

## **ESCUELA DE INGENIERÍA DE TELECOMUNICACIÓN Y ELECTRÓNICA**



### **TRABAJO FIN DE GRADO**

**Técnicas de aprendizaje máquina para detección y cuantificación  
de lluvia mediante radioenlaces terrenos a frecuencia de  
gigahercios**

**Titulación:** Grado en Ingeniería en Tecnologías de la  
Telecomunicación

**Mención:** Sonido e Imagen

**Autora:** Sonia Fuente González

**Tutores:** Dr. Juan Luis Navarro Mesa  
Dr. Eduardo Hernández Pérez

**Fecha:** Julio de 2020



## ESCUELA DE INGENIERÍA DE TELECOMUNICACIÓN Y ELECTRÓNICA



### TRABAJO FIN DE GRADO

**Técnicas de aprendizaje máquina para detección y cuantificación  
de lluvia mediante radioenlaces terrenos a frecuencia de  
gigahercios**

#### HOJA DE EVALUACIÓN

**Calificación:** \_\_\_\_\_

**Presidente:**

**Vocal:**

**Secretario:**

**Fecha: Julio de 2020**



## Agradecimientos

Me gustaría mostrar mi agradecimiento a mis tutores Dr. Juan Luís Navarro Mesa y al Dr. Eduardo Hernández Pérez por aconsejarme, guiarme y enseñarme en estos últimos meses.

Al Dr. Juan Luís Navarro Mesa, por proponerme y ayudarme en la investigación de este tema tan innovador que, sin su ayuda, no habría sido posible abordar. Por todas las tutorías, los e-mails enviados, la resolución a la cantidad de dudas que me han ido surgiendo y por las correcciones hechas. Con todo ello, ha hecho posible que la tarea de realizar este trabajo haya sido más fácil y llevadera, todo con un trato inmejorable.

Al COITT, en especial a Don Sebastián Suárez Cano por la información aportada para la realización del presupuesto

Mi agradecimiento también a mis familiares y amigos que han estado animándome y apoyándome durante estos cuatro años.

## Resumen

En este Trabajo de Fin de Grado se ha realizado la implementación de algoritmos de Machine Learning para la detección y cuantificación de precipitación de lluvia basándonos en los niveles de potencia involucrados en comunicaciones terrenas a frecuencias de microondas distribuidas por las zonas de observación. Nos apoyamos en el hecho de que cuando llueve se produce una atenuación de la señal de potencia, esto es especialmente significativo para frecuencias mayores a 10GHz.

Se ha utilizado cuatro métodos para abordar este objetivo, estos son: regresión lineal, regresión logística, red neuronal y máquina de vectores soportados. Asimismo, se ha llevado a cabo un balanceo de las clases pues nos encontramos ante un problema de clases desbalanceadas y un posprocesado en el que se han calibrado los resultados. Todo esto se enmarca en un proceso de aprendizaje máquina en el que se ha utilizado un 70% de las muestras (datos) para entrenar y un 30% para test. Para estudiar la validez de nuestra metodología, se hicieron repeticiones de los experimentos en los que cada vez se asignaban de forma aleatoria qué ejemplos eran para entrenar y cuáles para testear. Como resultado hemos visto que nuestra metodología es muy estable.

Tras todo el trabajo se puede concluir que los modelos obtenidos son válidos, y dan muy buenos resultados en los test de detección para los cuatro métodos, destacando en valores medios y desviaciones de sensibilidad, exactitud, especificidad y F1-Score la red neuronal. En lo referente a la cuantificación podemos destacar que los valores obtenidos son muy exactos.



## Abstract

In this Final Degree Project, we have implemented Machine Learning algorithms for the detection and quantification of rain precipitation. It has been carried out based on the power levels involved of terrestrial radiolink communications at microwave frequencies distributed throughout an observation area. We rely on the fact that when there is a rainfall the power signal suffers an attenuation, which is especially significant for frequencies above 10GHz.

Four methods have been used to achieve this objective: linear regression, logistic regression, neural network and supported vector machine. Likewise, a balancing of the classes has been carried out since we are facing a problem of imbalanced classes, and a post-processing of the detection-quantification in which the results have been calibrated in order to improve the performance. All this is framed in a machine learning process in which 70% of the samples (data) have been used for training and 30% for testing. To study the validity of our methodology, repetitions of the experiments were made in which each time the train or test samples were randomly assigned. As a result, we have seen that our methodology is very stable.

After all the work, it can be concluded that the models obtained are valid, and give good detection results with all four methods, highlighting the values of neural network for mean values and deviations of: sensitivity (recall), accuracy, specificity and F1-Score. Regarding rain level quantification, we have achieved very accurate results.



# Índice General

## Parte I

<b>Memoria .....</b>	<b>3</b>
<b>Capítulo 1 - Visión Global.....</b>	<b>3</b>
<b>1.1 Introducción .....</b>	<b>3</b>
<b>1.2 Estado de la técnica de observación meteorológica .....</b>	<b>3</b>
<b>1.3 El radioenlace de comunicación como alternativa.....</b>	<b>6</b>
<b>1.4 Objetivos .....</b>	<b>7</b>
<b>1.5 Estructura de la memoria.....</b>	<b>8</b>
<b>Capítulo 2 - Fundamentos teóricos acerca de los efectos de la lluvia sobre las ondas electromagnéticas .....</b>	<b>11</b>
<b>2.1 Condiciones meteorológicas .....</b>	<b>11</b>
2.1.1 El aire .....	12
2.1.2 Precipitación .....	13
2.1.3 La temperatura .....	15
2.1.4 Humedad .....	16
<b>2.2 Efectos de las condiciones meteorológicas en señales de radiofrecuencia.....</b>	<b>16</b>
2.2.1 Atenuación por hidrometeoros .....	16
2.2.2 Atenuación por nubes, niebla, arena y polvo. ....	17
2.2.3 Atenuación por gases atmosféricos .....	18
<b>2.3 Atenuación específica de la lluvia .....</b>	<b>19</b>
2.3.1 Tipos de lluvia .....	20
2.3.2 Altura de la lluvia .....	21
2.3.3 Tamaño, forma y orientación de las gotas.....	22
2.3.4 Velocidad de caída .....	23
2.3.5 Temperatura .....	23
2.3.6 Fórmula de la atenuación específica de la lluvia .....	23
<b>Capítulo 3 - Adquisición de datos y Preprocesado .....</b>	<b>27</b>
<b>3.1 Fuentes de datos .....</b>	<b>27</b>
3.1.1 Región de Trabajo .....	28

