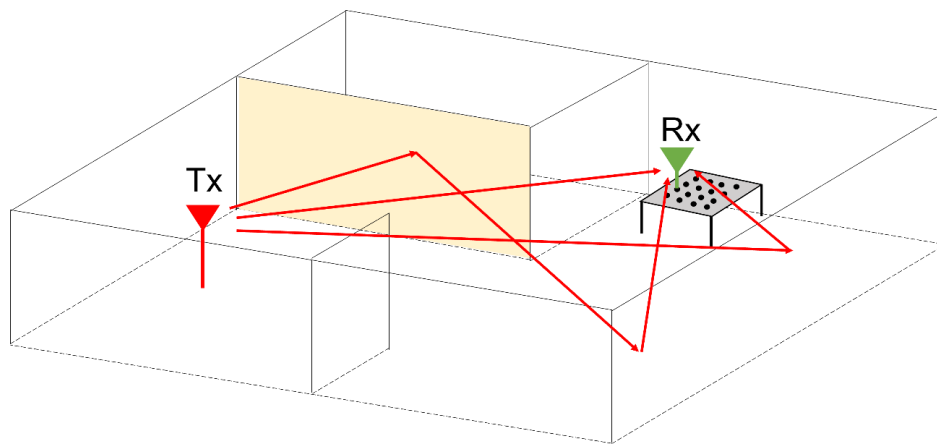


Figura 2. Mapa de la propagación de las ondas EM en un entorno indoor.

Donde τ es el exceso de retardo, tomando como referencia el instante en el que la primera contribución alcanza al receptor.

Si se considera la señal de entrada emitida por el transmisor como $z(t)$ y la señal de salida como $w(t)$, se puede definir el efecto del canal radio a través de la expresión (1).

$$w(t) = \int_{-\infty}^{\infty} z(t - \tau) \cdot h(t, \tau) d\tau \quad (1)$$

La expresión anterior está caracterizada para el dominio continuo, en la práctica los elementos medidores de campo discretizan las señales recibidas. En (2) se representa de forma discreta.

$$w(t) = \Delta\tau \sum_{m=1}^n z(t - m\Delta\tau) \cdot h(t, m\Delta\tau) \quad (2)$$

Si la señal de entrada es un pulso (delta), la salida del canal, desde el punto de vista discreto, es un sumatorio compuesto por la superposición de todas las réplicas o contribuciones de la señal original, cada una con un retardo y ponderada por un peso, es decir, atenuadas. Dicha atenuación dependerá de los caminos que haya tomado el frente de ondas emitido por el transmisor. De forma esquemática se puede definir el canal radio como un filtro FIR (*Finite Impulse Response*) tal y como se muestra en la Figura 3.

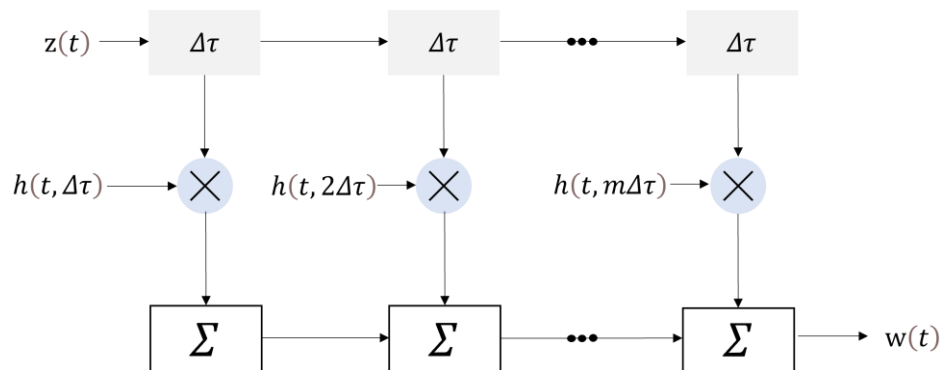


Figura 3. Filtro FIR por el que se representa el efecto dispersivo del canal.

Producto de la superposición de las réplicas surgidas por el fenómeno multicamino en la propagación surge la dispersión temporal, íntimamente relacionada con el ensanchamiento de los símbolos. Desde el punto de vista del receptor, si se representan todas las contribuciones que lo alcanzan, se distinguiría un pulso ensanchado o varios pulsos, ya que la señal recibida se corresponde con la superposición de los ecos de la señal original retardada y atenuada, tal y como se puede observar en la Figura 4. Por tanto, es fundamental que el receptor sea capaz de distinguir las contribuciones multicamino ya que el ensanchamiento de los pulsos provocará IES (*Interferencia Entre Símbolos*), obligando a reducir la velocidad de la transmisión, según lo establecido por el teorema de Shannon, reduciendo la capacidad de transmisión del canal.

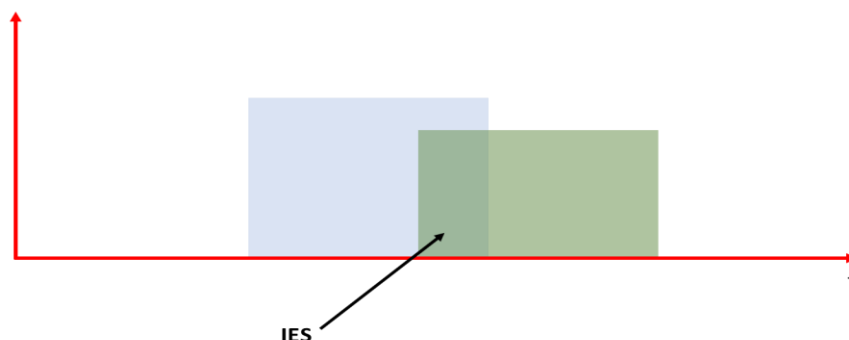


Figura 4. Representación de Interferencia Entre Símbolos.

Por otro lado, el movimiento entre emisor y receptor provoca selectividad en frecuencia, desde el punto de vista de las mediciones realizadas se debe tener en cuenta que el receptor realiza medidas dentro de un array de posiciones, de esta forma se considera que el receptor está cambiando su posición. Lo cual permitirá, mediante herramientas como SAGE, poder reconstruir el canal radio con información relacionada de los ángulos de llegada. Por lo que los modelos no solo pueden distinguirse a partir de los retardos y la amplitud de la señal recibida, sino que también con los ángulos. La respuesta impulsional queda definida como $h(t, \tau, \theta, \varphi)$.

Como consecuencia de la propagación multicamino se obtiene en el receptor la superposición de las múltiples réplicas de la señal transmitida. La superposición de las réplicas actúa como suma coherente de las componentes individuales. Dicha suma puede ser en fase o en contrafase. Si el canal fuera estático, dicha suma no cambiaría con el tiempo.

Los caminos que siguen las réplicas tienen retardos distintos en función de la distancia que recorren durante la propagación de la señal. Está estrechamente relacionado con los elementos dispersivos que se encuentra la señal por el camino y que producen efectos de reflexión y difracción.

Cuando existe un único camino de propagación y la señal que se transmite es un pulso en el dominio del tiempo, la respuesta impulsional del canal recibida por el receptor será el mismo pulso pero retardado en tiempo, lo que en el dominio frecuencial vendría representado como una respuesta plana del canal. Conforme aparecen nuevos caminos y réplicas de la señal original, éstas se podrían representar como un tren de deltas, lo que en dominio frecuencial hará que aparezcan desvanecimientos debido a que un tren de deltas en el tiempo se comporta como una función 'sinc' en frecuencia (con valles y nulos) variando la potencia media recibida, fenómeno conocido como shadowing. Puede verse representado en la Figura 5.

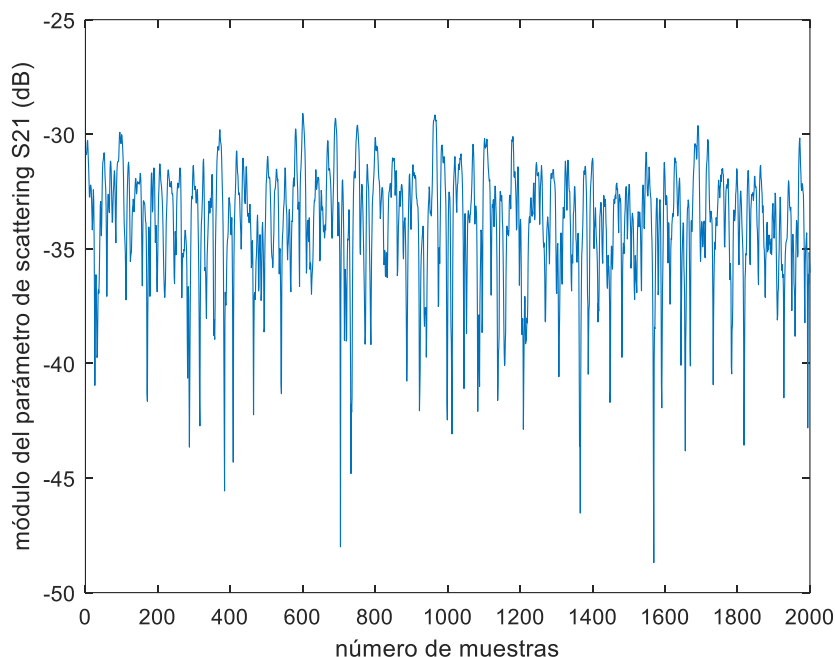


Figura 5. Representación de la función de transferencia de las 2000 primeras contribuciones discretizadas del canal.

Para describir la propagación de señales en el canal radio, se propone emplear los modelos estadísticos TDL. Dichos modelos se basan en que la respuesta impulsional del canal puede ser representada como la superposición de componentes exponenciales complejas. Existen numerosos modelos TDL en la literatura, como por ejemplo el modelo de Saleh-Valenzuela, Jakes, Clarke...etc. El modelo de Saleh-Valenzuela es ampliamente utilizado en las comunicaciones móviles y describe el canal a partir de características de amplitud, retardo y fase. Este modelo es útil para simular canales dispersivos en tiempo, donde la señal puede sufrir cambios rápidos debido a los múltiples caminos y reflexiones en la propagación. En la Figura 6 se muestra un ejemplo que sigue las condiciones del modelo estadístico SV.

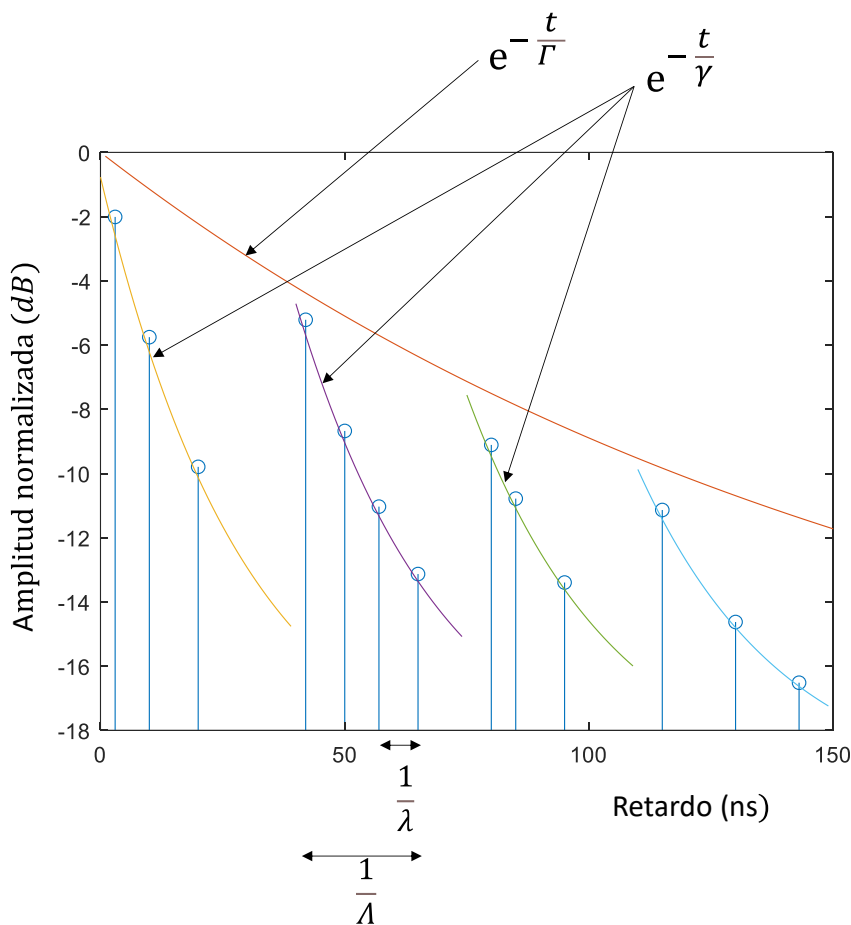


Figura 6. Identificación de los parámetros principales del modelo de propagación estadístico de Saleh-Valenzuela.

El modelo Saleh -Valenzuela [3] viene definido por cuatro parámetros que permitirán obtener los modelos estadísticos SV que caracterizarán el canal. A continuación, se identifica cada parámetro con su significado físico en el contexto de la propagación de las ondas EM.

- Retardo propagación medio (λ): Indica el espaciado temporal que existe entre las MPC de un mismo clúster.
- Retardo de los clústeres (Λ): Indica el espaciado temporal que ocupa cada clúster. Se estima una media entre todos los clústeres encontrados.
- Decaimiento medio (Γ): Indica el espaciado temporal que ocupa cada clúster. Se estima una media entre todos los clústeres encontrados.



Decaimiento de los clústeres (γ): Indica la constante por la que las MPC de un mismo clúster decaen en amplitud con el tiempo.

Capítulo 4. CAMPAÑA DE MEDIDAS

4.1 Sonda de medidas

Se han realizado medidas de propagación entre dos antenas, una transmisora y otra receptora, con la intención de extraer información de canal y conseguir caracterizarlo a partir de la descripción de modelos típicos como SV. Para ello se han empleado antenas de tipo omnidireccional en el plano horizontal y polarización vertical, tanto en el transmisor como en el receptor. La transmisión consistirá en realizar un barrido desde la frecuencia de 15GHz hasta la frecuencia de 40 GHz, estas frecuencias se encuentran en el rango de milimétricas, en referencia a la longitud de onda utilizada, y están pensadas para el despliegue de comunicaciones de quinta generación [4].

El receptor está formado por un *array* rectangular uniforme (URA, *Uniform Rectangular Array*) que consta de 12 filas y 12 columnas. La antena ubicada en la posición del receptor se desplazará entre los 144 puntos dentro del *array*.

Las medidas del canal se obtendrán a partir de un analizador de redes vectorial (VNA, *Vector Analyzer Adapter*), donde en cada posición del *array* se medirá el parámetro de *scattering* $S_{21}(f)$ en su forma compleja. Este parámetro es de carácter complejo y representa la ganancia de transmisión. La información captada en las 144 posiciones del URA se procesará a través del algoritmo SAGE, el cual será capaz de reconstruir, a partir de los datos reales, los ángulos de llegada de elevación y acimut de las contribuciones. A continuación, Tabla 1, se muestra un resumen con la configuración de las medidas.

BW	25 GHz a 40GHz
SPAN	15 GHz
Posiciones array	12x12
Separación	3.04 mm
Número de puntos	8192

Tabla 1. Resumen campaña de medidas en milimétricas.

4.2 Entorno y bandas de frecuencia

El entorno de las medidas es una sala de un edificio de la Universidad Politécnica de Valencia (UPV), que actúa como oficina de becarios del Grupo de Radiación Electromagnética (GRE) perteneciente al iTEAM. En la oficina, cuyo plano se muestra en la Figura 7, se pueden encontrar numerosos elementos de mobiliario como mesas, sillas, ordenadores, estanterías...etc. Estos elementos pueden influir en la propagación a través de la reflexión y refracción de los frentes de onda.

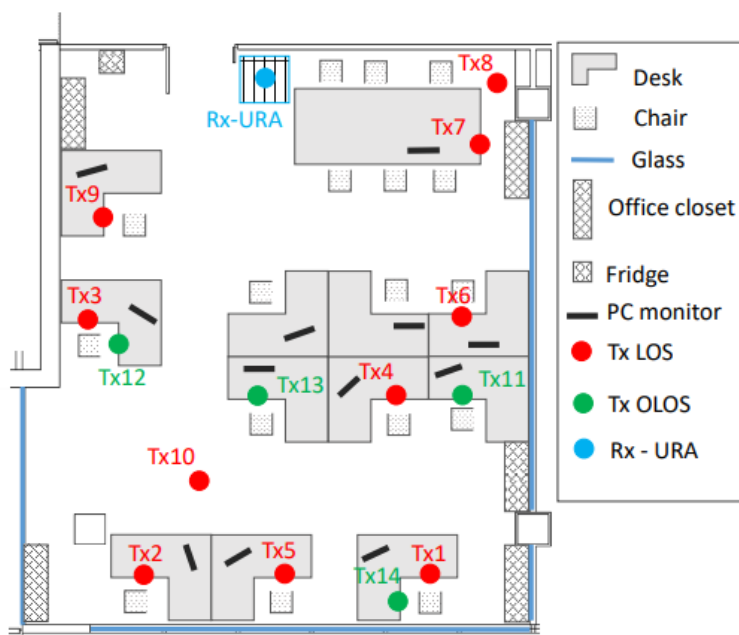


Figura 7. Plano de la sala de becarios del GRE, donde se realizarán las medidas en milimétricas.

A partir de la distribución de los elementos dispersivos en la oficina, se elegirán 14 posiciones donde se realizarán las medidas. El transmisor permanecerá fijo en la misma ubicación. De esta forma existirán medidas que tengan visión directa con el transmisor (LOS, *Line Of Sight*) y otras que no tengan visión directa (NLOS, *Non Line Of Sight*). En total se realizarán medidas en 10 posiciones LOS y 4 en posiciones NLOS.

Se distinguirán 4 bandas dentro del ancho de banda de 25 GHz a 40 GHz para realizar las medidas. Teniendo en cuenta que el receptor muestrea cada 1,831 MHz, si recorre todo el ancho de banda de 25 GHz a 40GHz, se obtienen 8192 observaciones. Las cuales se reparten entre las diferentes bandas. De forma que la primera banda está compuesta por las observaciones 1 a 1365, la segunda banda está comprendida entre las observaciones 1366 a 2457, la tercera banda va de la 3714 a



4587 y la cuarta banda va de la observación 6554 a 8192. En la Tabla 2 se recoge toda la información sobre las bandas B1, B2, B3 y B4.

	Frecuencia inicial (GHz)	Frecuencia final (GHz)	Número de puntos
B1	25	27,5	1365
B2	27,5	29,5	1092
B3	37,8	33,4	874
B4	37	40	1639

Tabla 2. Resumen de las bandas medidas en milimétricas.



Capítulo 5. TÉCNICAS DE CLUSTERIZACIÓN

Para agrupar las contribuciones MPC en clústeres es necesario emplear algoritmos que distingan aquellas características que las hacen semejantes y por tanto sean clasificadas en iguales clústeres. De la misma forma que es necesario que dos contribuciones MPC que no comparten características comunes sean clasificadas en clústeres diferentes. En este apartado se retoma el trabajo realizado por otros autores [2];[5] y se agrupan las MPC a partir del algoritmo lineal. Además, se determinará la idoneidad de emplear información adicional en la clusterización.

5.1 Clusterización basada en amplitud y retardo

En este apartado se empleará el algoritmo de clusterización lineal basado en amplitud y retardo de llegada. La función de este algoritmo será agrupar las contribuciones MPC que han alcanzado al receptor en grupos o *clústeres* con la finalidad de distinguir aquellas contribuciones que compartan características similares. El algoritmo lineal utilizado ya ha sido puesto a prueba en trabajos anteriores [2];[7];[8], también es conocido como el algoritmo de Corrigan. Este algoritmo agrupa, en primer lugar, con respecto a un criterio de espaciado temporal que afectará al retardo de las contribuciones y, en segundo lugar, agrupa en función de la amplitud. Para ello realizará una regresión con la línea de tendencia de mínimos cuadrados a las amplitudes y a partir de valor umbral que actuará como error máximo, se agruparán las contribuciones en clústeres. En resumen, se clasifican las contribuciones atendiendo a dos valores umbrales, uno para el retardo y otro para el error máximo en el ajuste de regresión. Dichos valores umbrales serán escogidos empíricamente poniendo a prueba el algoritmo con un PDP sintético.

A continuación, se representa en la Figura 8 el perfil de retardo de potencia diseñado para poner a prueba el algoritmo lineal de clusterización. Se puede observar que consta de cuatro *clústeres* bien diferenciados, los cuales están contruidos siguiendo los criterios definidos por Saleh-Valenzuela.

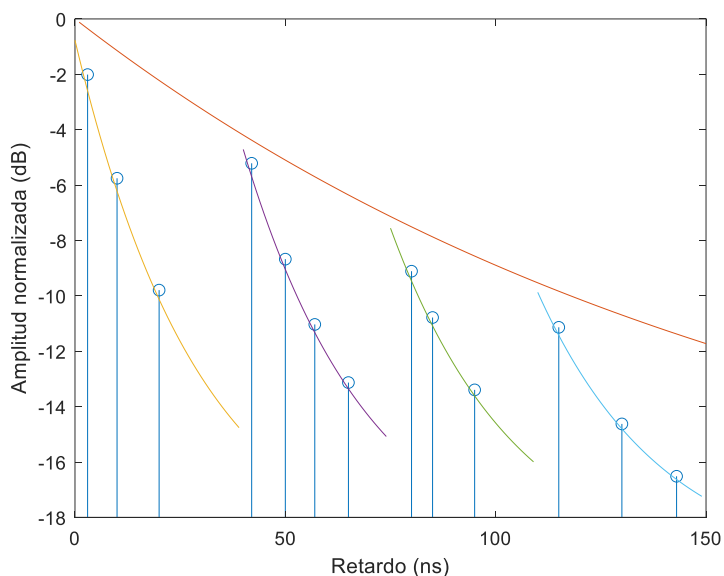


Figura 8. Perfil de potencia sintético a partir de parámetros ideales de Saleh-Valenzuela.

El PDP diseñado se puede considerar ideal y por tanto servirá de prueba para verificar que los parámetros del algoritmo están bien ajustados. En la Figura 9 se representa la nube de puntos que recibirá el algoritmo y que deberá clusterizar. Una vez los valores umbrales del algoritmo permitan diferenciar correctamente los clústeres definidos en el PDP ideal ya se podrá aplicar a la información real del canal, recogida por el VNA.

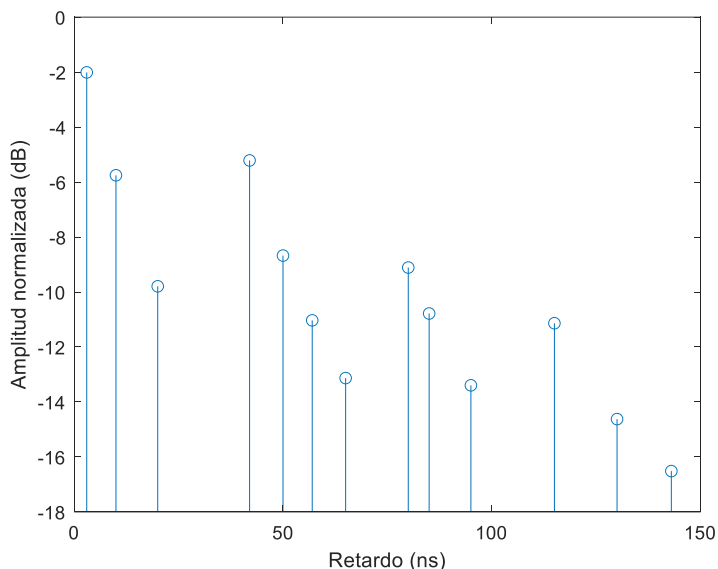


Figura 9. Nube de puntos PDP ideal.

Los resultados del algoritmo de Corrigan aplicados al perfil de potencia ideal se muestran en la Figura 10. Como se puede comprobar, este algoritmo funciona correctamente cuando se utilizan

datos ideales, por tanto, es posible aplicar este algoritmo sobre los datos recogidos en el canal y considerar que la clusterización obtenida permitirá obtener los parámetros de caracterización del canal SV.

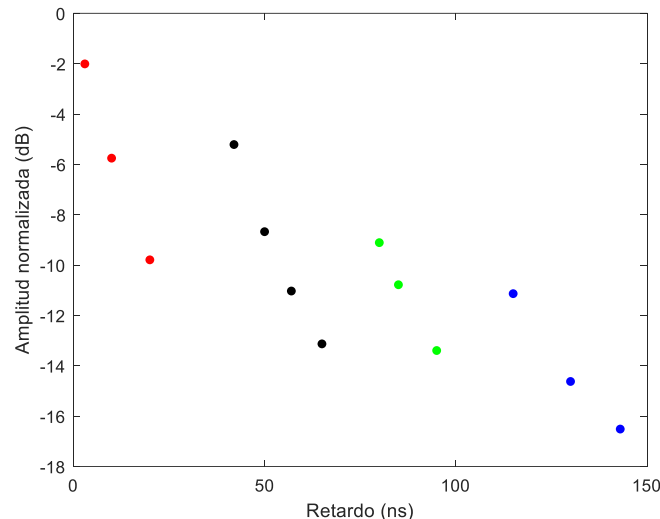


Figura 10. Clusterización del PDP ideal con algoritmo de Corrigan..

El análisis a partir de datos reales del canal se realizará teniendo en cuenta una posición LOS y otra posición NLOS entre las ubicaciones escogidas para realizar las medidas. Las posiciones LOS y NLOS serán la 1 y 11 del plano de medidas, referenciadas en la Figura 7.

Las medidas recogidas se organizan en una matriz de $12 \times 12 \times 8192$, debido a que se almacena toda la información recogida en cada posición del URA. El VNA ha medido la densidad de potencia que recibía el en su forma compleja, en concreto el parámetro de *scattering* S_{21} , de forma que se puede extraer el perfil de potencia en cada posición del *array*. A partir de dichas medidas se obtendrá el perfil de potencia medio, es decir, un *Average Power Delay Profile*. En (3) se muestra la formulación seguida para obtener el perfil de potencia medio.

$$APDP = \frac{1}{N} \sum_{i=1}^N |h_i(t, \tau)|^2 \quad (3)$$

En la Figura 11 se muestra el APDP de la banda 1 en la primera posición LOS. Introducir un valor de threshold entre -20 y -30 dB con respecto a la amplitud normalizada permitirá obtener la información más relevante recibida por el receptor.

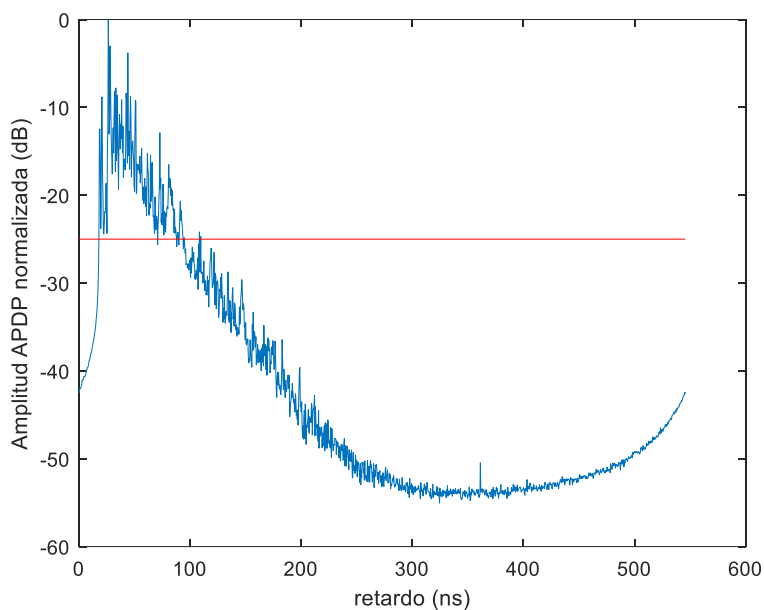


Figura 11. Representación del APDP y la línea de threshold (-25 dB).

La siguiente parte del capítulo recoge los resultados alcanzados por el algoritmo de Corrigan en la clusterización de los APDP de la primera posición LOS y NLOS. Para ello se distinguirá entre las 4 bandas que componen el espectro radioeléctrico entre los 25 GHz y los 40 GHz y se colorearán del mismo color las contribuciones que pertenecen al mismo clúster.

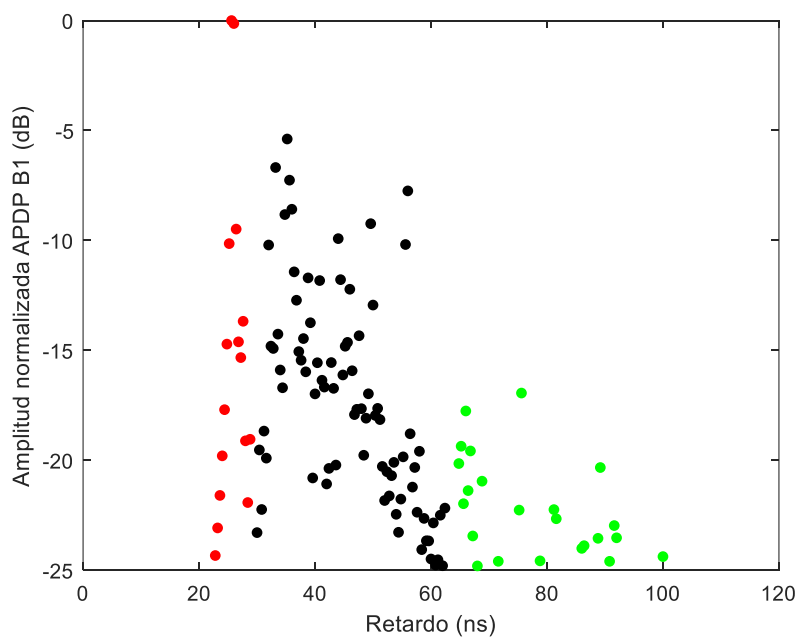


Figura 12. Clusterización a partir de Corrigan del APDP para primera posición LOS en B1 del URA.

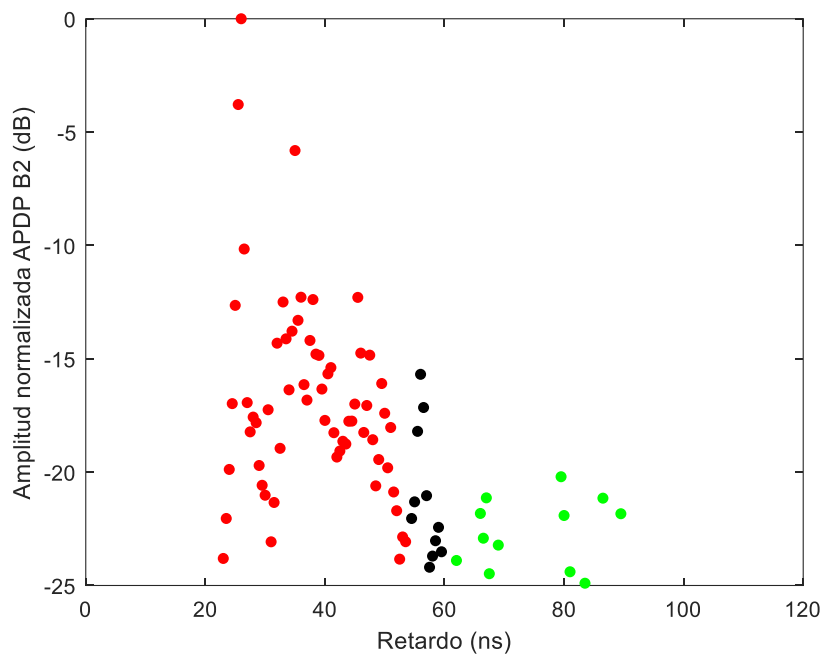


Figura 13. Clusterización con Corrigan del APDP para primera posición LOS en B2 del URA.

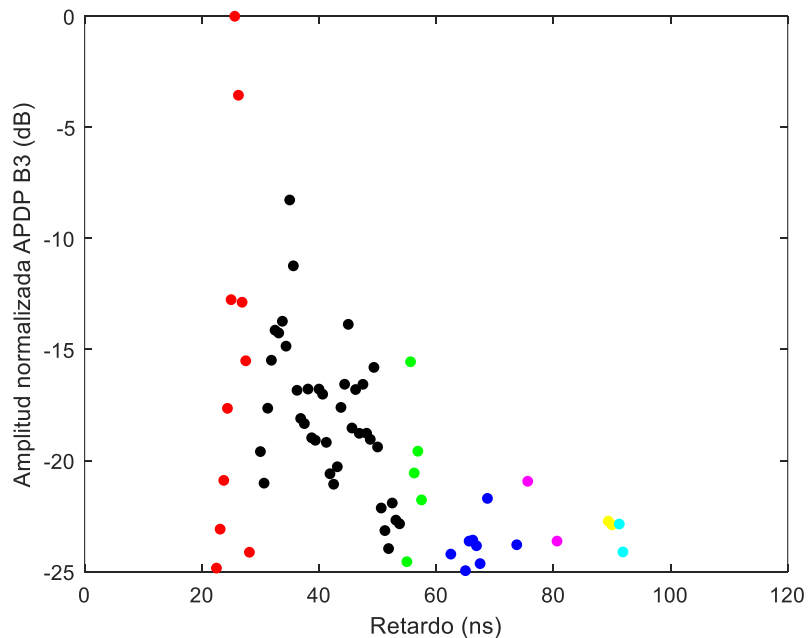


Figura 14. Clusterización con Corrigan del APDP para primera posición LOS en B3 del URA.

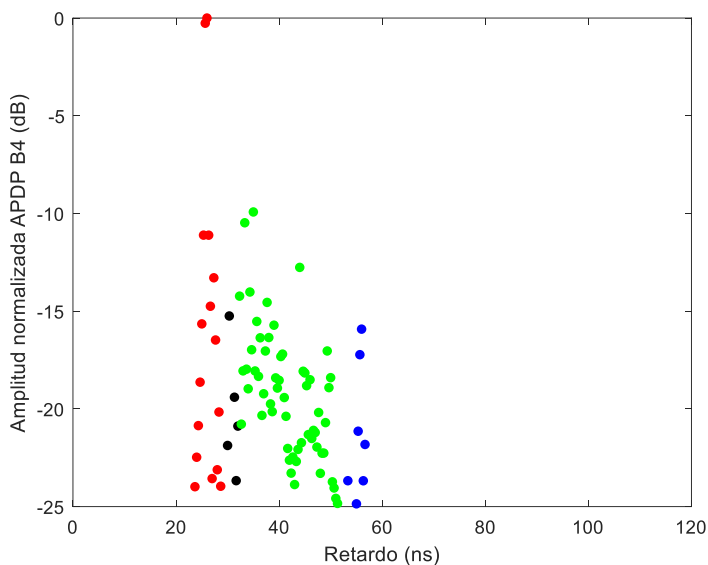


Figura 15. Clusterización con Corrigan del APDP para primera posición LOS en B4 del URA.