<b>3.3 Radioenlaces .....</b>	<b>29</b>
3.3.1 Características específicas .....	30
3.3.2 Adquisición de los datos de los radioenlaces.....	31
3.3.3 Localización de vacíos .....	32
<b>3.4 Estaciones meteorológicas.....</b>	<b>33</b>
3.4.1 Características específicas .....	34
3.4.2 Adquisición de los datos meteorológicos .....	35
3.4.3 Tratamiento de los datos meteorológicos .....	37
<b>3.5 Sincronización de los datos .....</b>	<b>40</b>
3.5.1 Eliminación de GAPS .....	42
<b>Capítulo 4 - Marco metodológico .....</b>	<b>44</b>
<i>En este capítulo realizamos una breve introducción al aprendizaje máquina y sus diferentes formas. También, explicamos brevemente los distintos métodos de clasificación que utilizaremos en este trabajo de fin de grado.....</i>	
<b>4.1 Introducción .....</b>	<b>44</b>
<b>4.2 Generalidades sobre el aprendizaje máquina.....</b>	<b>45</b>
4.2.1 Formas de aprendizaje.....	47
4.2.2 Bases de datos desbalanceadas y estrategias de balanceo. ....	48
<b>4.3 El problema de regresión. Descripción general.....</b>	<b>49</b>
4.3.1 Regresión lineal.....	50
4.3.2 Regresión logística .....	51
<b>4.4 Redes neuronales .....</b>	<b>53</b>
4.4.1 Propagación hacia atrás de errores o backpropagation. ....	56
4.4.2 Redes neuronales como método de regresión .....	56
<b>4.5 Máquinas de vectores soportados. ....</b>	<b>57</b>
4.5.1 El clasificador de vectores soportados.....	57
4.5.2 El concepto de Núcleo de la máquina.....	59
4.5.3 Máquinas de vectores soportados para la regresión.....	60
<b>Capítulo 5 - Algoritmos de detección y cuantificación de precipitación .....</b>	<b>62</b>
<b>5.1 Eliminación baja frecuencia de la señal .....</b>	<b>62</b>
<b>5.2 Extracción de características .....</b>	<b>64</b>
5.2.1 Características transformadas. ....	67

5.3 Balanceo de la base de datos .....	68
5.4 Definición de niveles de referencia .....	68
5.5 Calibrado de la regresión .....	69
<b>Capítulo 6 - Experimentos y Resultados.....</b>	<b>73</b>
6.1 Parámetros de calidad .....	73
6.1.1 Definición de parámetros de detección.....	73
6.1.2 Definición de parámetros de cuantificación .....	75
6.1.3 Estudio de la fiabilidad.....	76
6.2. Experimentos con regresión lineal .....	77
6.3 Experimentos con regresión logística .....	83
6.4 Experimentos con red neuronal .....	89
6.5 Experimentos con máquina de vectores soportados (SVM) .....	94
6.6 Resultados finales.....	100
<b>Capítulo 7 – Conclusiones y Líneas Futuras.....</b>	<b>104</b>
7.1 Conclusiones.....	104
7.2 Líneas futuras .....	105
<b>Bibliografía.....</b>	<b>107</b>

## Parte II

<b><i>Pliego de Condiciones .....</i></b>	<b>113</b>
Condiciones legales y administrativas.....	113
Derechos de explotación y restricciones de uso .....	113
Garantías y consideraciones adicionales .....	114
Herramientas utilizadas.....	115

## Parte III

<b><i>Presupuesto.....</i></b>	<b>120</b>
Declaración Jurada .....	120

<b>Desglose del presupuesto .....</b>	<b>121</b>
C-1. Trabajo tarifado por tiempo empleado .....	121
C-2. Amortización del inmovilizado material .....	122
C-3. Costes de redacción del Trabajo Fin de Grado .....	123
C-4. Derechos de visado.....	124
C-5. Gastos de tramitación y envío .....	125
<b>Resumen del presupuesto .....</b>	<b>126</b>

## Índice de Figuras

Figura 1 Distintos sistemas de observación. Fuente: AEMET .....	5
Figura 2 Diferentes frentes: Frío (izquierda), Cálido (derecha). Fuente: <a href="https://www.canaltiempo21.com">https://www.canaltiempo21.com</a> ...	12
Figura 3 Atenuación específica debida a los gases atmosféricos. Fuente: (ITU-R', 1997) .....	19
Figura 4 Media Anual de la isoterma a 0°C por encima del nivel del mar (km). Fuente: (ITU_R, 2013).....	22
Figura 5 Variación de los parámetros $k$ y $\alpha$ . Fuente: (ITU, 2005).....	25
Figura 6 Estación meteorológica de la AEMET situada en Niembro (Asturias). Fuente: AEMET.....	27
Figura 7 Estaciones meteorológicas de la AEMET en el archipiélago canario. Fuente: AEMET .....	28
Figura 8 Señal de potencia recibida en el radioenlace Monteluz-Hospital Militar. ....	29
Figura 9 Señal de potencia en la que se han producido GAPS .....	33
Figura 10 Nivel de lluvia acumulada en la estación Las Palmas de GC Tafira-Zurbarán .....	34
Figura 11 Pluviómetro: Vaisala Rain Gauge RG13. Fuente: <a href="https://www.vaisala.com">https://www.vaisala.com</a> .....	35
Figura 12 Datos de niveles de precipitación en los que se han sucedido GAPS .....	38
Figura 13 Señal de potencia y nivel acumulado de lluvia en el mismo intervalo temporal .....	40
Figura 14 Portada del periódico Cararias7 del día 22 de octubre de 2015.....	41
Figura 15 Ejemplo de función de regresión logística. Fuente: elaboración propia .....	52
Figura 16 Esquema básico de una neurona artificial con cuatro características de entrada. Fuente: elaboración propia .....	54
Figura 17 Representación gráfica de la estructura de un perceptrón multicapa de dos niveles. ....	55
Figura 18 Ejemplo de SVM con dos predictores X y Z y dos clases linealmente separables .....	58
Figura 19 Resultados binarios no separables. Fuente: elaboración propia .....	59
Figura 20 Señal RSL a la que se le ha aplicado una apertura y cierre morfológico .....	63
Figura 21 Eliminación de baja frecuencia de la señal mediante cierre morfológico .....	64
Figura 22 Ventanas de correlación de 8 máscaras Fuente: (Guerra Moreno, 2017).....	65
Figura 23 Representación de los vectores de características de no lluvia (izquierda) y lluvia(derecha) con 8 máscaras .....	66
Figura 24 Representación de los 10 primeros vectores de características de no lluvia (izquierda) y lluvia(derecha).....	67
Figura 25 Gráfica superior: datos de entrenamiento; Gráfica intermedia: .....	70
Figura 26 Señal calibrada del radioenlace a Monteluz - Hospital .....	71
Figura 27 Resultados para el radioenlace Agaete – Montaña de Gáldar.....	78
Figura 28 Resultados para el radioenlace Agüimes – Tabaiba .....	79
Figura 29 Resultados para el radioenlace Monteluz - Hospital .....	79
Figura 30 Resultados para el radioenlace Moriscos - Teror .....	80
Figura 31 Resultados para el radioenlace Anocheza - Birmagen .....	81
Figura 32 Resultados para el radioenlace Anocheza – El Reverón .....	81
Figura 33 Resultados para el radioenlace Suculum - Bailadero .....	82