Una vez realizada la clusterización con datos del canal, se propone emplear el algoritmo SAGE con el fin de reconstruir el canal y obtener datos de ángulos de llegada. El algoritmo reconstruye el canal muestreando la información recibida con el fin de determinar las 200 contribuciones que mejor describen el canal, con información de retardo, ángulo de acimut, ángulo de elevación y amplitud.

En la Figura 16 se muestra la superposición del APDP del URA con el APDP obtenido con SAGE. Se puede observar que los datos de amplitud de SAGE se ajustan a los datos originales del canal.

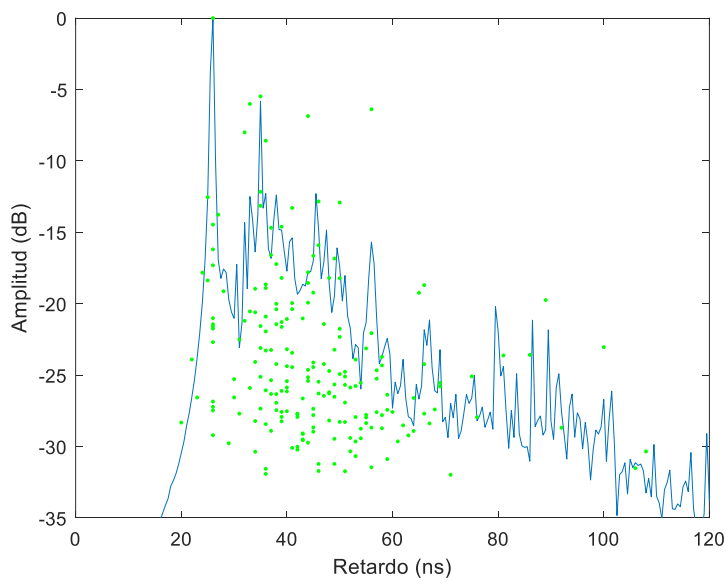


Figura 16. Superposición del APDP del URA (azul) con los datos obtenidos de SAGE (verde).

El siguiente paso es emplear el algoritmo de Corrigan con los datos de SAGE. Para ello se utilizarán los mismos valores umbrales empleados en la clusterización de los APDP anteriores con datos de canal. El objetivo es determinar si existe diferencia entre emplear datos de canal y datos con SAGE, de modo que si los resultados no se asemejan reflejaría la necesidad de introducir información adicional en la clusterización.

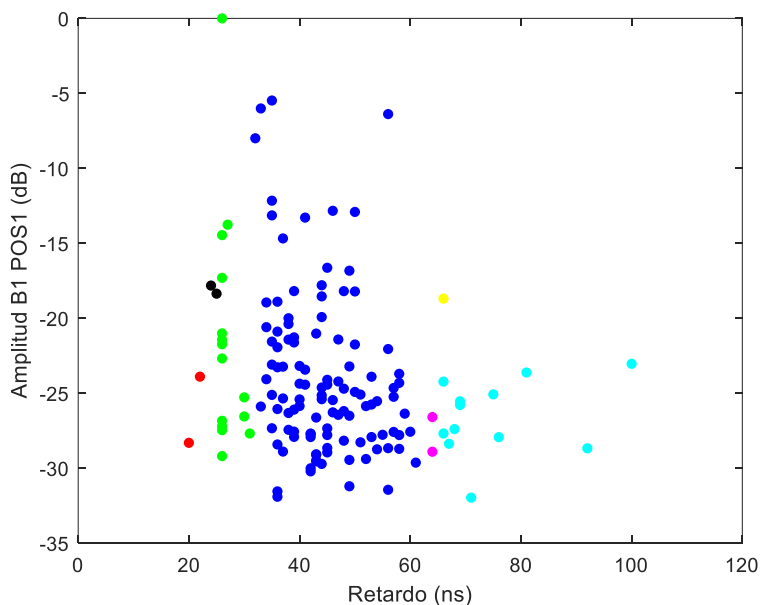


Figura 17. Clusterización con Corrigan a partir de los datos de SAGE para primera posición LOS en B1.

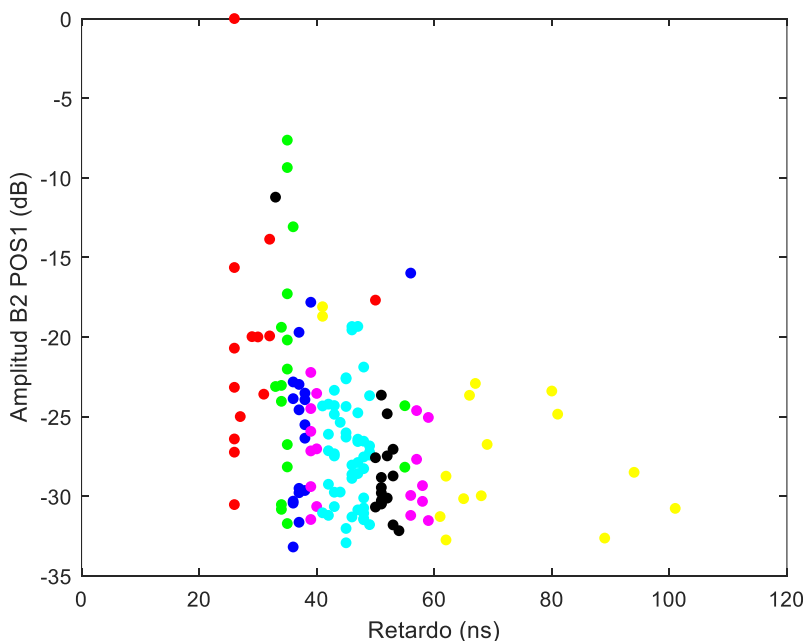


Figura 18. Clusterización con Corrigan a partir de los datos de SAGE para primera posición LOS en B2.

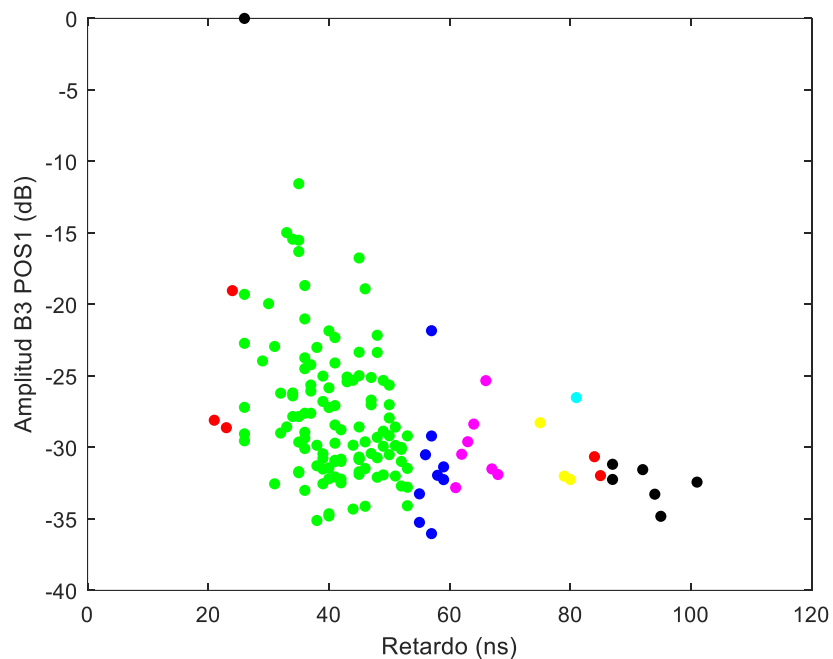


Figura 19. Clusterización con Corrigan a partir de los datos de SAGE para primera posición LOS en B3.

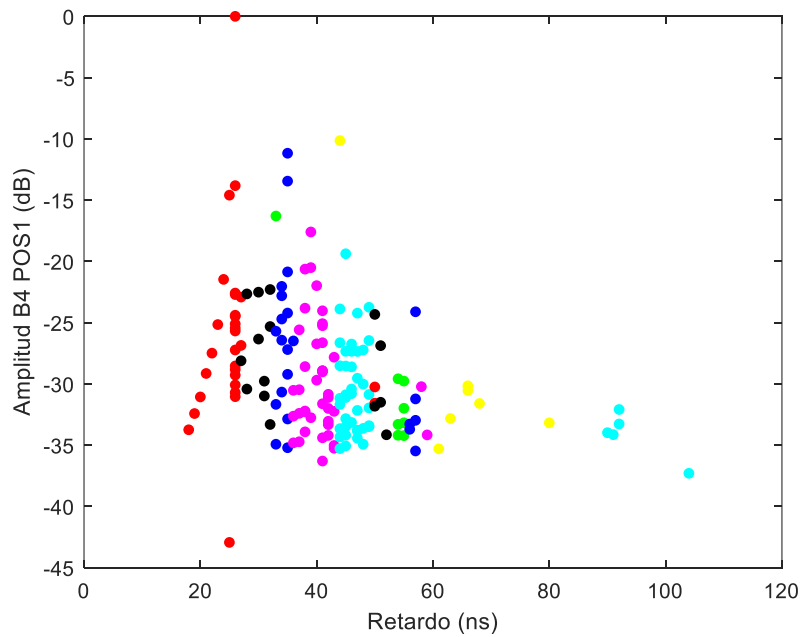


Figura 20. Clusterización con Corrigan a partir de los datos de SAGE para primera posición LOS en B4.

A partir de la clusterización de los datos de SAGE con el algoritmo de Corrigan, se extrae que el resultado dista de lo obtenido cuando se emplea información del canal (URA). Se justifica a partir del número de clústeres finales, mostrándose en las tablas 3 y 4. Analizando los resultados se tiene que el número de clústeres es bajo si se emplean datos de canal y muy alto si se emplea SAGE, tanto en posiciones LOS como posiciones NLOS.

Esto puede deberse al propio muestreo que realiza SAGE, ya que al reducir el número de contribuciones el espaciado temporal entre ellas aumenta, permitiendo la proliferación de más divisiones entre las contribuciones cuando se emplean los mismos parámetros umbrales del algoritmo de Corrigan con los datos del URA, dando lugar a un mayor número de clústeres.

Posición 1 LOS	B1	B2	B3	B4
Data_URA	3	3	7	4
SAGE	7	13	9	14

Tabla 3. Número de clústeres obtenidos con Corrigan a datos de canal y SAGE en posición LOS.

Posición 1 NLOS	B1	B2	B3	B4
Data_URA	2	4	4	3
SAGE	6	15	9	17

Tabla 4. Número de clústeres obtenidos con Corrigan a datos de canal y SAGE en posición NLOS.

Se extraen los parámetros intra-clúster e inter-clúster del modelo de Saleh-Valenzuela, siguiendo los criterios descritos en el Capítulo 3 donde se describían los parámetros del modelo estadístico de Saleh-Valenzuela. En las Tablas 5 y 6 se muestran los parámetros estadísticos obtenidos tras la clusterización de los datos de SAGE. Se ordenan en función de la ubicación LOS y NLOS y se diferencian entre las 4 bandas medidas.

Banda	Posición	γ	Γ	λ	Λ
B1	LOS1	0.00 ¹	8.90	1.80	0.10
	LOS2	0.43	9.62	1.59	0.17
	LOS3	0.00	12.58	1.70	0.17
	LOS4	0.60	14.27	1.39	0.17
	LOS5	4.68	5.37	1.31	0.26
	LOS6	0.62	12.80	1.67	0.31
	LOS7	0.00	31.93	1.78	0.30
	LOS8	1.22	17.52	1.20	0.22
	LOS9	4.09	46.25	1.94	0.35
	LOS10	0.00	10.12	1.79	0.23
B2	LOS1	0.61	6.86	1.96	0.06
	LOS2	0.70	6.89	1.80	0.24
	LOS3	3.29	9.42	1.31	0.04
	LOS4	0.53	8.22	1.77	0.25
	LOS5	0.00	4.07	2.65	0.67
	LOS6	2.64	8.40	1.49	0.21
	LOS7	6.51	17.22	1.33	0.20
	LOS8	0.00	23.36	1.73	0.33
	LOS9	0.00	26.71	1.80	0.36
	LOS10	0.67	12.62	1.52	0.17
B3	LOS1	23.80	7.67	1.81	0.13
	LOS2	3.59	8.70	1.72	0.21
	LOS3	3.41	12.64	1.53	0.07
	LOS4	0.43	12.78	1.89	0.23

¹ Aquellos parámetros asociados al valor 0.00 indican que son valores numéricos pequeños y la primera cifra significativa se encuentra por detrás del cuarto decimal.

Banda	Posición	γ	Γ	λ	Λ
B3	LOS5	0.00	5.05	1.42	0.41
	LOS6	1.44	26.02	0.37	0.07
	LOS7	1.03	16.33	1.57	0.25
	LOS8	1.32	22.71	1.24	0.21
	LOS9	0.00	25.07	1.60	0.24
	LOS10	1.05	12.03	2.00	0.30
B4	LOS1	0.88	6.89	1.94	0.07
	LOS2	0.00	11.91	1.76	0.40
	LOS3	9.04	14.35	1.95	0.08
	LOS4	1.23	12.35	2.28	0.26
	LOS5	0.00	5.62	1.56	0.31
	LOS6	1.64	10.56	1.74	0.31
	LOS7	3.05	25.52	1.53	0.34
	LOS8	0.00	11.84	1.47	0.27
	LOS9	0.00	32.97	2.00	0.36
	LOS10	1.26	20.85	0.53	0.06
Promedio		1.99	14.87	1.64	0.23
Desviación típica		3.99	8.97	0.39	0.12

Tabla 5. Parámetros de canal del modelo SV en cada banda de las posiciones LOS obtenidos con algoritmo de Corrigan.

Banda	Posición	γ	Γ	λ	Λ
B1	NLOS1	0.00	18.51	2.69	0.38
	NLOS2	1.17	11.83	2.01	0.20
	NLOS3	2.14	17.80	2.16	0.25
	NLOS4	0.00	7.31	2.43	0.27

Banda	Posición	γ	Γ	λ	Λ
B2	NLOS1	0.00	14.28	2.69	0.27
	NLOS2	0.75	10.68	2.62	0.34
	NLOS3	3.50	11.93	2.37	0.24
	NLOS4	1.65	6.85	2.58	0.31
B3	NLOS1	3.84	13.83	2.16	0.27
	NLOS2	1.43	17.73	2.01	0.15
	NLOS3	1.16	18.28	1.84	0.21
	NLOS4	4.60	10.61	2.73	0.37
B4	NLOS1	0.67	14.32	2.12	0.20
	NLOS2	0.00	12.81	2.62	0.23
	NLOS3	1.89	9.06	3.02	0.36
	NLOS4	1.46	8.13	2.65	0.25
Promedio		1.52	12.75	2.42	0.27
Desviación típica		1.37	3.79	0.32	0.07

Tabla 6. Parámetros de canal del modelo SV en cada banda de las posiciones NLOS obtenidos con algoritmo de Corrigan.

En definitiva, aplicando el algoritmo de mínimos cuadrados a los datos de canal y de SAGE no se consigue deducir un número de clústeres óptimo ya que los resultados muestran que hay disparidad en la cantidad de clústeres finales. Por esta razón se hace imprescindible el uso de información adicional en la clusterización. Gracias a que SAGE permite obtener los ángulos de llegada de las contribuciones, se comprobará la idoneidad de utilizar esta información cuando se complementa con la clusterización de la amplitud y el retardo (PDP).

5.2 Clusterización por separado: Ángulos y retardo.

Como resultado del apartado anterior, se tiene que la información de potencia y retardo no es suficiente para caracterizar completamente el canal, ya que el número de clústeres óptimos presentan distribuciones irregulares, existiendo sistemáticamente un clúster principal que aglutina a la gran mayoría de contribuciones si utilizamos datos del canal o un número muy elevado de clústeres, alguno de ellos con un número muy pequeño de contribuciones, si se utilizan datos de SAGE. Por tanto, este algoritmo tendería a no ofrecer buenos resultados. Entonces, a partir de la herramienta SAGE, es posible ampliar el estudio de caracterización con información adicional sobre los ángulos. En estudios anteriores se ha demostrado que los ángulos de llegada son especialmente importante en escenarios de interior o *indoor* [12].

El procedimiento será el siguiente: se clusterizará con el algoritmo de Corrigan los datos de ángulo de llegada en función de su retardo, sustituyendo así los valores de amplitud de la señal recibida. Una vez obtenidos los resultados de la clusterización, se combinará con los resultados del apartado anterior 5.1 donde se clusterizaba en función de la amplitud.

En las siguientes figuras se recogen los resultados obtenidos de aplicar Corrigan a los ángulos de elevación (180°) y acimut (360°) extraídos en las posiciones LOS en las bandas B1, B2, B3 y B4.

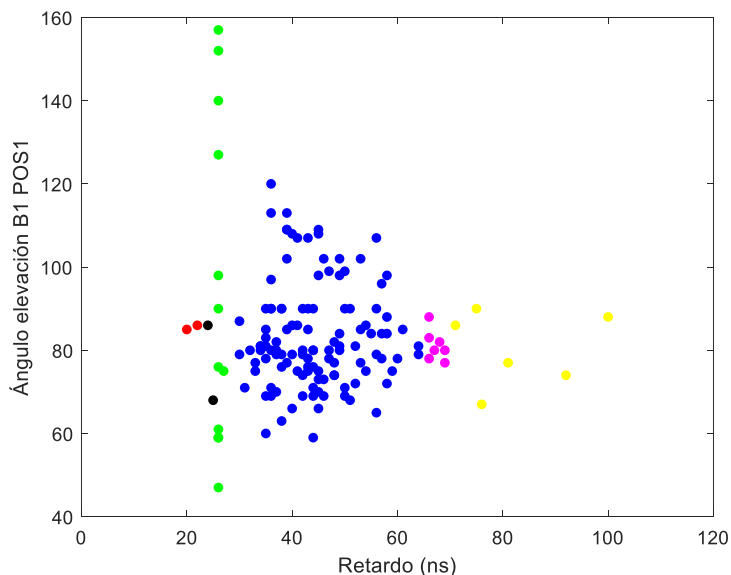


Figura 21. Clusterización con Corrigan de los datos de ángulo de elevación de SAGE para la primera posición LOS en B1.

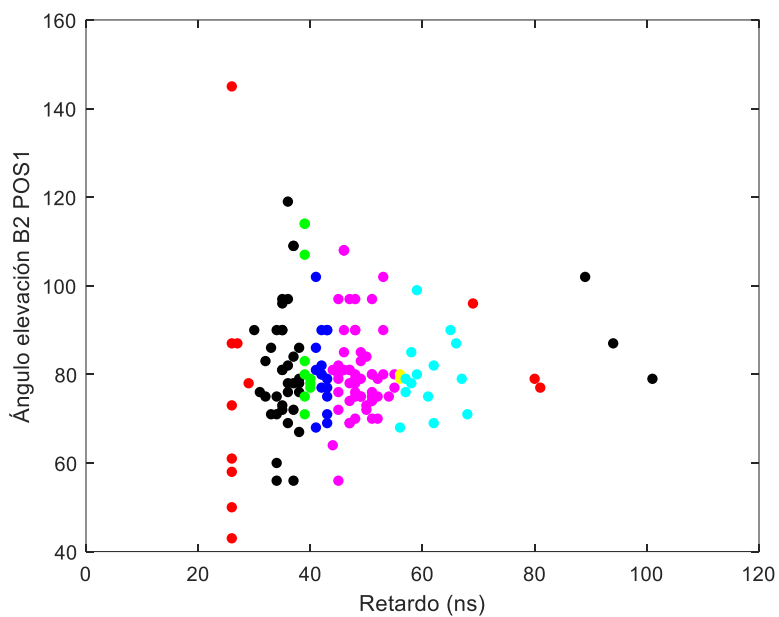


Figura 22. Clusterización con Corrigan de los datos de ángulo de elevación de SAGE para la primera posición LOS en B2.

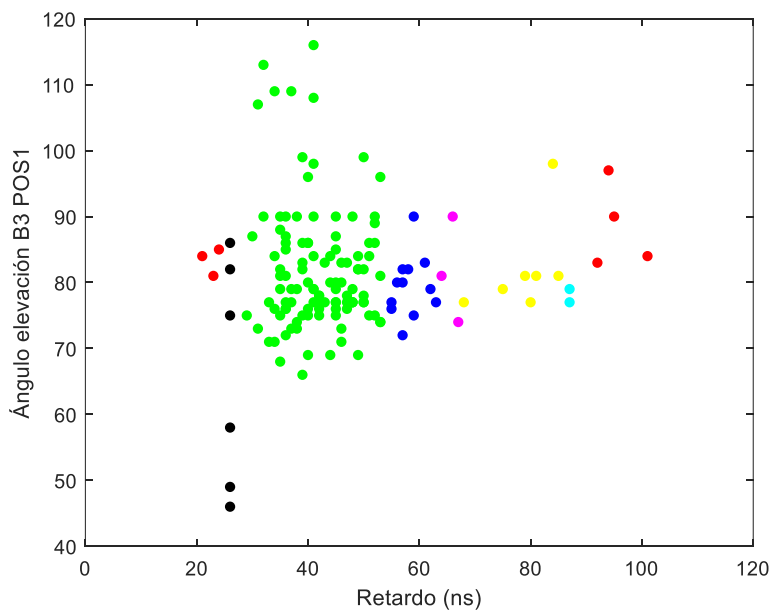


Figura 23. Clusterización con Corrigan de los datos de ángulo de elevación de SAGE para la primera posición LOS en B3.

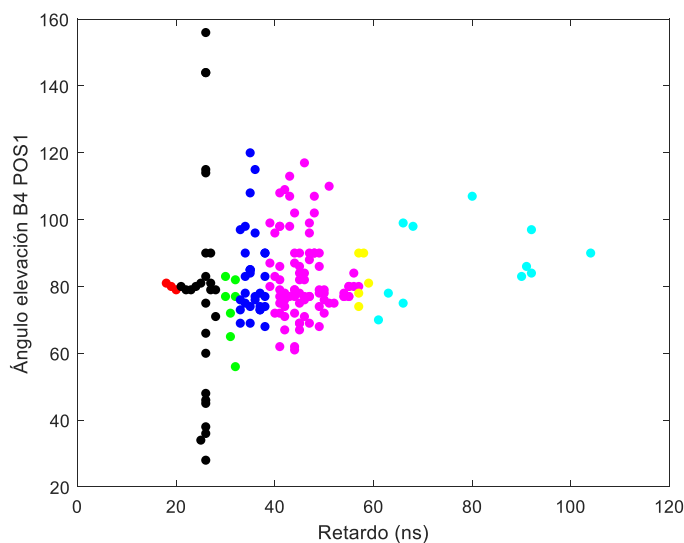


Figura 24. Clusterización con Corrigan de los datos de ángulo de elevación de SAGE para la primera posición LOS en B4.

De los datos de ángulo de elevación cabe decir que el algoritmo SAGE sobreestima las contribuciones que llegan con ángulo de 90° . Para ello se debe realizar un proceso de filtrado con el fin de seleccionar las contribuciones que no estén repetidas. Se deberá escoger todas aquellas contribuciones y prescindir de aquellas que estén muy cerca entre ellas, incluso adyacentes en el tiempo de llegada. Los resultados indican que el número de clústeres es similar al obtenido cuando se empleaba información de amplitud y retardo. A continuación, se clusterizan los datos de ángulo de acimut.

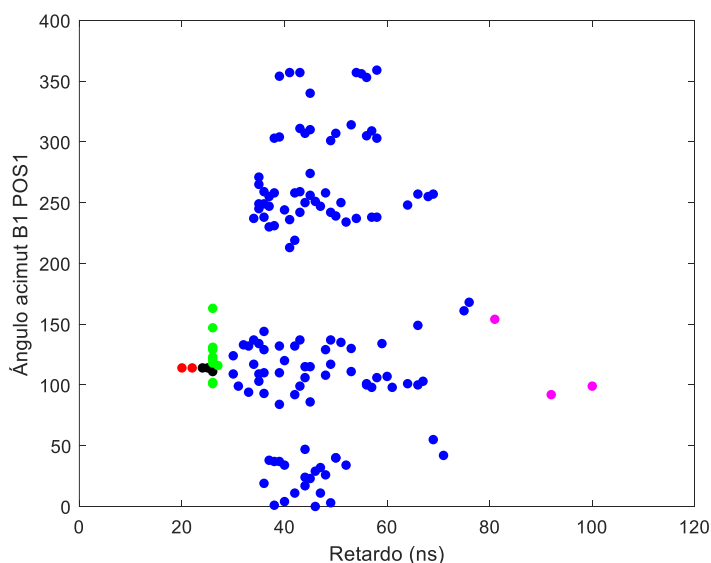


Figura 25. Clusterización con Corrigan de los datos de ángulo de acimut de SAGE para la primera posición LOS en B1.

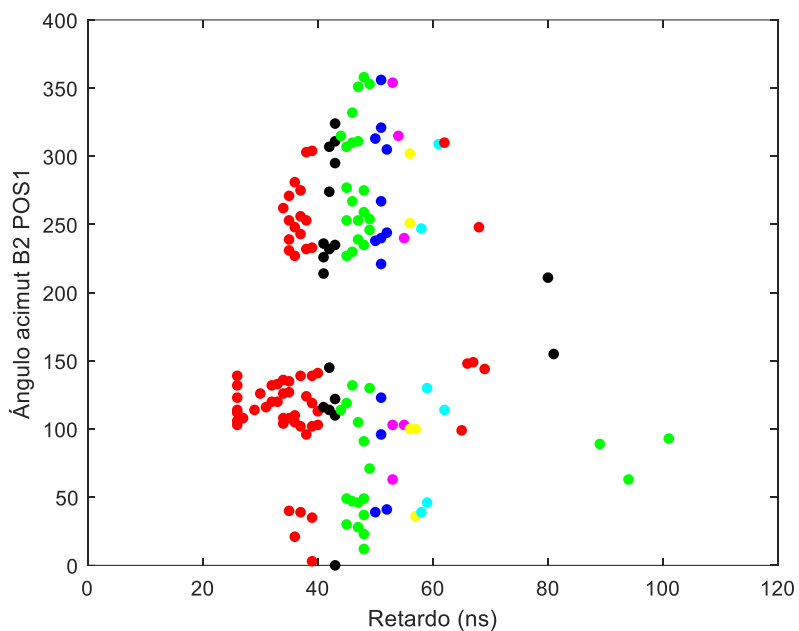


Figura 26. Clusterización con Corrigan de los datos de ángulo de acimut de SAGE para la primera posición LOS en B2.

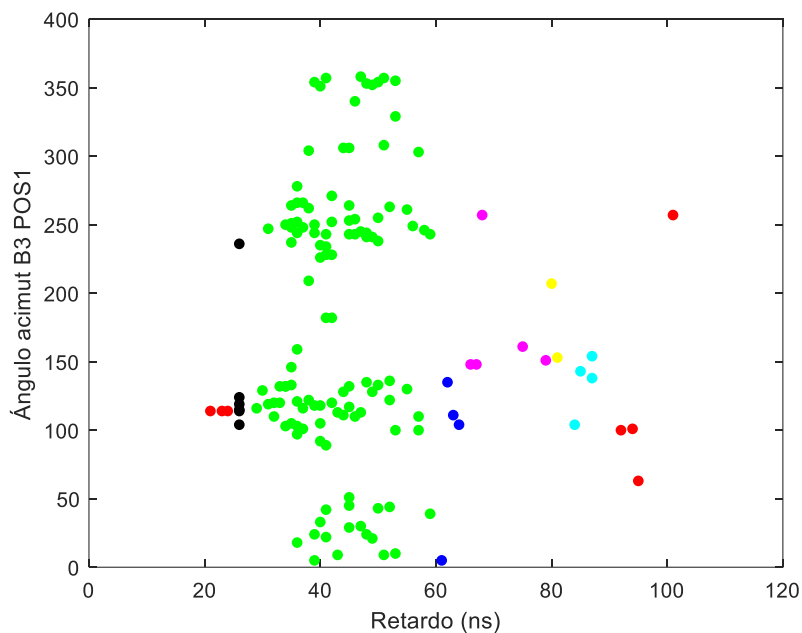


Figura 27. Clusterización con Corrigan de los datos de ángulo de acimut de SAGE para la primera posición LOS en B3.

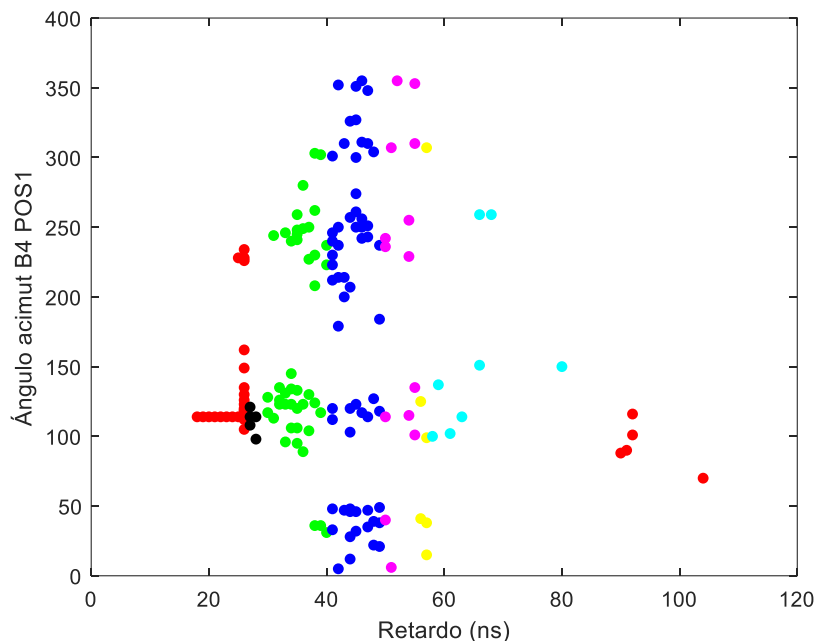


Figura 28. Clusterización con Corrigan de los datos de ángulo de acimut de SAGE para la primera posición LOS en B4.

A partir de la clusterización con ángulos se obtiene una conclusión evidente. Los ángulos aportan información adicional en la clusterización. En posiciones LOS queda demostrado que la información de ángulos refuerza la clusterización con datos de amplitud ya que el número de clústeres óptimos es mayor que en los casos donde se utilizaba información de canal (URA). Por otro lado, también demuestra que un mismo algoritmo sea capaz de clasificar ciertas contribuciones en distintos clústeres cuando se emplea información del ángulo de elevación o acimut con respecto a la amplitud. Por tanto, queda determinar en qué medida afecta que una contribución pertenezca a un clúster cuando se emplea la variable amplitud en la clusterización y pertenece a otro clúster cuando se emplea el ángulo de elevación ó acimut.

Para tratar de determinarlo se propone combinar los resultados de todas las clusterizaciones para cada banda y posición. De forma que estas contribuciones que no pertenecen a un clúster concreto serán consideradas como contribuciones con características ‘híbridas’. Por tanto, si se trata de superponer los resultados de la clusterización cuando se emplea información de amplitud, ángulos y retardo, se puede observar cómo los clústeres se solapan entre ellos.

Para determinar el solape se implementará un algoritmo que identifique las contribuciones que forman parte de varios clústeres a la vez. Para ello se deben comprender qué es un solape y cómo se pueden distinguir.

5.3 Algoritmo de identificación de solapes

A partir de los resultados obtenidos de la clusterización con amplitud y retardo, se puede combinar con los resultados de la clusterización de las direcciones de llegada. En la Figura 29 se representa la clusterización del PDP para la primera posición LOS en B1, junto con las pendientes calculadas de decaimiento de los estadísticos que describe el modelo de SV. Identificar solapes entre las gráficas será de ayuda a la hora de agrupar aquellas contribuciones que no sean fácilmente clasificables.

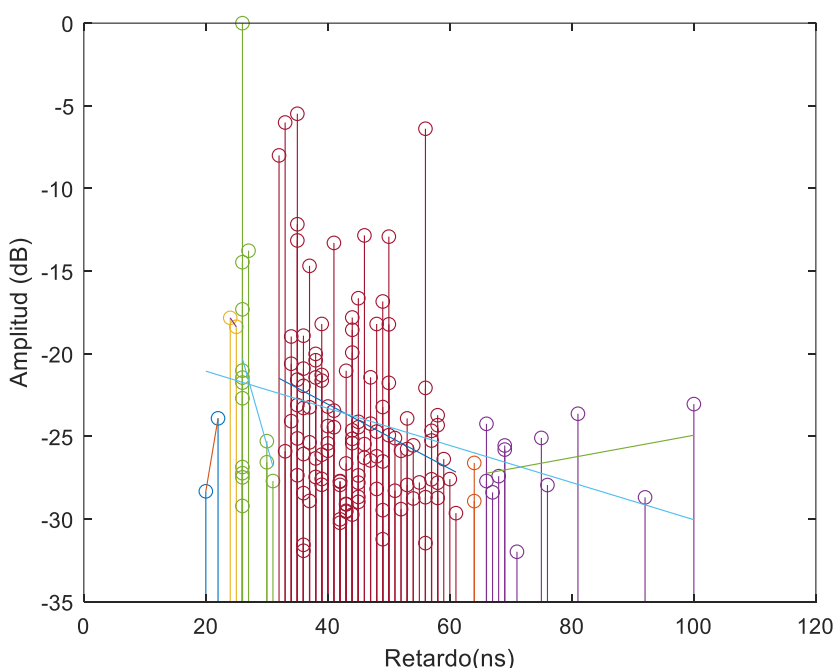


Figura 29. Clusterización con Corrigan del PDP en LOS1 B1.

En primer lugar, los solapes aparecen cuando una contribución se clasifica en dos clústeres a la vez, existiendo contribuciones que pertenecen a un clúster cuando se utiliza la variable amplitud y a otro clúster si se emplea la información de dirección de llegada. En tal caso, se considerará que poseen propiedades compartidas entre ambos grupos y formarán un clúster nuevo constituido por todas aquellas contribuciones solapadas.