---

<i>Figura 34 Resultados para el radioenlace Suculum – Casas Cumbre .....</i>	<i>83</i>
<i>Figura 35 Resultados para el radioenlace Agaete - Montaña de Gáldar .....</i>	<i>84</i>
<i>Figura 36 Resultados para el radioenlace Agüimes – Tabaiba .....</i>	<i>85</i>
<i>Figura 37 Resultados para el radioenlace Monteluz - Hospital .....</i>	<i>85</i>
<i>Figura 38 Resultados para el radioenlace Moriscos - Teror .....</i>	<i>86</i>
<i>Figura 39 Resultados para el radioenlace Anocheza - Birmagen .....</i>	<i>87</i>
<i>Figura 40 Resultados para el radioenlace Anocheza – El Reverón .....</i>	<i>87</i>
<i>Figura 41 Resultados para el radioenlace Suculum - Bailadero .....</i>	<i>88</i>
<i>Figura 42 Resultados para el radioenlace Suculum - Casas Cumbre .....</i>	<i>89</i>
<i>Figura 43 Resultados para el radioenlace Agaete - Montaña de Gáldar .....</i>	<i>90</i>
<i>Figura 44 Resultados para el radioenlace Agüimes – Tabaiba .....</i>	<i>90</i>
<i>Figura 45 Resultados para el radioenlace Monteluz – Hospital .....</i>	<i>91</i>
<i>Figura 46 Resultados para el radioenlace Moriscos – Teror.....</i>	<i>92</i>
<i>Figura 47 Resultados para el radioenlace Anocheza – Birmagen.....</i>	<i>92</i>
<i>Figura 48 Resultados para el radioenlace Anocheza - El Reverón .....</i>	<i>93</i>
<i>Figura 49 Resultados para el radioenlace Suculum – Bailadero.....</i>	<i>93</i>
<i>Figura 50 Resultados para el radioenlace Suculum - Casas cumbre.....</i>	<i>94</i>
<i>Figura 51 Resultados del radioenlace Agaete - Montaña de Gáldar.....</i>	<i>95</i>
<i>Figura 52 Resultados para el radioenlace Agüimes – Tabaiba .....</i>	<i>95</i>
<i>Figura 53 Resultados para el radioenlace Monteluz - Hospital .....</i>	<i>96</i>
<i>Figura 54 Resultados para el radioenlace Moriscos – Teror.....</i>	<i>97</i>
<i>Figura 55 Resultados para el radioenlace Anocheza – Birmagen.....</i>	<i>97</i>
<i>Figura 56 Resultados para el radioenlace Anocheza - El Reverón .....</i>	<i>98</i>
<i>Figura 57 Resultados para el radioenlace Suculum – Bailadero.....</i>	<i>99</i>
<i>Figura 58 Resultados para el radioenlace Suculum - Casas Cumbre .....</i>	<i>99</i>

## Índice de Tablas

<i>Tabla 1 Distintos tipos de lluvia según su intensidad.</i>	13
<i>Tabla 2 Coeficientes de las fórmulas (2.3) y (2.4) para polarización horizontal:</i>	24
<i>Tabla 3 Coeficientes de las fórmulas (2.3) y (2.4) para polarización vertical:</i>	24
<i>Tabla 4 Radioenlaces seleccionados para Gran Canaria y Tenerife.</i>	29
<i>Tabla 5 Características técnicas de los radioenlaces</i>	31
<i>Tabla 6 Datos de la estación meteorológica de Tafira.</i>	37
<i>Tabla 7 Eventos de lluvia y de no lluvia de los radioenlaces de estudio</i>	68
<i>Tabla 8 Definición gráfica de True Positive, True Negative, False Positive y False Negative.</i>	74
<i>Tabla 9 Matriz de confusión del modelo</i>	74
<i>Tabla 10 Resultados finales para el radioenlace Agaete – Montaña de Gáldar</i>	100
<i>Tabla 11 Resultados finales para el radioenlace Agüimes - Tabaiba</i>	100
<i>Tabla 12 Resultados finales para el radioenlace Monteluz - Hospital</i>	101
<i>Tabla 13 Resultados finales para el radioenlace Moriscos - Teror</i>	101
<i>Tabla 14 Resultados finales para el radioenlace Anocheza - Birmagen</i>	101
<i>Tabla 15 Resultados finales para el radioenlace Anocheza – El Reverón</i>	102
<i>Tabla 16 Resultados finales para el radioenlace Suculum - Bailadero</i>	102
<i>Tabla 17 Resultados finales para el radioenlace Suculum – Casas Cumbre</i>	102



## Acrónimos

AEMET	Agencia Española de Meteorología
COITT	Colegio Oficial de Graduados e Ingenieros Técnicos de Telecomunicación
EMA	Estación Meteorológica Automática
FN	False Negative
FP	False Positive,
GOS	Global Observing System
IDeTIC	Instituto para el Desarrollo Tecnológico y la Innovación en Comunicaciones
IoT	Internet of Things
ML	Machine Learning
ODU	Output Devices Units
PFC	Proyecto de Fin de Carrera
RAE	Real Academia Española
SHF	Super High Frequency
TETRA	Terrestrial Trunked Radio
TFG	Trabajo de Fin de Grado
TN	True Negative
TP	True Positive
WCS	Wireless Communication System
WMO	World Meteorological Organization



# Parte I:

# Memoria



# Memoria

## Capítulo 1 - Visión Global

### 1.1 Introducción

En este capítulo de la memoria describimos los distintos sistemas de observación meteorológica que existen en la actualidad. A continuación, hablamos sobre las alternativas para dicha observación basadas en radioenlaces terrenos a frecuencias de microondas. Finalmente, describimos los objetivos de este proyecto y presentamos la estructura de esta memoria.

Antes de proceder hemos de decir que este trabajo se desarrolla en el marco del proyecto VIMETRI-MAC, del que la ULPGC es Jefa de Filas, y participan como socios beneficiarios también la Universidad de Madeira y el Laboratorio de Ingeniería Civil de Madeira. Como socios colaboradores participan la Agencia Estatal de Meteorología, y las Direcciones Generales del Seguridad y Emergencias y la de Telecomunicaciones y Nuevas Tecnologías del Gobierno de Canarias.

### 1.2 Estado de la técnica de observación meteorológica

La ciencia que estudia la atmósfera, sus propiedades y de los fenómenos que tienen lugar en la misma, los denominados meteoros, es la Meteorología (Rodríguez Jiménez, 2004). Todo estudio de la atmósfera supone disponer de datos meteorológicos precisos y comparables entre sí (Núñez & Gentile, 2004) y se basa en el conocimiento de una serie de variables meteorológicas como, por ejemplo, la temperatura, la presión atmosférica o la humedad, las cuales varían tanto en el espacio como en el tiempo. Por otro lado, hablamos de el tiempo atmosférico cuando describimos las condiciones atmosféricas en un momento y lugar concretos (Rodríguez Jiménez, 2004).

Según la Real Academia Española (RAE) un meteoro es un fenómeno atmosférico que puede ser aéreo, como los vientos, acuoso, como la lluvia o la nieve, luminoso, como el arco iris, el parhelio o la paraselene, y eléctrico, como el rayo y el fuego de San Telmo (RAE, 2020). Este Trabajo de Fin de Grado (TFG) se centra en el hidrometeoro lluvia. Las mediciones de lluvia han sido investigadas en todo el mundo debido a sus importantes implicaciones en meteorología, hidrología,

advertencias de inundaciones y gestión de recursos de agua dulce (Rayitsfeld, 2012).