La Figura 30 ejemplifica visualmente un solape. Existe un dato que pertenece tanto al clúster verde como al clúster azul. Por tanto, hay un conflicto entre dos clústeres para definir una misma muestra.

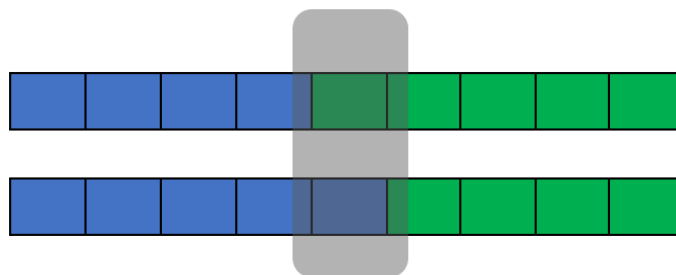


Figura 30. Ejemplo de solape 1.

El análisis de los clústeres nos obliga a introducir una tasa de solape, la cual indica el porcentaje de muestras que entran en conflicto con otros clústeres. En este ejemplo la tasa de solape sería de $1/9$.

En la Figura 31 se propone otra casuística de solape. En rojo se representa un nuevo clúster, indicando que existen diferencias tanto desde el punto de vista de la amplitud como del ángulo con respecto al clúster anterior. En total tenemos tres clústeres y el solape de una contribución. De forma que la tasa de solape de este segundo ejemplo es $1/10$.

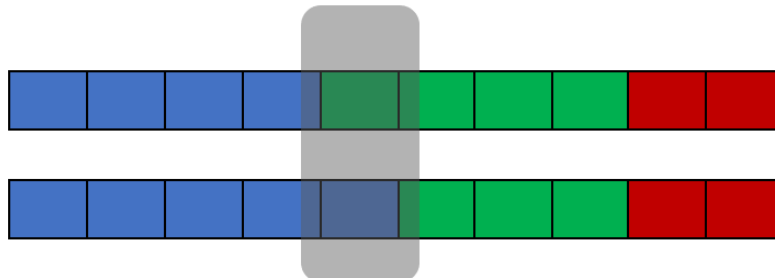


Figura 31. Ejemplo de solape 2.

Por último, y cerrando la explicación de solape, se muestra un nuevo caso donde existe otro solape entre clústeres. En este ejemplo, la contribución 9 comparte propiedades del clúster verde y propiedades nuevas del clúster rojo. La tasa de solape será de $2/10$. Y el número total de clústeres será cinco: tres clústeres predominantes (azul, verde y rojo) y dos clústeres nuevos nacidos del solape.

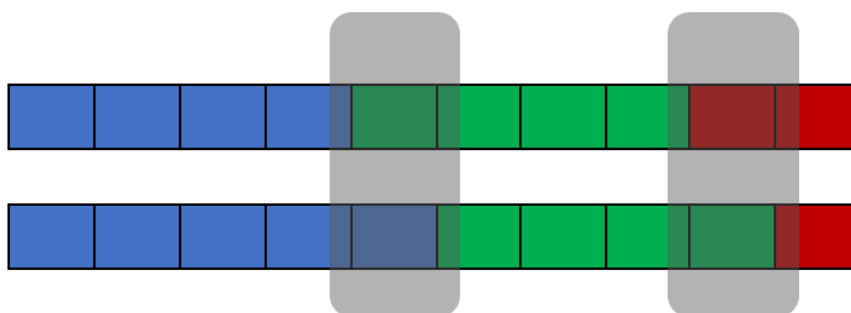


Figura 32. Ejemplo de solape 3.

A partir de los resultados de la clusterización del algoritmo de Corrigan con medidas de amplitud y de ángulos en las cuatro bandas de las primeras posiciones LOS y NLOS, se tienen las tasas de solape recogidas en las Tablas 7 y 8.

LOS	Banda 1	Banda 2	Banda 3	Banda 4
Amplitud- Ángulo 1	5,52%	10,14%	6,16%	7,14%
Amplitud- Ángulo 2	3,45%	10,81%	5,48%	8,93%

Tabla 7. Tasa de solape en las bandas de la primera posición LOS.

NLOS	Banda 1	Banda 2	Banda 3	Banda 4
Amplitud- Ángulo 1	3,61%	15,24%	1,17%	11,46%
Amplitud- Ángulo 2	2,41%	12,8%	4,09%	12,1%

Tabla 8. Tasa de solape en las bandas de la primera posición NLOS.

El solape demuestra que tomar como referencia la amplitud en la clusterización no es razón suficiente para describir el canal radio ya que la clusterización con información de ángulos permite extraer otras conclusiones sobre la agrupación de las contribuciones multicamino. Si se combinan por separado los resultados, se puede concluir que el número de clústeres aumenta significativamente ya que aparecen clústeres nuevos nacidos de las MPC que no pertenecen siempre al mismo clúster.

En las Tablas 9 y 10 se recogen el número final de grupos que surgen de la puesta en común de los resultados de clusterizar por separado los datos extraídos de SAGE de amplitud y ángulos de elevación y acimut tanto en la primera posición LOS como NLOS.

LOS	B1	B2	B3	B4
Amplitud-acimut	16	21	12	17
Amplitud- elevación	10	19	13	18

Tabla 9. Número total de clústeres en las bandas de la primera posición LOS.

NLOS	B1	B2	B3	B4
Amplitud- acimut	7	23	9	22
Amplitud- elevación	9	21	10	22

Tabla 10. Número total de clústeres en las bandas de la primera posición NLOS.

Como por separado aumenta enormemente el número de clústeres, siendo una clasificación ineficiente de las contribuciones, se probará a estudiar si es mejor utilizar de forma conjunta los datos de amplitud, retardo y ángulo de llegada.

Capítulo 6. APLICACIÓN DE MACHINE LEARNING A LA CLUSTERIZACIÓN

En este apartado se comprobará que la forma de mejorar la clusterización es utilizando la información de la amplitud, retardo y ángulos conjuntamente. Para ello es necesario la implementación de modelos de clusterización más sofisticados que el algoritmo de Corrigan, ya que éste sólo permite clusterizar en función de dos variables. Entre los algoritmos más sofisticados se encuentra el modelo de clusterización basado en Machine Learning no supervisado llamado Kmeans. Existen otros algoritmos, como por ejemplo los modelos jerárquicos, los modelos basados en densidad o el Knn (*K nearest neighbor*). A continuación, se detallará la implementación del algoritmo Kmeans con los datos de SAGE.

6.1 Algoritmo Kmeans

Generalmente los clústeres pueden ser definidos haciendo uso de una mera observación de los datos, pero esta tarea se complica cuando el número de observaciones es muy elevado, o la estructura de los mismos imposibilita la interpretación debido a su complejidad, como es el caso de representaciones 3D. Cuando se tratan datos que no pueden ser supervisados, como es el caso de los perfiles de potencia, donde las medidas se obtienen de un canal que no es estático, hace necesario el uso de algoritmos que validen el número de clústeres apropiados y clasifiquen los datos a través de criterios o heurísticas, que hagan uso, generalmente, de las distancias entre los datos [9]. Por tanto, la validación es esencial para conseguir una división eficiente de los datos cuando se desconoce el número de clústeres.

La validación del número de clústeres puede llevarse a cabo a través de dos métodos, external validation y relative validation [10]. La validación externa es utilizada cuando los clústeres obtenidos pueden ser comparados con respecto a una estructura predeterminada, lo que vendría de la mano de una clusterización ‘supervisada’ y por tanto no podría ser empleada en este estudio. Por otro lado, la validación relativa trata de enfrenar los resultados de las clusterizaciones entre ellos con el fin de obtener una puntuación, lo cual permitiría elegir el número k de clústeres más eficiente.

Proliferan numerosos métodos de validación relativa en la literatura, por ejemplo, *Davies-Bouldin*, *Calinski-Harabasz*, *Silhouette*, *Sum-of-squares*...etc. Todos ellos tienen como finalidad encontrar el número apropiado de clústeres en los que se clasifican todas las contribuciones. Los

estudios de estos criterios [10], resaltaron que los más efectivos para predecir el valor apropiado de k , fueron Calinski-Harabasz y Silhouette. Ambos ofrecían buenos resultados cuando se trataban datos cuyos clústeres eran desconocidos.

Por su parte el algoritmo Kmeans tiene la función de identificar y asignar a cada contribución un clúster en función de la similitud que se encuentra con otras contribuciones. Para determinar que el número de clústeres es el óptimo, se deberá validar con los criterios de Calinski-Harabasz, Silhouette, Davies-Bouldin, etc. Como algoritmo de Machine Learning no supervisado, no requiere un entrenamiento previo, y por tanto sus resultados dependerán de criterios internos o heurísticas. Estas heurísticas emplean la distancia como variable de decisión.

Los procesos internos del algoritmo Kmeans se pueden dividir en cuatro, explicados a continuación.

6.1.1 Inicialización

En primer lugar, la inicialización es el proceso por el cual se escoge un número k predefinido de centroides o semillas. Estos centroides indican el número de clústeres que debe haber. Para ello se podrá utilizar una asignación aleatoria entre la nube de puntos de los datos o también pueden aplicarse criterios que permitan identificar las mejores ubicaciones para los centroides. En Matlab pueden utilizarse 4 criterios de inicialización ó 'Start':

Clúster	Escoge aleatoriamente una submuestra equivalente al 10% de todos los datos a clusterizar y escoge un número k de ellas aleatoriamente como ubicaciones de inicio de los centroides. Si k es mayor que el número de datos en la submuestra, se aplica 'sample'.
Plus	Mediante el algoritmo k -medias++ se seleccionan k semillas entre la totalidad de los datos. (Criterio que se adopta por defecto gracias a la mejora del tiempo de ejecución que introduce y la calidad de la solución final).
Sample	Selecciona aleatoriamente un número k de entre todos los datos como ubicación de los centroides.
uniform	Selecciona k observaciones de forma aleatoria e uniformemente dentro de todos los datos.

Tabla 11. Tabla resumen de los métodos de inicialización del algoritmo Kmeans.

6.1.2 Cálculo de distancias

El segundo proceso será la asignación de los datos al centroide más cercano. Para ello se deberá calcular la distancia entre ellos. Existen distintas fórmulas matemáticas que calculan la distancia dentro de un espacio p-dimensional, donde p indica la dimensión. La Tabla 12 resume las métricas de ‘distance’ que pueden ser implementadas en Matlab [11]. Existe más información relacionada con el cálculo de distancias, para ello se puede emplear la documentación que ofrece el soporte de Matlab sobre la función Kmeans y los cálculos internos que tienen lugar en la aplicación de la función.

Squeclidean	Distancia euclidiana cuadrada.	$d(x, c) = (x - c) \cdot (x - c)'$
Cityblock	Suma de diferencias absolutas.	$d(x, c) = \sum_{j=1}^P x_j - c_j $
Cosine	Uno menos el coseno del ángulo que forma el vector entre los puntos y el centroide.	$d(x, c) = 1 - \frac{xc'}{\sqrt{(xx') \cdot (cc')}}$
Correlation	Métrica más compleja que emplea la correlación de la muestra entre los puntos, normalizando los puntos con una media de cero y desviación típica la unidad.	
Hamming	Aplicable únicamente a datos binarios, no aplicable a información del canal y por tanto desechada del análisis.	

Tabla 12. Tabla resumen con los tipos de cálculo de distancia en el algoritmo Kmeans.

6.1.3 Iteraciones y repeticiones

El tercer proceso es un cálculo iterativo, donde se modifica la ubicación de los centroides con la finalidad de encontrar nuevas ubicaciones que mejoren la clasificación de los clústeres. Para ello se utiliza la opción ‘MaxIter’ especificando el número de iteraciones máximas que se pueden realizar. Puede ocurrir que el número de iteraciones indicado no sean suficientes para encontrar una solución que converja, en tal caso se debe indicar al algoritmo que aumente el número de iteraciones, lo cual puede incrementar el coste computacional enormemente. En cualquier caso, es conveniente que el algoritmo de Kmeans encuentre una solución convergente sin superar el número de iteraciones máximas.

No existe un número de iteraciones concreto de referencia que se pueda utilizar, simplemente se deduce de las ejecuciones y la calidad del resultado obtenido.

Por último, el algoritmo tratará de ‘replicar’ los tres procesos anteriores desde el principio (asignación aleatoria de la ubicación de los clústeres, asignación de datos a los clústeres y modificaciones en la ubicación de los centroides) con la finalidad de que el nuevo resultado coincida con los anteriores, de esta forma se considerará que la solución converge y se da como válida.

Puede darse que no haga falta ejecutar todas las iteraciones del algoritmo Kmeans indicados en la función debido a que se ha alcanzado el mismo resultado en dos iteraciones seguidas, parando la ejecución con el fin de optimizar el coste computacional. Se debe tener en cuenta, además, que cuantos más datos se desea clusterizar más carga computacional se debe asumir, pudiendo afectar directamente al número de iteraciones y repeticiones que se deben realizar con el fin de encontrar la clasificación óptima de las contribuciones.

Por ejemplo, se puede añadir la opción ‘Display’, dentro de la función ‘Kmeans’, con el fin de poder visualizar por pantalla los resultados alcanzados en cada iteración y repetición junto con la suma total de distancias conseguidas. En la Figura 33 se muestra una captura que servirá como ejemplo.

```
Replicate 1, 4 iterations, total sum of distances = 145926.  
Replicate 2, 2 iterations, total sum of distances = 145926.  
Replicate 3, 2 iterations, total sum of distances = 145926.  
Replicate 4, 8 iterations, total sum of distances = 229920.  
Replicate 5, 5 iterations, total sum of distances = 145926.  
Replicate 6, 2 iterations, total sum of distances = 145926.  
Replicate 7, 7 iterations, total sum of distances = 229920.  
Replicate 8, 7 iterations, total sum of distances = 229883.  
Replicate 9, 3 iterations, total sum of distances = 145926.  
Replicate 10, 2 iterations, total sum of distances = 145926.  
Best total sum of distances = 145926
```

Figura 33. Ejemplo de proceso de iteraciones y repeticiones. Captura de Matlab.

6.1.4 Proceso de validación.

La validación es un proceso complejo que integra todos los resultados obtenidos previamente cuando se clusteriza a partir de k clústeres de forma aislada. Por ello, el objetivo de la validación es hallar el número k óptimo de entre todos los k posibles.

El criterio de evaluación empleado para encontrar el número apropiado de clústeres será Calinski-Harabasz (CH). Este criterio es uno de los más exitosos (Baarsch & Emre, 2012) además de que funciona correctamente cuando se utiliza el PDP ideal. Su funcionamiento se basa en el cálculo de distancias inter-clúster, (BCSM, Between Cluster Scatter Matrix) obtenidas a partir de la suma de los cuadrados entre el centro del clúster y los centroides.

Todo ello ponderado por el tamaño del clúster. También se incluyen distancias intra-clúster, (WCSM, Within Cluster Scatter Matrix) obtenidas a partir de la suma de los cuadrados de las distancias entre el verdadero centro de cada clúster y los puntos que lo conforman. Una vez calculadas las distancias se normaliza por el factor $(N - K) / (K - 1)$. El cual hará que vayan disminuyendo las puntuaciones conforme el número k de clústeres aumenta. Se muestra detalladamente en (4)

$$CH = \left(\frac{BCSM}{K - 1} \right) / \left(\frac{WCSM}{N - K} \right) = \left(\frac{\sum_{k=1}^K n_k \cdot |c_k - c|^2}{K - 1} \right) / \left(\frac{\sum_{k=1}^k \sum_{i=1}^{n_k} |d_i - c_k|^2}{N - K} \right) \quad (4)$$

La validación usando el criterio de Calinski-Harabasz buscará la maximización de la distancia inter-clúster y la minimización de la distancia intra-clúster, de forma que se maximizará el coeficiente CH obtenido como resultado de la función de las distancias para que la clusterización converja en una solución óptima. Se considerará una buena clusterización cuando los puntos que presenten similitudes en sus características de forma cuantificable, por ejemplo, a partir de las distancias euclídeas, estén agrupados en el mismo clúster. De la misma forma que cuando dos puntos estén alejados el uno del otro, éstos estarán clasificados en clústeres distintos.

En la Figura 34 se muestra el resultado del proceso de optimización de la función Kmeans con el fin de encontrar la combinación de operaciones internas que consiguen clusterizar el PDP ideal de forma correcta. El resultado de la optimización indica que se debe emplear el el método de validación CH, la inicialización 'Sample' y como cálculo de la distancia se utilizará 'Squeclidean'.

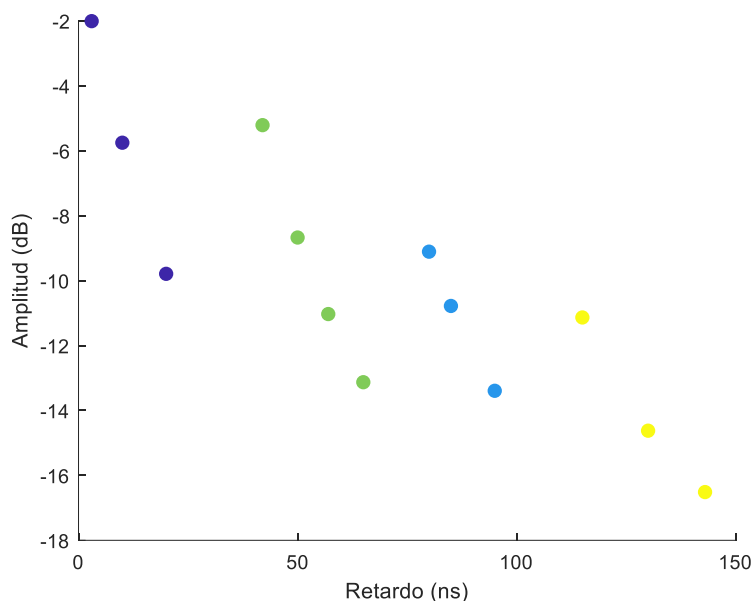


Figura 34. Clusterización con el algoritmo Kmeans del PDP sintético ideal.

A continuación, se muestran otros casos de validación. En la Figura 35 y 36, se muestra un ejemplo de validación donde no se alcanza la convergencia y en el segundo, sí. Resaltando que el principal problema de este criterio de validación surge al incrementar el número de clústeres k , originando que la distancia intra-cluster (denominador) se reduzca tanto que eclipse por completo el efecto de la reducción de la distancia inter-clúster (numerador), aumentando el coeficiente CH. Esto ocurre cuando los datos medidos están extremadamente cerca los unos de los otros siendo prácticamente imposible discernir los clústeres a primera vista.

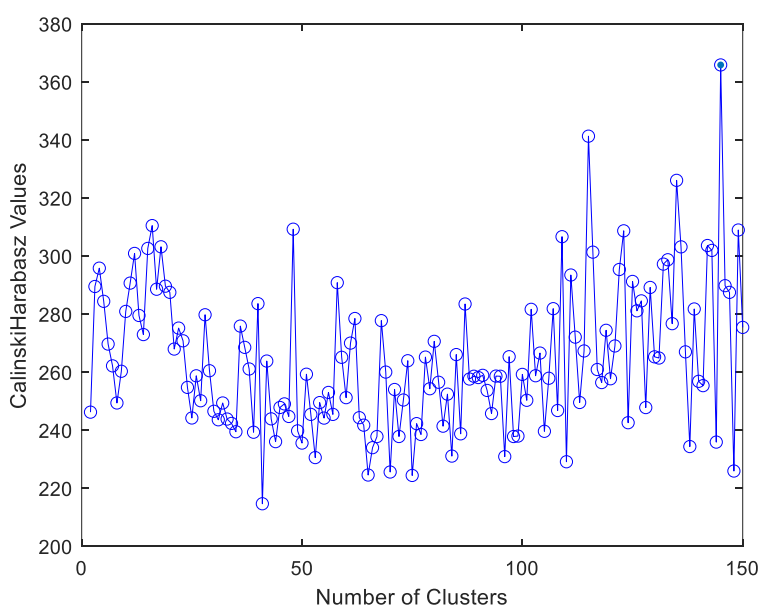


Figura 35. Proceso de validación con el método de Calinski-Harabasz cuyo resultado no es convergente.

En la Figura 34 Se puede observar que la validación no converge hacia un número de clústeres óptimo ya que conforme el orden k de clústeres aumenta, el coeficiente CH también aumenta. Por tanto, la solución arrojada no será considerada como eficiente.

Por otro lado, en la figura 35 se muestra un proceso de validación donde se maximiza el coeficiente, siendo óptimo cuando el orden k de clústeres es cuatro. En este caso la validación converge ya que se ha encontrado un máximo global.

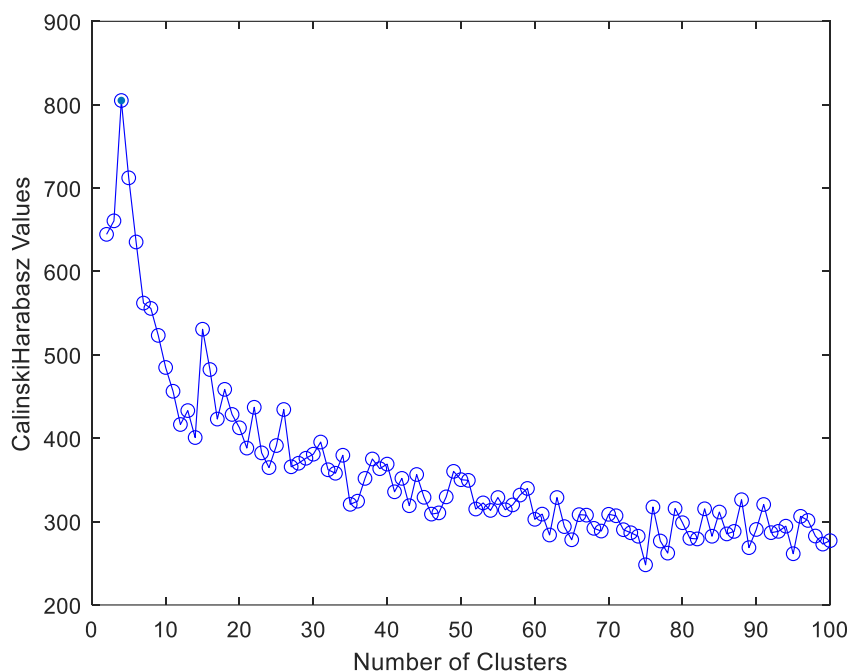


Figura 36. Proceso de validación con el método de Calinski-Harabasz cuyo resultado sí es convergente.

También existen otros métodos de validación como el Silhouette que relaciona distancias inter-clúster e intra-clúster, aunque emplea una normalización que conforme mejores son los resultados de la clusterización, la función de validación se acerca asintóticamente hacia el 1. Por otro lado, cabe destacar el método de validación de Davies-Bouldin, también conocido como el criterio del codo, cuya función de validación otorga más importancia a la minimización de la distancia intra-clúster.

De todos los métodos de validación descritos, el que mejor resultados ha dado de forma general con los datos de canal es el de Calinski-Harabasz. De forma que será el que se empleará para la validación de los datos de SAGE de amplitud, retardo, ángulo de acimut y de elevación.

6.1.5 Implementación de Kmeans en Matlab

Kmeans está implementado en una función dentro de la Toolbox de Matlab de Machine Learning. Dicha función recoge todas las especificaciones necesarias para el proceso de clusterización, por ejemplo: el criterio de cálculo de distancias, la identificación de los centroides iniciales ó el número de iteraciones que realizará el algoritmo hasta encontrar la solución óptima de la clusterización.

Como principales datos de entrada está la matriz de datos a clusterizar, y el orden 'k' de clústeres a emplear. Además, también se puede optimizar la clasificación de los datos empleando los criterios de distancias entre clústeres, de asignación de semillas o centroides iniciales y un número máximo de iteraciones y repeticiones que la función podrá ejecutar.

Por otro lado, la validación requiere de un proceso de comprobación iterativo, por lo que se implementará un bucle que determine en primer lugar la clasificación óptima de las contribuciones con la función Kmeans para después, mediante una función de evaluación índice el orden 'k' de clústeres que optimiza la clasificación de los datos.

Dentro de la función 'kmeans' puede identificarse las siguientes especificaciones para el proceso de clusterización, recogidas en la Tabla 13.

```
[idx,Centroids] = kmeans(matriz_clusterizar, eva.OptimalK, 'MaxIter', 100,  
'Replicates',10, 'Distance', 'sqeuclidean', 'Display', 'final', 'Start', 'sample');
```

Idx: Vector que contiene la clasificación en clústeres de cada dato de forma individual. La asignación corresponde con un número entero positivo desde 1 hasta el orden k.

Centroids: Matriz que contiene las coordenadas de los datos que se corresponden con los centroides de cada clúster.

Identificación	Función
'data'	Matriz de datos de contribuciones de retardo y amplitud medidos por el URA ó la matriz de datos obtenidos con SAGE.
'OptimalK'	Orden óptimo de clústeres obtenido tras la evaluación de todos los posibles ordenes de clusterización en el proceso de validación.
Identificación	Función

'Maxiter'	Número máximo de iteraciones que se realizarán durante el proceso de obtención de la clasificación óptima, en la práctica se han utilizado un número máximo de 100 iteraciones.
'Replicates'	Repeticiones del proceso de iteración, el objetivo es realizar pasadas a la función y con la aleatorización del centroide inicial repetir el proceso de iteración. Se han seleccionado 10 repeticiones.
'Distance'	Se debe indicar de forma explícita el método para calcula la distancia entre los datos. La selección del método lleva consigo la elección de una u otra ecuación matemática, para el cálculo de las distancias de los datos del URA y de SAGE se ha empleado sgeuclidean.
'Start'	Proceso inicial de asignación de centroides. Los mejores resultados se han obtenido con el criterio Sample.
'Display'	Representa por pantalla los resultados de cada iteración y repetición.

Tabla 13. Identificación de los criterios internos de la función Kmeans de Matlab.

Capítulo 7. RESULTADOS

7.1 Clusterización con Kmeans

Kmeans tiene la capacidad de clusterizar más de dos variables, a diferencia del método lineal basado en la línea de tendencia de mínimos cuadrados que sólo podía utilizar dos variables, por ejemplo: amplitud y retardo. De esta forma será posible introducir en forma de matriz los datos de amplitud, retardo, ángulo de acimut y elevación a la vez. Esto es debido a la posibilidad que tiene el algoritmo para calcular distancias en p-dimensiones.

La clusterización en Kmeans se realizará en primer lugar sobre el conjunto de datos de SAGE para las primeras posiciones LOS y NLOS. Una vez clusterizados los datos, serán parametrizados con el fin de obtener los modelos estadísticos SV que caracterizarán el canal.

En la Figura 37 se representan los datos de SAGE en B1 de la posición 1 LOS. Se ha escogido un gráfico 3D donde el eje x representa el ángulo de elevación (0° a 180°), el eje y contiene el ángulo de acimut (0° a 360°), el eje z está el retardo (ns) y, por último, se define el eje w como la amplitud de las contribuciones, de forma que el tamaño de los puntos está directamente relacionado con la amplitud de las MPC.

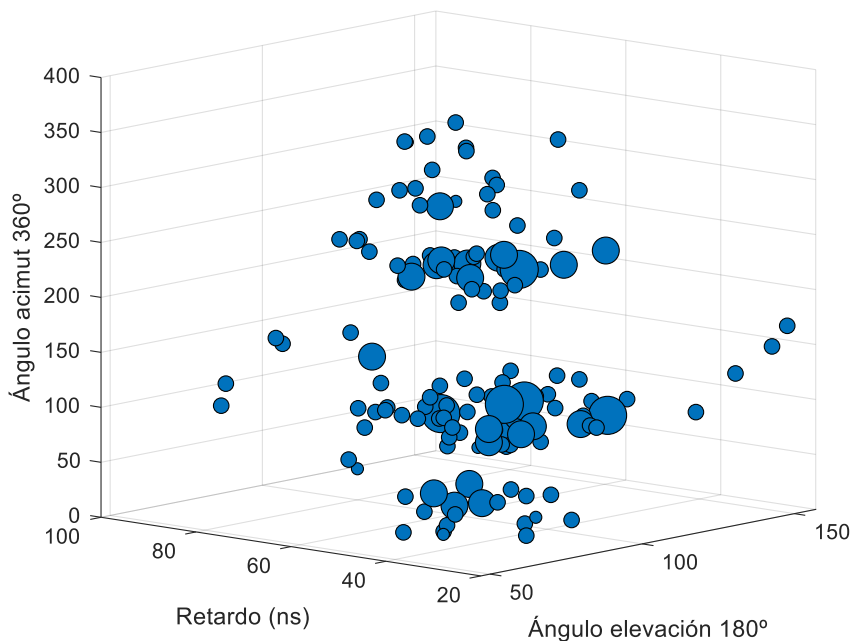


Figura 37. Nube de puntos representados en 3D con el conjunto de variables de SAGE.

En la Figura 38 se muestra el proceso de validación con el método CH. Se puede observar que existe un máximo bien diferenciado cuando el número de clústeres es 4.

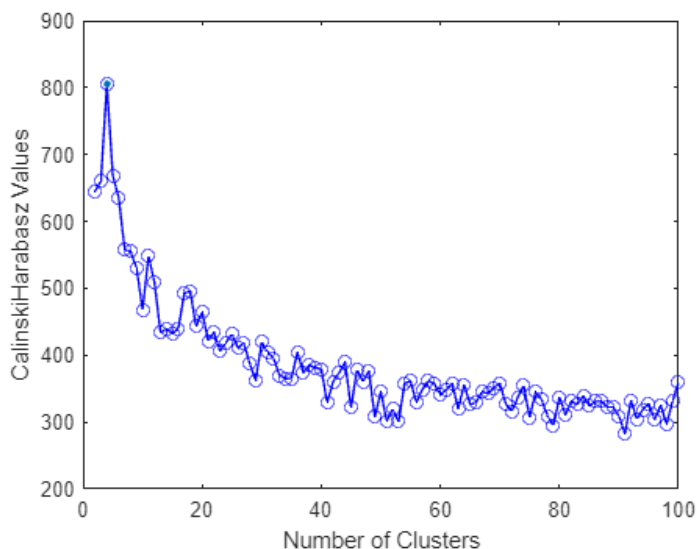


Figura 38. Proceso de validación con el método de Calinski-Harabasz, donde la solución converge en 4 clústeres.

En la Figura 39 se puede identificar el resultado de la clusterización con Kmeans, donde cada clúster tiene asignado un color y se representan las MPC en función del clúster al que pertenecen.

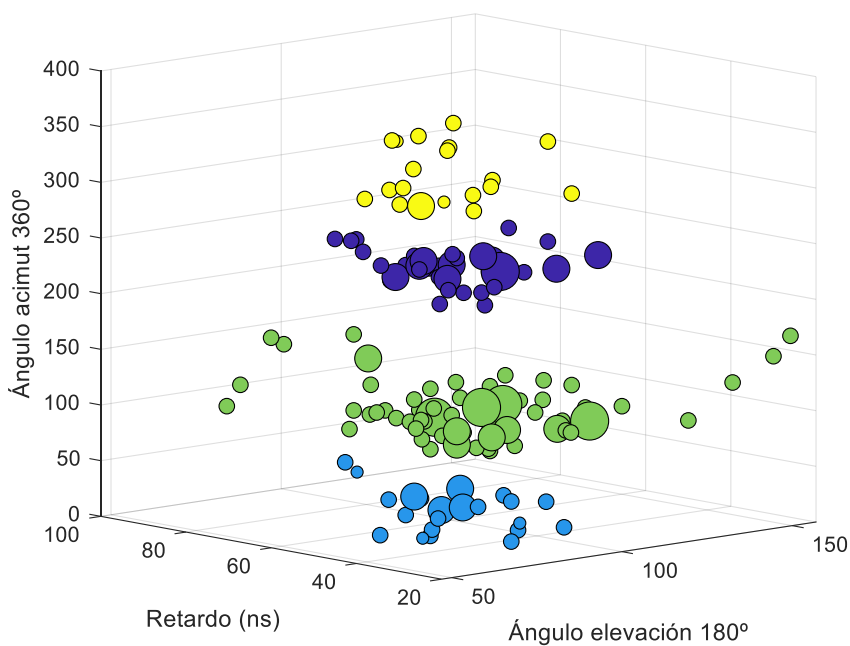


Figura 39. Resultados de la clusterización con Kmeans empleando el conjunto de los datos de SAGE para la primera posición LOS en B1.

Una vez realizada la clusterización de los datos en su conjunto, es útil representar las proyecciones de las variables de canal clusterizadas con respecto al retardo. En la Figura 40 se muestra la proyección de la gráfica 3D con respecto al ángulo de elevación y el retardo.

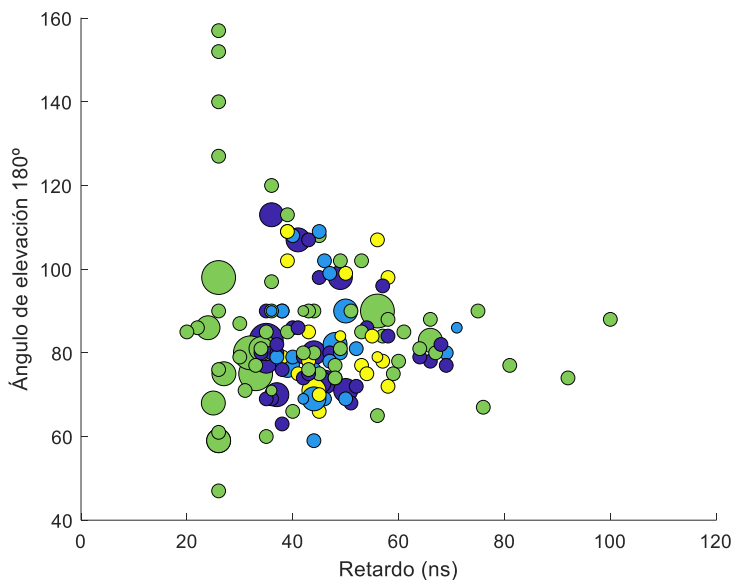


Figura 40. Proyección del ángulo de elevación y el retardo de la figura 38.

En la Figura 41 se representa la proyección de la Figura 39 de la amplitud con respecto al retardo. Esta representación se corresponde con la solución que ofrece Kmeans en la clusterización del PDP del canal para la banda 1 en la primera posición LOS, cuando se ha tenido en cuenta información de direcciones de llegada, amplitud y retardo durante el proceso de clusterización.