Como ya mencionamos anteriormente, todo estudio de la atmósfera supone disponer, ante todo, de datos meteorológicos precisos y comparables entre sí. Así, definimos observación meteorológica como la técnica que consiste en la medición de forma sistemática, uniforme, ininterrumpida y a horas establecidas, de todas las variables atmosféricas que representan las condiciones meteorológicas en un momento dado y en un determinado lugar. Estas mediciones deben realizarse utilizando instrumental adecuado y complementado por los sentidos del observador, si fuera necesario, principalmente la vista. Solo así podremos conocer las características y las variaciones de las variables atmosféricas con exactitud, las cuales constituyen los datos básicos que utilizan los servicios meteorológicos, tanto en tiempo real como en diferido (Núñez & Gentile, 2004).

La observación de los diversos elementos meteorológicos se hace en puestos especialmente denominados estaciones meteorológicas y se conforman tanto de observaciones in-situ como por medición remota (Renom, 2011), tal y como podemos observar en la Figura 1.

La siguiente forma de clasificación de los sistemas de observación pertenece al grupo de observación global (Global Observing System, GOS) que forma parte de la red de observación de la Organización Meteorológica Mundial (World Meteorological Organization, WMO): (WMO, 2020)

- Observaciones en superficie.
- Observaciones de altura.
- Observaciones marinas.
- Observaciones provenientes de aviones.
- Observaciones satelitales.

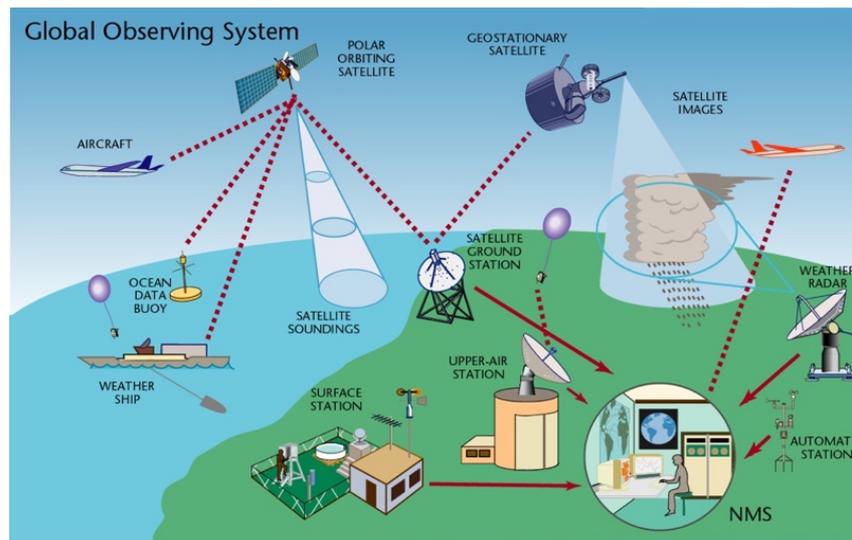


Figura 1 Distintos sistemas de observación. Fuente: AEMET

Uno de los sistemas de captura de datos más difundidos y de los que hemos recogido los datos para este TFG son las estaciones meteorológicas automáticas (EMA) (WMO, 2020), una EMA es una estación meteorológica en la que se realizan y se transmiten observaciones automáticamente. Las variables meteorológicas medidas mediante los sensores que se encuentran integrados son leídas o recibidas por una unidad central de adquisición de datos.

Otros dos tipos de sistemas de observación meteorológica importantes son los radares y los satélites meteorológicos. Los radares meteorológicos pueden detectar y localizar zonas de precipitación, medir su intensidad, desplazamiento y su tipo (granizo, lluvia, nieve, etc.). Además, detectan las variaciones del índice de refracción en la atmósfera generadas por variaciones locales de la temperatura o de la humedad. También a través de ellos se puede estimar la velocidad y dirección del viento a alturas bajas. Los satélites meteorológicos suministran datos actualizados continuamente, de las condiciones meteorológicas que afectan a grandes áreas geográficas. Se sitúan a altitudes de unos 36000 km (geoestacionarios) o de 850 km, (polares).

Además de los sistemas anteriores existen otros sistemas de observación como son: las boyas meteorológicas, los buques de observación marítima, las radiosondas, etc. Estos sistemas presentan unas desventajas pues los datos que recogen son en puntos muy específicos, no proporcionan con exactitud la hora de inicio y de fin de los eventos meteorológicos, necesitan de constante mantenimiento de forma específica, muchos casos de GAPS, etc. (González Gómez, 2017).

El auge del Internet de las Cosas (Internet of Things, IoT) y las técnicas de análisis de datos ha facilitado estas labores en los últimos años con técnicas como las basadas en Aprendizaje Máquina (Machine Learning, ML) entre otras, dando solución a distintos problemas con amplia aplicabilidad (Kaur, 2012).

### **1.3 El radioenlace de comunicación como alternativa**

La gran promesa de los sistemas de telecomunicación, en ciertos aspectos, se puede resumir en una sola palabra ubicuidad para ello la sociedad se ha dotado de múltiples y diversos soportes de comunicación que permiten hacer realidad dicha promesa. Entre ellos los radioenlaces, enmarcados en los servicios terrestres, facilitan la extensión del alcance de los sistemas y servicios de comunicación de forma inalámbrica.

Para la mayoría de los sistemas de comunicación un aspecto importante es el porcentaje de tiempo que el enlace permite la realización de comunicaciones con unas especificaciones de tasa de error determinada. Sorteando las averías, interferencias y desvanecimientos por multitrayecto, hay otros factores que afecta significativamente a dichas comunicaciones. Cuando hablamos de la banda de frecuencia superalta (Super High Frequency, SHF), banda comprendida entre 3 y 30 GHz (ITU-R, 2015) las atenuaciones producidas en la señal son debidas en múltiples ocasiones a la meteorología y se relacionan directamente con los llamados hidrometeoros; de ellos la lluvia que se puede considerar el factor más limitante.

Las perturbaciones en la propagación de radio por microondas causadas por la dispersión y la absorción de la precipitación han sido estudiadas para optimizar el diseño y la planificación confiables del Sistema de Comunicación Inalámbrica (Wireless Communication System, WCS). Recientemente, se ha propuesto que estos mismos impedimentos se pueden utilizar como una forma novedosa de medir la lluvia y otros fenómenos meteorológicos (Rayitsfeld, 2012) tales como los parámetros atmosféricos (temperatura, humedad, presión, etc.), los hidrometeoros (lluvia, nieve, granizo, etc.), los litometeoros (calima, humo, viento, etc.) (AEMET, 2020).

Ya en 2009, Artem Zinevich, Hagit Messer y Pinhas Alpert (Zinevich, 2009) publicaron un artículo donde se presentaba un enfoque distinto para la

reconstrucción de la dinámica espaciotemporal de la lluvia desde una red inalámbrica de microondas. Afirmaban que la técnica es capaz de agregar los datos en el tiempo y el espacio a lo largo de la dirección de movimiento del campo de lluvia, que se recupera de la observación simultánea de una multitud de enlaces de microondas (Zinevich, 2009).

Posteriormente, estos tres mismos investigadores afirman que este método basado en estimar la intensidad de la lluvia a partir de las variaciones de la potencia de la señal transmitida puede revelar la evolución a escala fina de la lluvia en el espacio y tiempo y permite la observación de precipitaciones cercanas a la superficie a una resolución espacial de hasta  $1 \times 1 \text{ km}^2$  y una resolución temporal de hasta 1 min, sin costos adicionales de instalación y mantenimiento (Rayitsfeld, 2012).

## 1.4 Objetivos

Este TFG es un proyecto que se basa en sistemas de telecomunicación terrenales desplegados con alguna finalidad (p.e., servicios de telefonía, emergencia) y trabajamos con la idea de darles valor añadido hacia la observación meteorológica. Planteamos un funcionamiento alternativo y complementario a los métodos tradicionales actuales de tal forma que se aporta nuevo valor a las redes de telecomunicación.

**Objetivo principal:** Implementación de algoritmos de Machine Learning (ML) para la detección y cuantificación de precipitación de lluvia basado en los niveles de potencia involucrados en comunicaciones terrenas a frecuencias de microondas distribuidas por las zonas de observación.

**Objetivos secundarios:** Para alcanzar el citado objetivo es necesario identificar otros más específicos que, vinculados a este permitan alcanzarlo. Así, distinguimos los siguientes cuatro objetivos:

**Objetivo 1: Estudio de las precipitaciones y su influencia sobre la atenuación de señales en radioenlaces terrenos.** Con este objetivo queremos estudiar cómo conociendo el par parámetros-degradación en un radioenlace terrenal podemos obtener información que nos permita hacer hipótesis acerca de los mecanismos de propagación. Nos centramos en las precipitaciones debidas a lluvia. En este sentido, vemos la atmósfera terrestre como un sistema oculto para

un observador situado en el emisor o el receptor. Este sistema tiene capacidad de modificar la señal transmitida, caracterizada esta por parámetros conocidos por diseño, y cuya respuesta es una señal afectada por el sistema atmósfera.

**Objetivo 2: Bases de datos de radioenlaces y estaciones meteorológicas.** Una vez estudiadas las precipitaciones y su influencia sobre la atenuación, debemos gestionar la base de datos de que disponemos a fin efecto de que sirva como elemento de aprendizaje en etapas posteriores donde usaremos técnicas de ML. Así, usaremos dos bases de datos: 1) Niveles de potencia proporcionados por la Dirección General de Telecomunicaciones del Gobierno de Canarias y 2) Datos de estaciones meteorológicas proporcionados por la Agencia Estatal de Meteorología. En este estudio se tendrá en cuenta aspectos como la ubicación de las antenas emisora y receptora, y la ubicación de las estaciones meteorológicas cercanas.

**Objetivo 3: Esquema general del sistema.** Dentro de este apartado de los objetivos diseñaremos los elementos específicos del sistema: adquisición de datos y su procesado, extracción de características, aplicación de técnicas de aprendizaje máquina. Particularmente, desarrollaremos los elementos principales del sistema que nos permitirán obtener resultados.

**Objetivo 4: Estudio de los algoritmos de detección lluvia y su cuantificación.** Particularmente, estudiaremos técnicas basadas en redes neuronales, máquinas de vectores soportados, etc. Pondremos énfasis también en el estudio de las técnicas más apropiadas para extraer las características más relevantes.

## 1.5 Estructura de la memoria

La redacción de este documento se ha estructurado en diferentes capítulos, estos son:

**Capítulo 1. Visión Global.** Hacemos una breve introducción sobre los sistemas de observación meteorológica y la alternativa basada en radioenlaces. Describimos los objetivos principales del proyecto.

**Capítulo 2. Fundamentos teóricos acerca de los efectos de la lluvia sobre las ondas electromagnéticas.** Indagamos en los posibles fenómenos

meteorológicos que pueden influir en la atenuación de la señal y nos centramos en la atenuación por el hidrometeoro lluvia.

**Capítulo 3. Adquisición de datos y Preprocesado.** Analizamos las distintas fuentes de datos al igual que la misma adquisición de los datos y el previo tratamiento que hay que hacer para utilizarlos.

**Capítulo 4. Marco metodológico.** Realizamos una breve introducción al aprendizaje máquina y sus diferentes formas. Explicamos brevemente los distintos métodos de clasificación que utilizaremos en este trabajo de fin de grado.

**Capítulo 5. Algoritmos de detección y cuantificación de precipitación.** Describimos el marco general del método desarrollado en los programas

**Capítulo 6. Experimentos y Resultados.** Se presenta los parámetros que hemos utilizado para definir la calidad de los experimentos, se detallan los experimentos realizados y se muestran los resultados obtenidos.

**Capítulo 7. Conclusiones y Líneas Futuras.** En este capítulo se exponen las conclusiones obtenidas del trabajo y las posibles líneas futuras.

Al final de la memoria se encuentra la Bibliografía en donde se encuentran las referencias bibliográficas empleadas en este TFG.



## Capítulo 2 - Fundamentos teóricos acerca de los efectos de la lluvia sobre las ondas electromagnéticas

En este capítulo vamos a estudiar los posibles fenómenos meteorológicos que pueden influir en la atenuación de la señal, centrándonos en la atenuación por el hidrometeoro lluvia., el cual lo analizaremos con más detalle.

Para la realización de este capítulo se ha utilizado la información proporcionada por la tesis de Iván Daniel Guerra Moreno (Guerra Moreno, 2017), así como el Proyecto de Fin de Carrera (PFC) de Patricia del Carmen González Gómez (González Gómez, 2017) que, al igual que este TFG, se enmarcan en el proyecto VIMETRI-MAC (Navarro Mesa & Ravelo García, 2017-2019) aprobado en la primera convocatoria dentro del Programa de Cooperación Territorial. INTERREG V A España- Portugal. MAC 2014-2020. Este proyecto es desarrollado por investigadores del Instituto para el Desarrollo Tecnológico y la Innovación en Comunicaciones (IDeTIC), División de Procesado Digital de Señales, de la Universidad de Las Palmas de Gran Canaria.

### 2.1 Condiciones meteorológicas

Como vimos en el capítulo anterior, la meteorología es la ciencia que estudia la atmósfera y los mecanismos que producen cambios en ella. Ésta, en compañía de la climatología que es la ciencia que estudia las variaciones a largo plazo del tiempo meteorológico, vienen determinadas principalmente por la posición y los movimientos del planeta respecto al Sol.

Los movimientos del planeta Tierra son periódicos o cuasi-periódicos, éstos provocan variaciones en la luz recibida y por tanto en la cantidad de calor recibido en las distintas zonas del planeta. Estas variaciones de radiación solar y calor originan cambios en las temperaturas de la atmósfera que producen gradientes de temperatura. El aire se desplaza debido a los gradientes de temperatura de unas zonas a otras creando variaciones de presión.

Nos solemos referir al tiempo meteorológico cuando hablamos del estado de la atmosfera en un momento específico. Normalmente en términos de luminosidad, lluvioso, factores aéreos y factores eléctricos, fundamentalmente, términos que engloban entre otros a los hidrometeoros, lluvia, nieve, granizo, etc.

### 2.1.1 El aire

La atmósfera se puede definir como la envoltura de gases que rodea la Tierra. El aire es una mezcla de gases caracterizados por ser transparente, inodoro e insípido; está constituido principalmente por oxígeno en un 20.946% y nitrógeno en un 78.09%, y por cantidades variables de otros componentes como dióxido de carbono, vapor de agua, argón, etc.