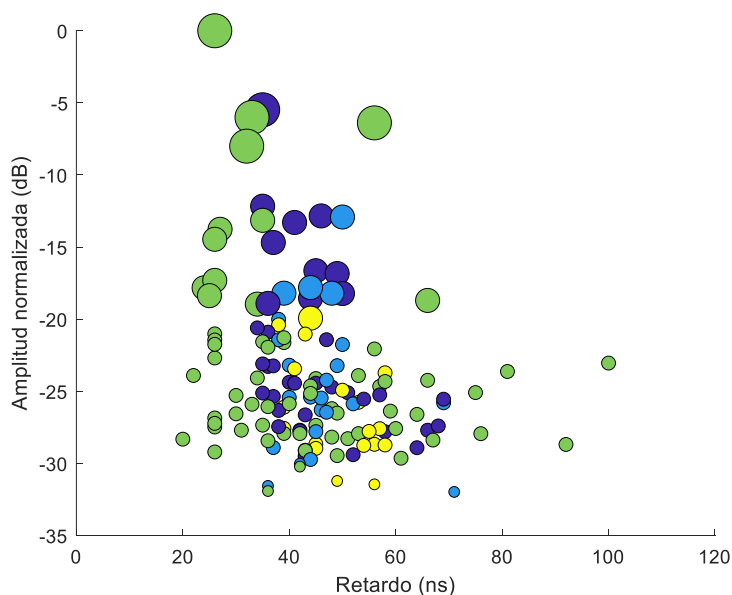


Figura 41. Proyección de la amplitud normalizada y el retardo de la figura 38

En la Tabla 14 se muestra el resultado de la parametrización del PDP de la Figura 41 siguiendo el modelo estadístico SV cuando se tienen en cuenta los datos de SAGE en conjunto y se utiliza el algoritmo Kmeans.

Parámetros Saleh-Valenzuela	γ	Γ	λ	Λ
B1 LOS1	6.87	8.91	1.80	0.04

Tabla 14. Tabla resumen con los resultados de emplear todos los datos de SAGE.

En la Figura 42 se puede determinar que la clusterización toma mayor sentido cuando se emplea el ángulo de acimut y no otras variables como la amplitud o el ángulo de elevación. Se puede concluir que cuando se clusterizan todas las variables de SAGE en su conjunto, el ángulo de acimut es determinante en el proceso ya que los resultados obtenidos pueden ser verificados a simple vista.

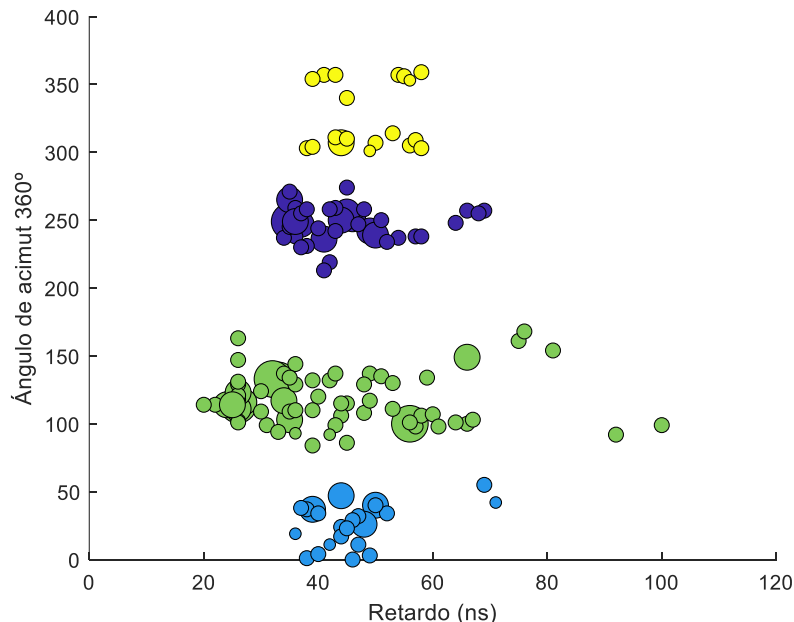


Figura 42. Proyección del ángulo acimut y retardo de la figura 38.

Se realiza el mismo procedimiento en la primera posición NLOS en B1. Se observa que el número de clústeres óptimo es 4, al igual que las posiciones LOS. El resultado se muestra en la Figura 43. También se pueden identificar los principales clústeres que surgen a simple vista en la Figura 44,

aunque en la posición NLOS puede parecer óptimo el uso de un número de clústeres mayor, originado por aquellas contribuciones que están más alejadas de la nube de puntos principal.

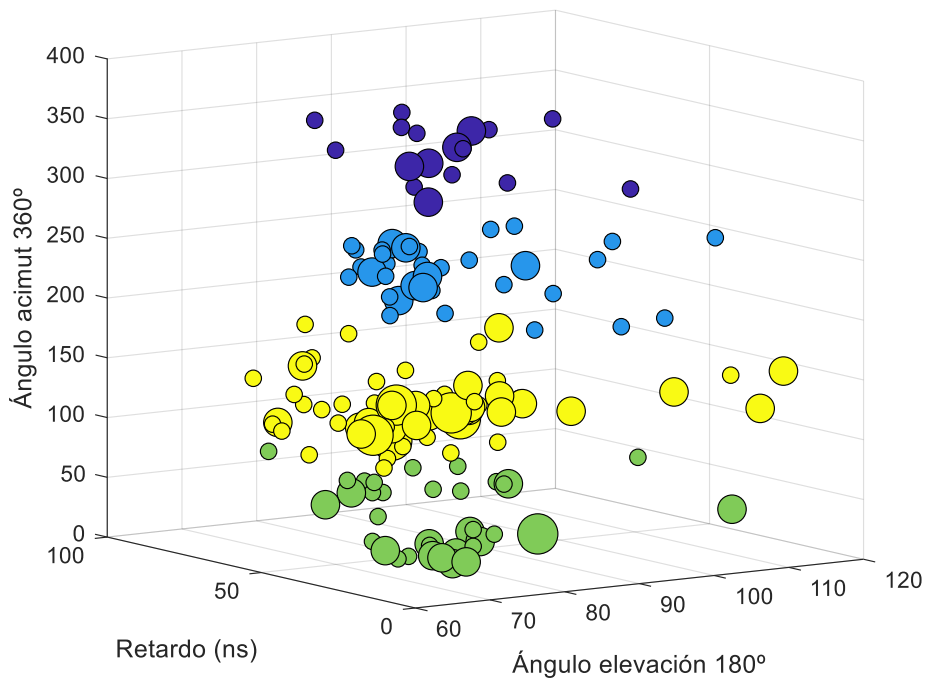


Figura 43. Resultados de la clusterización con Kmeans empleando el conjunto de los datos de SAGE para la primera posición NLOS en B1.

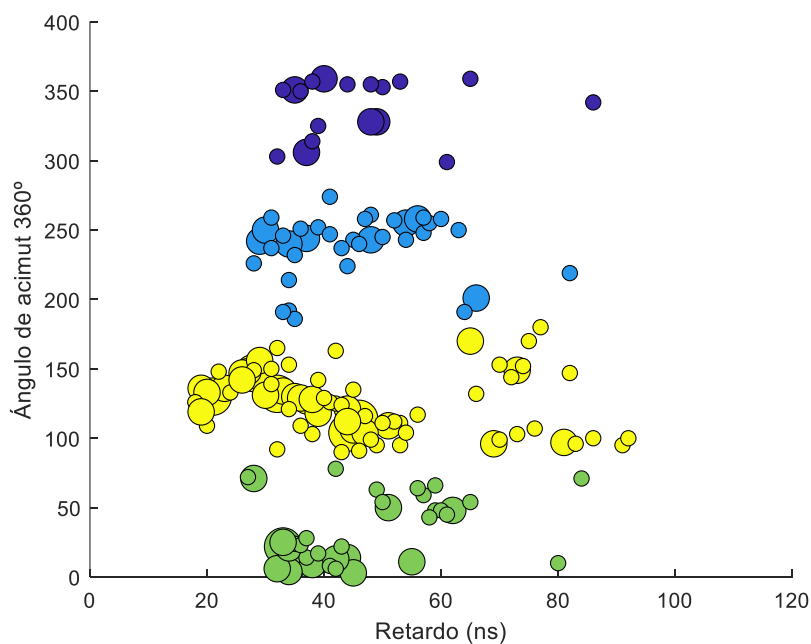


Figura 44. Proyección del ángulo acimut y retardo de la figura 42.

7.2 Modelizado del canal

En este apartado se toma la proyección de la gráfica en 3D con respecto a la amplitud y el retardo. A partir de estos datos y el resultado de la clusterización con Kmeans se calcularán los parámetros que caracterizarán los modelos estadísticos SV. En la Tabla 15 se recogen todos los resultados de la parametrización en cada posición LOS y las bandas medidas: B1, B2, B3 y B4.

Banda	Posición	γ	Γ	λ	Λ
B1	LOS1	6.88	8.90	1.80	0.04
	LOS2	10.75	9.62	1.59	0.03
	LOS3	12.18	12.58	1.70	0.03
	LOS4	13.25	14.27	1.39	0.06
	LOS5	7.31	5.37	1.31	0.05
	LOS6	7.86	12.80	1.67	0.05
	LOS7	25.87	31.93	1.78	0.05
	LOS8	4.54	17.52	1.20	0.50
	LOS9	34.98	46.25	1.94	0.03
	LOS10	10.20	10.12	1.79	0.04
B2	LOS1	7.25	6.86	1.96	0.12
	LOS2	7.68	6.89	1.80	0.67
	LOS3	10.75	9.42	1.31	0.12
	LOS4	6.65	8.22	1.77	0.04
	LOS5	6.81	4.07	2.65	0.07
	LOS6	5.94	8.40	1.49	0.04
	LOS7	10.61	17.22	1.33	0.03
	LOS8	39.55	23.36	1.73	0.38
	LOS9	20.97	26.71	1.80	0.05
	LOS10	11.84	12.62	1.52	0.03

Banda	Posición	γ	Γ	λ	Λ
B3	LOS1	9.28	7.67	1.81	0.04
	LOS2	9.98	8.70	1.72	0.02
	LOS3	20.53	12.64	1.53	0.02
	LOS4	8.94	12.78	1.89	0.05
	LOS5	6.70	5.05	1.42	0.67
	LOS6	14.68	26.02	0.37	0.16
	LOS7	9.44	26.33	1.57	0.10
	LOS8	36.33	22.71	1.24	0.23
	LOS9	23.42	25.07	1.60	0.07
	LOS10	11.78	12.03	2.00	0.02
B4	LOS1	6.98	6.89	1.94	0.09
	LOS2	10.74	11.91	1.76	0.07
	LOS3	28.69	14.35	1.95	0.17
	LOS4	18.59	12.35	2.28	0.07
	LOS5	8.22	5.62	1.56	0.06
	LOS6	17.64	10.56	1.74	0.14
	LOS7	17.46	25.52	1.53	0.07
	LOS8	8.16	11.84	1.47	0.09
	LOS9	12.95	32.97	2.00	0.04
	LOS10	16.87	20.85	0.53	0.03
Promedio		13.98	15.12	1.64	0.12
Desviación típica		8.64	9.14	0.39	0.16

Tabla 15. Parámetros de canal del modelo SV en cada banda de las posiciones LOS obtenidos con algoritmo Kmeans.

En la Tabla 16 se muestran los parámetros SV que caracterizan el canal en las posiciones NLOS, diferenciando entre las distintas bandas en las que se divide el espectro radioeléctrico medido.

Banda	Posición	γ	Γ	λ	Λ
B1	NLOS1	15.82	15.57	2.23	0.11
	NLOS2	10.41	11.35	1.26	0.09
	NLOS3	27.91	25.95	1.44	2.50
	NLOS4	6.33	7.27	1.67	0.06
B2	NLOS1	12.87	13.00	2.20	0.07
	NLOS2	9.93	9.63	1.81	0.09
	NLOS3	18.89	15.72	1.52	0.19
	NLOS4	6.81	6.23	2.07	0.12
B3	NLOS1	9.86	14.31	1.85	0.16
	NLOS2	27.25	31.99	1.45	0.06
	NLOS3	9.15	24.10	1.76	0.09
	NLOS4	5.98	10.07	2.16	0.08
B4	NLOS1	6.79	12.63	1.66	0.31
	NLOS2	24.66	15.02	1.61	1.00
	NLOS3	7.21	7.82	2.18	0.06
	NLOS4	6.28	8.55	2.09	0.13
Promedio		12.89	14.33	1.81	0.32
Desviación típica		7.48	7.03	0.30	0.61

Tabla 16. Parámetros de canal del modelo SV en cada banda de las posiciones NLOS obtenidos con algoritmo Kmeans.

En las tablas siguientes se muestra un resumen que recoge la media y la desviación típica de todos los resultados de las parametrizaciones del canal en todas las posiciones LOS y NLOS diferenciando los resultados cuando se emplea el algoritmo de Corrigan y el algoritmo Kmeans.

Parámetros Saleh-Valenzuela	γ	Γ	λ	Λ
Promedio LOS Corrigan	1.99	14.87	1.64	0.23
Desv. típica LOS Corrigan	3.99	8.97	0.39	0.12

Tabla 17. Resumen media y desviación típica de los parámetros de las posiciones LOS obtenidos a partir del algoritmo de Corrigan.

Parámetros Saleh-Valenzuela	γ	Γ	λ	Λ
Promedio NLOS Corrigan	1.52	12.75	2.42	0.27
Desv. típica NLOS Corrigan	1.37	3.79	0.32	0.07

Tabla 18. Resumen media y desviación típica de los parámetros de las posiciones NLOS obtenidos a partir del algoritmo de Corrigan.

Parámetros Saleh-Valenzuela	γ	Γ	λ	Λ
Promedio LOS Kmeans	13.98	15.12	1.64	0.12
Desv. típica LOS Kmeans	8.64	9.14	0.39	0.16

Tabla 19. Resumen media y desviación típica de los parámetros de las posiciones LOS obtenidos a partir del algoritmo Kmeans.

Parámetros Saleh-Valenzuela	γ	Γ	λ	Λ
Promedio NLOS Kmeans	12.89	14.33	1.81	0.32
Desv. típica NLOS Kmeans	7.48	7.03	0.30	0.61

Tabla 20. Resumen media y desviación típica de los parámetros de las posiciones NLOS obtenidos a partir del algoritmo Kmeans.

7.3 Clusterización a partir del ángulo acimut y retardo

Queda claro que la clusterización con todas las variables de SAGE: amplitud, ángulos de llegada y retardo, ofrece mejores resultados cuando se utiliza conjuntamente mediante el algoritmo Kmeans que por separado. Aunque los resultados alcanzados demuestran que los ángulos de llegada son más determinantes a la hora de identificar los clústeres que la amplitud de las MPC's.

Este hecho deja en evidencia la practicidad que tiene emplear la amplitud de las MPC como variable principal (junto el retardo) en el estudio del canal radio en entornos *indoor*. De los resultados obtenidos en el capítulo sexto, cabe deducir que la clasificación en clústeres no es adecuada desde el punto de vista de la propagación, ya que existen contribuciones que alcanzan al receptor con mismo ángulo pero distinto retardo, es decir, toma un camino distinto (más largo) siendo propio de una clasificación en clústeres diferentes. Por tanto, es evidente que la clusterización conjunta de los datos no ofrece buenas soluciones desde el punto de la propagación de las ondas EM. Como justificación, se adjunta en la figura 45 la representación en polares a partir del ángulo de acimut junto con el retardo y los resultados de la clusterización cuando se emplean los datos de SAGE en conjunto, ya que ofrece otro punto de vista interesante que estudiaremos en este apartado.

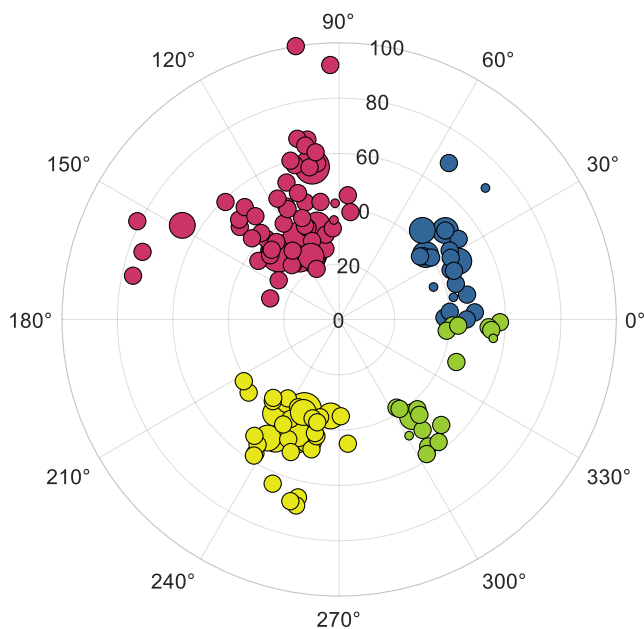


Figura 45. Diagrama en polares del ángulo acimut y el retardo de llegada para la primera posición LOS en B1, clusterización Kmeans con el conjunto de datos.