Debido a los cambios de temperaturas las masas de aire se calientan o se enfrían. Se denomina frente a la línea imaginaria que separa dos masas de aire de temperatura diferentes. Los frentes se suelen desplazar guiados por la dirección del viento o por las formas de la superficie terrestre. Hay tres clases de frentes (Figura 2): frente cálido, frente frío y ocluido.

En el frente cálido una masa de aire caliente empuja una masa de aire frío. El aire frío es más pesado y adopta una forma de cuña por debajo del aire caliente que asciende siguiendo la especie de rampa que constituye la línea frontal. En la zona de contacto el aire caliente se enfría hasta condensarse y formar nubes que provocan precipitaciones, estas son débiles pero continuadas.

En el frente frío una masa de aire frío empuja una masa de aire caliente. El aire frío, al ser más denso, avanza con rapidez por la superficie y hace elevar con fuerza el aire más cálido, que se enfría y se condensa formando originan tormentas intensas.

Los frentes ocluidos se forman cuando un frente frío, al moverse más rápido, atrapa un frente cálido y se fusionan.

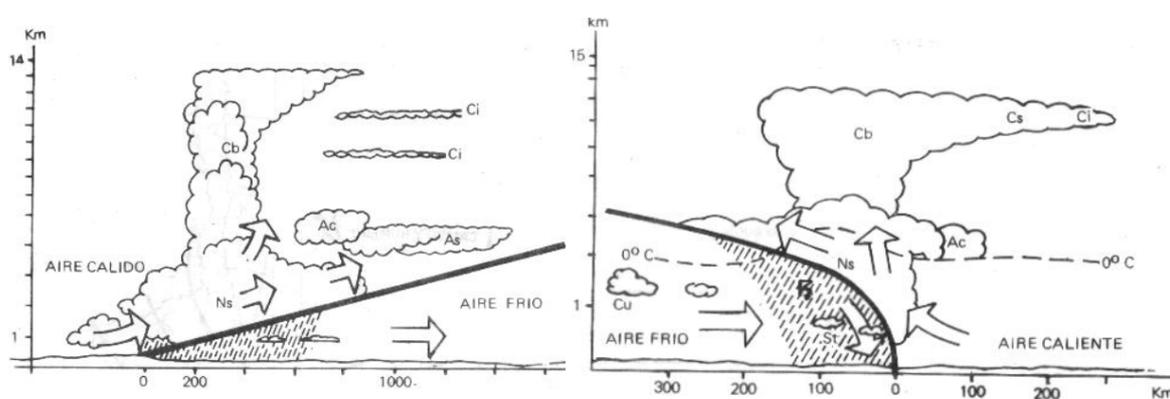


Figura 2 Diferentes frentes: Frío (izquierda), Cálido (derecha). Fuente: <https://www.canaltiempo21.com>

## 2.1.2 Precipitación

Una nube es un conjunto de finas partículas de agua en estado líquido o en estado sólido (cristales de hielo) que forman masas de espesor, color y formas variables.

Las gotas de agua que forman las nubes son esféricas y muy pequeñas (entre 0,004 y 0,1 mm). Estas gotas se encuentran suspendidas en el aire y sometidas a corrientes ascendentes y otras fuerzas, de tal forma que se encuentran en constante movimiento dentro de la nube, chocando unas con otras y agrupándose entre ellas. Según las condiciones atmosféricas existentes, se puede producir un aumento de su espesor hasta el punto de que su peso supere las fuerzas ascendentes y caigan hacia la tierra en forma de precipitación. Dicho término comprende entre otros la lluvia, el granizo, la nieve, el rocío y la escarcha.

### 2.1.2.1 Lluvia

La lluvia es la caída de gotas de agua que provienen de la condensación del vapor de agua en la atmósfera. El tamaño de las gotas de agua oscila entre 0.7 y 5mm de diámetro y estas caen a distintas velocidades según su peso. Como podemos observar en la Tabla 1 existen distintos tipos de lluvia según su intensidad.

Nombre	Descripción
<b>Llovizna</b>	Las gotas son pequeñas, con un diámetro < 0.5 mm y se presenta de una forma pulverizada, como flotando en el aire.
<b>Lluvia</b>	Si es continua, regular y el diámetro de sus gotas es > 0.5 mm
<b>Chubasco</b>	Si cae de golpe, con intensidad, y en un intervalo de tiempo pequeño.
<b>Tromba</b>	Si cae tan violenta y abundantemente que provoca riadas e inundaciones.

Tabla 1 Distintos tipos de lluvia según su intensidad.

### 2.1.2.2 Nieve

La nieve se forma cuando la temperatura es tan baja que el agua adquiere estado sólido, se produce cuando las gotas al caer atraviesan una capa de aire frío, por debajo de cero grados.

Cuando la temperatura al nivel del suelo es inferior a la de fusión, el conglomerado de cristales de hielo alcanza la superficie terrestre en forma de nieve. Cuando la temperatura es superior a 0° C, la nieve se funde y se convierte en lluvia.

### **2.1.2.3 Granizo**

Es la precipitación en forma de glóbulos de hielo que caen separados y con violencia sobre la superficie terrestre. El trozo de granizo está constituido por varios cientos de diminutos cristales de hielo. Su tamaño oscila entre unos milímetros y dos o más centímetros.

Se forman en el interior de los cumulonimbos, nubes de gran desarrollo vertical internamente formadas por una columna de aire cálido y húmedo que se eleva en forma de espiral rotatoria, es decir, existe un contraste entre las capas que la forman.

La formación del granizo se origina con la presencia de una partícula sólida; esta es arrastrada por fuertes vientos ascendentes dentro de la nube, a la que se le van adosando partículas de agua. Al ascender, se produce el enfriamiento de estas partículas, congelándose. Al llegar a la zona superior de la nube, el granizo cae hacia la tierra por su propio peso, en su caída, muchas de las capas de hielo que se formaron durante su ascenso, pueden descongelarse. Sin embargo, puede ser capturada nuevamente por otra corriente de aire ascendente y ser trasladada hacia las regiones altas de la nube. Este ciclo puede ocurrir varias veces, hasta que el granizo tome una dimensión y peso, que las corrientes ascendentes de aire dentro de la nube no tienen la fuerza suficiente para transportarlo, precipitándose así a tierra.

### **2.1.2.4 Rocío y escarcha**

Se originan directamente sobre la superficie terrestre. Si una masa de aire se enfría lo suficiente, alcanza una temperatura llamada punto de rocío, por debajo de la cual no puede mantener toda su humedad en estado de vapor y este se condensa, convirtiéndose en líquido. Si la temperatura es lo suficientemente baja se originan cristales de hielo.

Se puede distinguir entre tres tipos de rocío: como resultado del transporte hacia abajo de la humedad atmosférica condensada en superficies frías, por el

vapor de agua procedente del suelo y de las plantas y condensado en superficies frías, o como agua exudada por hojas.

La escarcha no es el rocío que se hiela, éste proviene directamente del vapor atmosférico sin pasar por el estado líquido y está formado por cristales de hielo entremezclados. De ahí que a este fenómeno también se le conozca por el nombre de helada.