En la Figura 45 se observa un clúster principal representado en magenta que puede estar compuesto por más de un clúster ya que existen MPC que alcanzan al receptor con retardos mayores desde ángulos similares, indicando caminos distintos (más largos) y por tanto deberían ser diferenciados.

Este hecho indica que la clusterización con la totalidad de los datos no es una práctica adecuada desde el punto de vista de la propagación. Para ello, se propone emplear el ángulo de acimut y el retardo como variables de clusterización. En la Figura 46 se representa en polares los resultados para la primera posición LOS en B1, y en la figura 48, los resultados para la primera posición NLOS en B1.

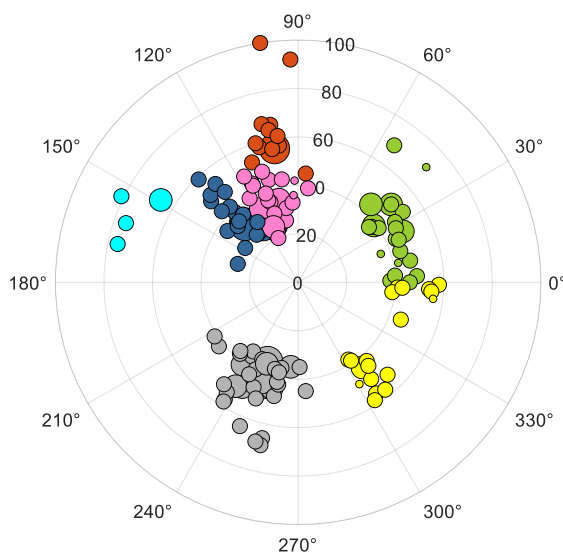


Figura 46. Representación en polares del ángulo de llegada acimut y el retardo, para la primera posición LOS en B1.

En primer lugar, se puede ver como el diagrama une las contribuciones que alcanzan al receptor con ángulos de 0° y de 360° , pudiendo identificar que dichas contribuciones forman un mismo clúster y no dos diferentes. Por otro lado, analizando el clúster naranja que contiene la primera MPC que llega al receptor, está rodeado de otros dos clústeres, de forma que el algoritmo Kmeans es capaz de realizar una clusterización mucho más precisa de los datos.

Por otro lado, se representa en verde un clúster compuesto por contribuciones aisladas que llegan con retardos mayores y ángulos ligeramente diferentes a los principales clústeres, lo cual corrobora que la clusterización tiene en cuenta las singularidades de las contribuciones en su propagación, los distintos caminos que sigue y los obstáculos con los que interactúa.

Se puede superponer este diagrama con el plano de la sala de becarios donde se han realizado las medidas en milimétricas. El resultado se muestra en la Figura 47.

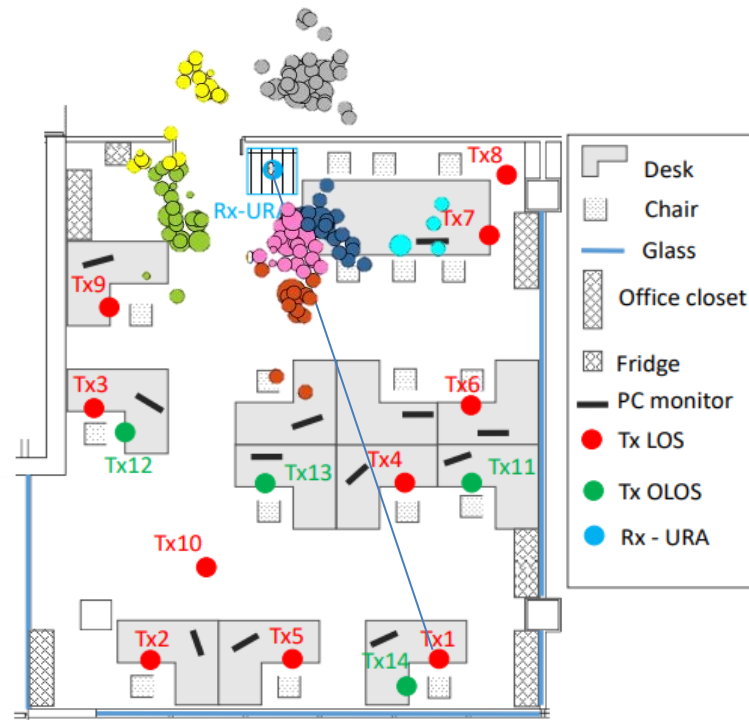


Figura 47. Superposición de la figura 43 en el plano de la sala de becarios, indicando la dirección de llegada de las contribuciones LOS.

Se puede distinguir claramente las contribuciones que llegan con visión directa (siguiendo la línea marcada que une el transmisor con el receptor), recorren la distancia más corta y por tanto llegan antes al receptor, es decir, con un retardo menor. Así como las reflexiones en las paredes de la izquierda y con la de atrás del receptor.

Análogamente se realiza la misma comparación con medidas realizadas en posiciones NLOS. El diagrama en polares se muestra en la Figura 48 y la superposición con el plano de la sala de becarios, en la Figura 49.

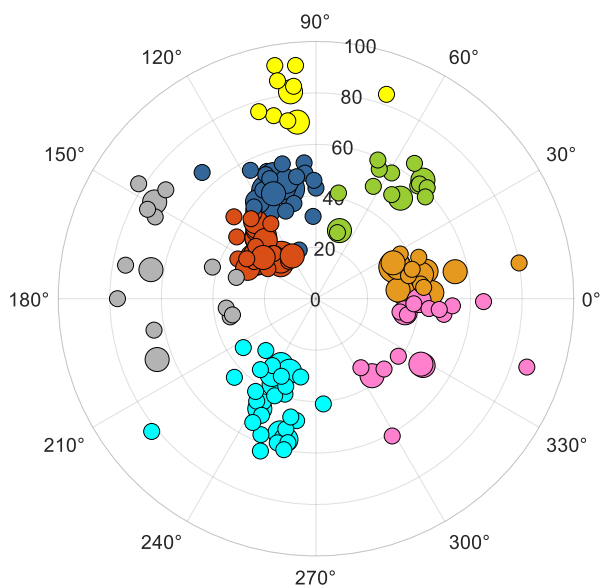


Figura 48. Representación en polares del ángulo de llegada acimut y el retardo, para la primera posición NLOS en B1.

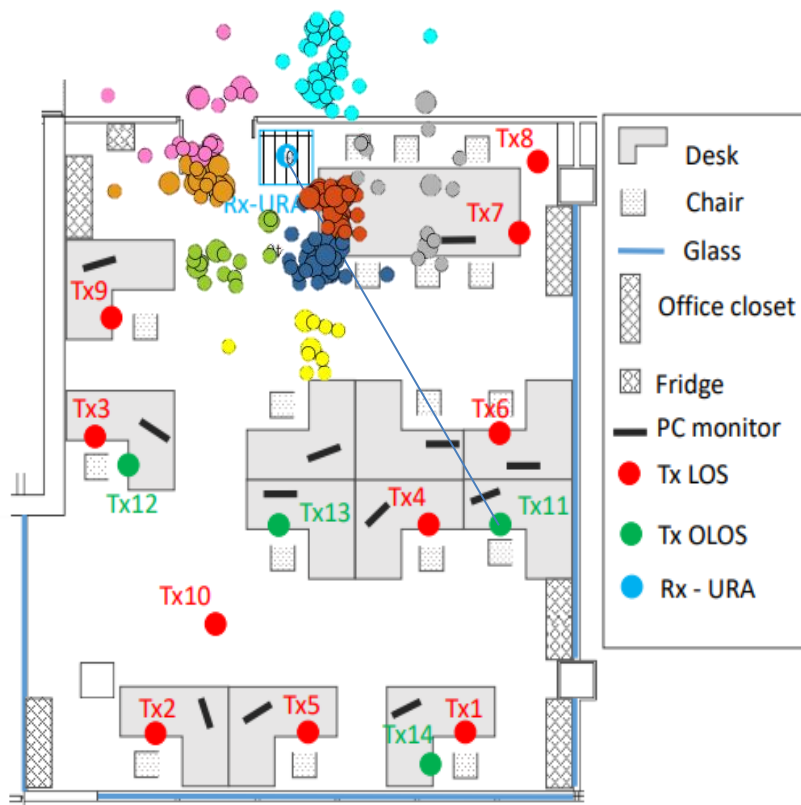


Figura 49. Superposición de la figura 45 en el plano de la sala de becarios, indicando la dirección de llegada de las contribuciones NLOS.



Analizando los resultados de las medidas NLOS cabe decir que surgen clústeres nuevos si se compara con los resultados de la clusterización que emplea el conjunto de datos. A simple vista se puede decir que se repite la necesidad de juntar las contribuciones que llegan con un ángulo de acimut de 360° y de 0° en un mismo clúster. Por otro lado, esta clusterización permite clasificar las MPC que alcanzan al receptor de forma aislada en un mismo clúster, por ejemplo, el clúster gris. De semejante manera con aquellas contribuciones que sí pertenecen claramente a un clúster concreto, como es el caso de los clústeres azul oscuro, azul claro, rojo y naranja. Por su parte, el clúster azul podría ser dividido en dos clústeres diferentes, ya que incluye MPC que llegan con un retardo bastante mayor que el resto de las contribuciones, representado dos grupos diferenciables a vista en un mismo clúster.

A partir del análisis visual de los resultados de la clusterización con el algoritmo Kmeans, se confirma una mejoría notable con respecto al algoritmo de Corrigan. Aunque necesita ser mejorado, desde el punto de vista de propagación, para clasificar adecuadamente las contribuciones MPC en los clústeres correctos.

Capítulo 8. CONCLUSIONES Y LÍNEAS FUTURAS

8.1 Conclusiones

En este trabajo se ha analizado la propagación de ondas EM en recintos *indoor* con el fin de estudiar el comportamiento de las bandas de frecuencia asignadas a la implantación de la tecnología 5G. Por un lado, se ha cumplido con el objetivo de desarrollar un algoritmo basado en ML que permita clusterizar las MPC medidas por el receptor, llegando a obtener resultados coherentes desde el punto de vista de la propagación de las ondas. Por otro lado, también se ha cumplido con el objetivo de determinar la relevancia de emplear información adicional de ángulos de llegada en la clusterización. Para ello se ha empleado el algoritmo lineal extrayendo la conclusión de que el número de clústeres obtenidos es irregular en función de la información que se tenga en cuenta. Desde el punto de vista de los datos, ha quedado patente que emplear información del URA o los datos de SAGE sea determinante en la clusterización, ya que el número total de clústeres finales es irregular. La discrepancia en el número de clústeres hace pensar que información adicional pudiera dar claridad, de tal forma que los ángulos ayuden a extraer conclusiones desde otro punto de vista.

Utilizando los ángulos de forma aislada y combinando los resultados de la clusterización del PDP se concluye que el número final de clústeres es demasiado alto, debido en gran medida por los solapes entre los clústeres. Teniendo en cuenta que por separado no se obtienen resultados coherentes, se propone emplear los datos de SAGE en conjunto. Para ello es necesario emplear un algoritmo que permita utilizar más de dos variables, por lo que se emplea el algoritmo Kmeans de ML. Dicho algoritmo permite obtener un número de clústeres menor, aunque la distribución de las MPC no se ajuste a la realidad desde el punto de vista de la propagación de las ondas, ya que confunde contribuciones que recorren caminos más largos con otras que llegan más rápido, cometiendo errores de clusterización si se analizan los datos a simple vista.

Una vez el algoritmo Kmeans ha dado buenos resultados se propone continuar con el análisis de las MPC con el fin de encontrar la clasificación de las contribuciones que más se ajuste a la propagación real de las ondas. Para ello se identifica el ángulo de acimut y el retardo como las variables más determinantes en la clusterización de los datos de SAGE en conjunto. Es por ello que en el último capítulo se propone clusterizar a partir de dichos datos y el algoritmo Kmeans, alcanzando unos resultados que se reflejan en los diagramas de polares. Analizando dichos diagramas tanto para posiciones LOS como NLOS se demuestra que la información de llegada del ángulo de acimut es determinante, y se consiguen identificar clústeres con MPC aisladas que



no pertenecen de forma clara a clústeres principales en otros clústeres de menor tamaño. Haciendo una clasificación más eficiente desde el punto de vista de la propagación.

Como se puede ver en las figuras de los mapas superpuestos al diagrama de polares, existe una clara unión entre las contribuciones que llegan al receptor con un ángulo de 360° y las que llegan con un ángulo de 0° , uniendo dichas MPC en un único clúster.

Otra conclusión que se extrae de la representación en polares es que se pueden identificar tanto las MPC que alcanzan al receptor con visión directa, reflejado en una amplitud más elevada y un retardo menor. De forma análoga se pueden identificar aquellos objetos que interfieren en la propagación, así como los elementos que generan las reflexiones dentro de la sala de medidas, como es el caso de la pared del lateral izquierdo del mapa ó la reflexión desde la pared de atrás del receptor.

En definitiva, el algoritmo basado en ML, Kmeans, ofrece mejores resultados con respecto al algoritmo lineal cuando se emplea información adicional sobre ángulos de llegada. Siendo el más determinante de ellos el ángulo de acimut, como se ha demostrado en la representación en polares de las MPC.



8.2 Líneas futuras

De las conclusiones se extrae que la clusterización de los datos de canal tiene como principal desventaja la imposibilidad de conocer a ciencia cierta la pertenencia de las contribuciones a un clúster o a otro, por lo que acaba derivándose en una parametrización estadística del canal, donde influyen los algoritmos empleados para conseguir los parámetros SV. Por ello la primera línea futura que se plantea es generar perfiles de retardo aleatorios a partir de los parámetros SV con el fin de caracterizar estadísticamente el canal, obteniendo estadísticos como la tasa de error.

Otra línea futura del trabajo es comparar el modelo Kmeans con otros modelos basados en ML, por ejemplo, haciendo uso de otro tipo de algoritmos de clusterización como es el modelo basado en densidad, o modelos jerárquicos de clasificación.

Por otro lado, también queda demostrado que la información empleada en la clusterización es determinante a la hora de obtener resultados, de forma que se puede mejorar el algoritmo de SAGE con el fin de optimizar la información del canal que extrae del URA, poniendo especial atención a identificar claramente las contribuciones que llegan al receptor sin sobreestimarlas.

REFERENCIAS

- [1] M. Corrigan, A. Walton, Weihong Niu, Jia Li, and T. Talty (2009). Automatic UWB clusters identification. *Hindawi*, pages 376–37. DOI:10.1109/rws.2009.4957359
- [2] Ferrando Navarro, A (2021). Caracterización del canal radio a través de técnicas de clusterización. Trabajo Final de Grado. Valencia: Universidad Politécnica de Valencia.
- [3] A.A.M. Saleh y R. Valenzuela. “A Statistical Model for Indoor Multipath Propagation”. En: *IEEE Journal on Selected Areas in Communications* 5.2 (1987), págs. 128-137. DOI:10.1109/jsac.1987.1146527
- [4] Lorenzo Rubio y col. “Contribution to the Channel Path Loss and TimeDispersion Characterization in an Office Environment at 26 GHz”. En: *Electronics* 8.11 (nov. de 2019), pág. 1261. ISSN: 20799292. DOI: 10.3390/electronics8111261
- [5] Cervera Marín, R (2020). Caracterización experimental de la dispersión temporal y angular del canal radio en un escenario de oficina. Trabajo Final de Grado. Valencia: Universidad Politécnica de Valencia.
- [6] Avellán Carrión, C (2019). Caracterización experimental en la banda de 25 a 40 GHz del canal MIMO multiusuario en un entorno de oficina. Trabajo Final de Grado Avellán Carrión, C (2019). Valencia: Universidad Politécnica de Valencia.
- [7] Jiménez Márquez, I (2019). Diseño y evaluación de algoritmos automáticos de clustering para modelos de canal radio en interiores. Trabajo Final de Grado. Madrid: Universidad Politécnica de Madrid.
- [8] X. Wu, Y. Zhang, C. -X. Wang, G. Goussetis, e. -H. M. Aggoune and M. M. Alwakeel, "28 GHz indoor channel measurements and modelling in laboratory environment using directional antennas," 2015 9th European Conference on Antennas and Propagation (EuCAP), Lisbon, Portugal, 2015, pp. 1-5. DOI: 10.23919/eucap51087.2021.9410912
- [9] Likas, A., Vlassis, N., & Verbeek, J. J. (2003). The global k-means clustering algorithm. *Pattern recognition*, 36(2), 451-461. DOI: 10.1016/s0031-3203(02)00060-2
- [10] Baarsch, J., & Celebi, M. E. (2012). Investigation of internal validity measures for K-means clustering. In *Proceedings of the international multiconference of engineers and computer scientists* (Vol. 1, pp. 14-16). DOI: 10.1201/b16763-19



- [11] Mathworks., (2022). Statistics and Machine Learning Toolbox (R2023a). [Consulta: 3/7/2023]. <https://es.mathworks.com/help/stats/kmeans.html>

- [12] Moreno García Loygorri, Juan Antón, Briso Rodríguez, César, Calvo Ramírez, César, Jiménez Muñoz, Florentino, González Posadas, Vicente, Arnedo, Israel, Laso, Miguel A. G. and He, Danping (2018). *Wideband Channel Modeling for mm-Wave inside Trains for 5G-Related Applications*. "Wireless Communications and Mobile Computing"; pp. 1-11. ISSN 1530-8669. DOI: 10.1155/2018/6916954