### 2.1.3 La temperatura

La temperatura representa el estado termodinámico de un cuerpo y su valor está determinado por la dirección del flujo neto de calor entre dos masas, se dice que la masa que pierde calor transfiriéndolo al otro se encuentra a mayor temperatura. Para nosotros la temperatura está relacionada con los cinco movimientos del planeta y su posición respecto al Sol. La temperatura varía según distintos factores:

- Variación diurna: Cambio de temperatura entre el día y la noche, producido por la rotación de la Tierra.
- Variación estacional: Se debe a la inclinación del eje terrestre y el movimiento de traslación de la Tierra alrededor del sol. El ángulo de incidencia de los rayos solares varía, estacionalmente, en forma diferente para los dos hemisferios. El hemisferio norte es más cálido en los meses de junio, julio y agosto, en tanto que el hemisferio sur recibe más energía solar en diciembre, enero y febrero.
- Variación con la latitud: La mayor inclinación de los rayos solares en altas latitudes, hace que éstos entreguen menor energía solar sobre estas regiones, siendo mínima dicha entrega en los polos. Por el contrario, sobre el Ecuador los rayos solares llegan perpendiculares, siendo allí máxima la entrega energética.
- Variaciones con el tipo de superficie: Las grandes masas de agua tienden a minimizar los cambios de temperatura, mientras que los continentes permiten variaciones considerables en la misma.
- Variaciones con la altura: La radiación solar es absorbida por la tierra, la cual a su vez calienta por contacto las capas inferiores de la atmósfera, y estas

luego transmiten su calor a las capas más altas, en virtud de las corrientes de convección que se establecen. Así pues, en general, las capas bajas de la atmósfera se hallan a mayor temperatura que las situadas encima de ellas.

### **2.1.4 Humedad**

Nos referimos al término humedad cuando hablamos de la cantidad de vapor de agua que puede haber en un volumen determinado de aire. Existen distintas magnitudes cuando hablamos de humedad como son: razón de mezcla, humedad absoluta o humedad relativa, estas las explicaremos brevemente a continuación.

- Razón de mezcla: consiste en la relación entre la masa de vapor de agua y la masa de aire seco
- Humedad absoluta: se refiere a la cantidad total de vapor de agua que se encuentra en el aire
- Humedad relativa: es la cantidad de agua que hay en el aire con relación a la máxima que puede tener a una determinada temperatura.

Cuando el contenido de vapor de agua del aire permanece constante, la disminución de la temperatura del aire provoca un aumento de la humedad relativa y el aumento de la temperatura del aire provoca una disminución de la humedad relativa.

## **2.2 Efectos de las condiciones meteorológicas en señales de radiofrecuencia**

Los diferentes fenómenos meteorológicos afectan a la señal de radio a frecuencia de microondas de diferente forma. por ello en los siguientes apartados los estudiaremos por separados.

### **2.2.1 Atenuación por hidrometeoros**

Los efectos de los hidrometeoros deben tenerse en cuenta a frecuencias altas de varios GHz. Estos producen absorción y dispersión en las ondas radioeléctricas. El hidrometeoro más importante dentro de las comunicaciones

troposféricas es la lluvia. La atenuación que produce dicho hidrometeoro aumenta a frecuencias superiores de 5 GHz.

Existe un modelo para la atenuación específica producida por la lluvia a partir de valores de intensidad de lluvia conocidos explicado en la Recomendación P.838-3 del UIT-R (ITU, 2005).

Aunque la lluvia es la precipitación con más impacto y en la cual nos centraremos en este TFG, no deben descartarse otros hidrometeoros como la nieve y el granizo de los cuales hablamos anteriormente en los apartados 2.1.2.2 y 2.1.2.3. La precipitación en forma de nieve húmeda puede provocar mayores atenuaciones que la equivalente de lluvia y la debida al granizo puede ser importante para frecuencias por debajo de 2 GHz. Sin embargo, estos fenómenos son poco frecuentes.

En posteriores apartados se llevará a cabo un estudio de la atenuación específica de la lluvia al igual que detallaremos características específicas de ella ya que este TFG trata de la detección y cuantificación de lluvia a frecuencias de gigahercios.

### **2.2.2 Atenuación por nubes, niebla, arena y polvo.**

Las nubes, la niebla, la arena y partículas de polvo afectan a las ondas electromagnéticas de una forma muy parecida a los hidrometeoros. Las nubes y la niebla son pequeñas gotas de agua suspendidas en la atmósfera. Por otro lado, cuando se habla de atenuación por arena y polvo se suele centrar el estudio en tempestades de arena compuestas por nubes densas de partículas de polvo en suspensión.

Las nubes pueden dar lugar a valores apreciables de atenuación en frecuencias milimétricas. El problema está en la complejidad física del modelado de las nubes y la escasez de datos precisos. Para poder obtener la influencia de las nubes sobre la propagación deberíamos tener en cuenta multitud de características algunas de las cuales no están suficientemente definidas como por ejemplo la extensión vertical y horizontal de la nube. Por ello Se prefiere realizar los cálculos a partir del contenido de agua líquida en una columna vertical, este procedimiento es el incluido en la recomendación UIT-R P.840-6 (09/2013).

La niebla es la suspensión de gotas pequeñas de agua que producen una visibilidad de menos de 1 km. La atenuación por niebla es solo significativa en frecuencias superiores a 100 GHz.

En el caso de la arena y el polvo la atenuación solo será apreciable para concentraciones de partículas altas o humedad elevada, éstas condiciones sólo se dan en zonas tropicales desérticas o semidesérticas. Se puede evaluar en longitudes de onda centimétricas, gracias a la teoría de dispersión de Rayleigh y a la teoría de dispersión de Mie, longitudes de onda cortas. Las partículas de polvo pueden tener unas dimensiones típicas de entre 0,05  $\mu\text{m}$  y 0,08  $\mu\text{m}$ .

### **2.2.3 Atenuación por gases atmosféricos**

La atmósfera está compuesta por gases variables y permanentes. Los gases variables cambian su concentración en la atmósfera respecto a la posición y la altura a la que se midan dichos gases son el vapor de agua, el dióxido de carbono, el ozono y los aerosoles, siendo el vapor de agua el más importante. Los permanentes, como el oxígeno, el nitrógeno y el argón, aparecen en la atmósfera en proporciones más o menos constantes dentro de la troposfera.

La atenuación debida a los gases es causada sobre todo por el oxígeno y el vapor de agua. Dicha atenuación comienza a ser significativa a partir de una frecuencia de 10 GHz.

El vapor de agua produce una atenuación muy variable puesto que está estrechamente relacionada con la lluvia y la atenuación que provoca ésta. Presenta máximos de atenuación en torno a 22,3, 183,3 y 325 GHz.

La atenuación por oxígeno es mucho más constante, variando algo con los cambios de temperatura. Tiene un pico de absorción en 118,7 GHz y varios entre 50 a 70 GHz.

Para su cálculo existen distintos modelos recogidos en la Recomendación UIT-R P.676-12. (UIT-R, 2019), para la estimación de la atenuación causada por los gases mediante la suma de las rayas de absorción individuales, válida para la gama de frecuencias entre 1-1 000 GHz. En la misma recomendación podemos encontrar dos métodos aproximados simplificados, para evaluar la atenuación debida a los gases válida en la gama de frecuencias de 1-350 GHz.

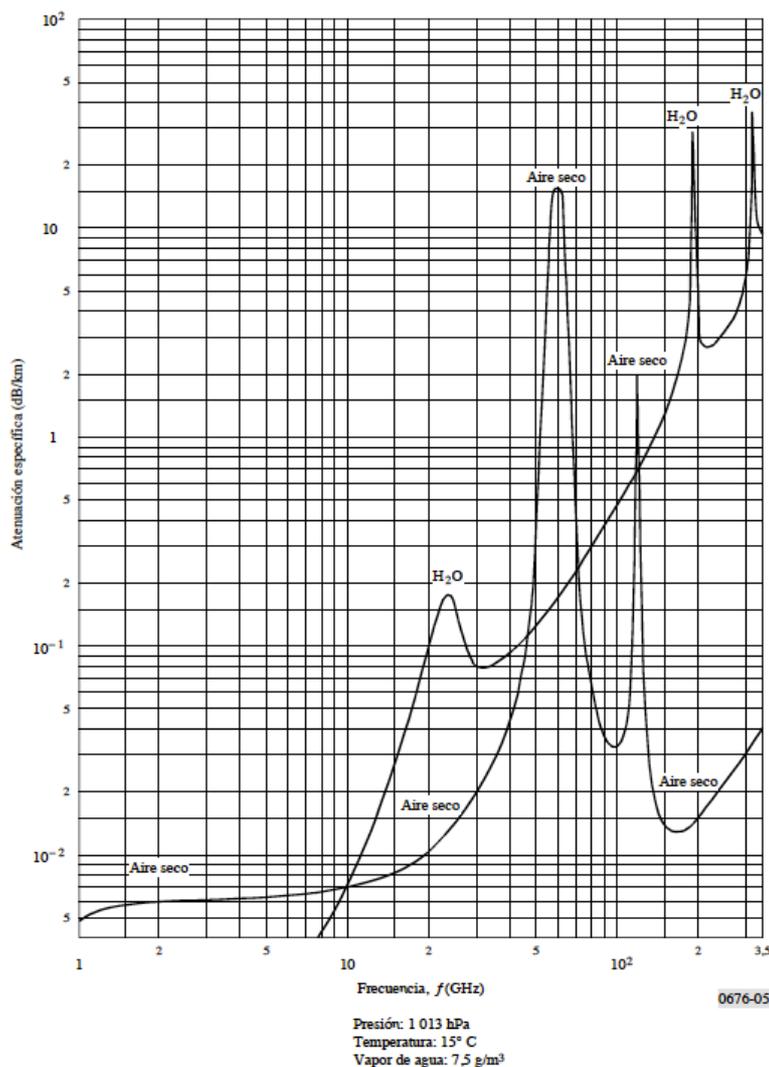


Figura 3 Atenuación específica debida a los gases atmosféricos. Fuente: (ITU-R', 1997)

### 2.3 Atenuación específica de la lluvia

Para el cálculo de la atenuación específica de lluvia se necesita información sobre la estructura horizontal y vertical de la precipitación. La estructura de la lluvia dependerá del área que cubra la precipitación y la duración de ésta, es decir, del tipo de lluvia, además de la intensidad y del tipo de clima. Por ello empezaremos describiendo los diferentes tipos de lluvia que existen.

Además, describiremos los parámetros físicos que caracterizan a las gotas de lluvia como son su tamaño, su forma y distribución, su velocidad de caída y su temperatura que son de gran importancia para el cálculo.

La altura donde se produce la lluvia es un parámetro de gran importancia en los modelos de predicción porque la atenuación total es casi proporcional a esta altura. Por ello lo trataremos en un apartado específico.

### 2.3.1 Tipos de lluvia

Aunque no siempre resulta sencillo definir un límite a la intensidad de precipitación que puede provocar cada lluvia ni su extensión horizontal, las precipitaciones por lluvia se clasifican principalmente en cuatro tipos (ITU., 1995):

- Precipitación estratiforme: ocupa extensiones muy grandes. Su estructura evoluciona con relativa lentitud. Descarga en forma de lluvias de baja intensidad y gran duración. Se suele producir en nubes tipo nimbostratos.
- Precipitación convectiva: corresponden a zonas localizadas. Se produce por un ascenso de aire húmedo y cálido como consecuencia del calentamiento de la superficie terrestre. Se forman nubes de desarrollo vertical como cúmulos y cumulonimbos. Generan lluvias intensas en un periodo relativamente corto. En el radar se visualizan como regiones en forma de columnas casi verticales de mayor reflectividad.
- Lluvias monzónicas: son combinaciones de bandas de precipitación convectiva intensa seguidas de lluvia estratiforme. Suelen tener una extensión de unos 50 km y centenares de km de largo con una duración de varias horas.
- Tormentas tropicales: parecidas a las anteriores, aunque estas son más extensas, de unos cientos de km, y las bandas están dispuestas en forma de espiral.

En climas templados, como es el estudiado en este TFG, sólo existen dos tipos de lluvias, estratiforme y convectiva ya que los otros dos raramente se dan. Sin embargo, no siempre se dan de manera aislada, sino que muchas veces el fenómeno de precipitación es una combinación de ambos tipos. En este TFG trabajaremos con precipitaciones totales, que será la suma de la cantidad de lluvia estratiforme y de la convectiva.

### 2.3.2 Altura de la lluvia

La altura de la lluvia supone uno de los factores más importantes que provocan la atenuación de la señal; es el límite para la extensión vertical de la atenuación. La altura de la lluvia no coincide con la isoterma (línea que une los vértices, en un plano cartográfico, que presentan las mismas temperaturas en la unidad de tiempo considerada) a 0 °C ya que en la realidad sigue existiendo precipitación por lluvia a una altura algo superior, pero están directamente relacionadas. La isoterma a 0 °C es muy variable según las características atmosféricas del lugar, pero se puede medir fácilmente por radiosondas.

Hay modelos que consideran la altura de la isoterma a 0 °C como la máxima altura de la lluvia para precipitaciones estratiformes, ya que por encima de esta altura se supone que las precipitaciones son en forma de hielo y nieve, las cuales no afectan de manera significativa a la atenuación para frecuencias inferiores a 60GHz.

La Recomendación UIT-R P.839-4 (ITU\_R, 2013) , para obtener la altura de la lluvia durante los periodos de precipitación, propone usar la siguiente fórmula, donde  $h_0$  es el valor medio de la altura de la isoterma a 0 °C por encima del nivel del mar (km).

$$h_R = h_0 + 0,36 \text{ km} \quad (2.1)$$

El parámetro  $h_0$  se calcula a través de un fichero de texto que proporciona la UIT y que contiene estos valores para una resolución de 1,5 °C. Junto con los ficheros de longitud y latitud se realiza una interpolación bilineal para obtenerla en el punto necesario.

Por otro lado, en las zonas del mundo donde no se dispone de información específica, propone unos mapas mundiales estacionales para el cálculo de la isoterma 0 °C. Dichos mapas ofrecen un valor medio anual de la altura de la lluvia sobre el nivel del mar para cada emplazamiento. Sin embargo, esta no es la altura de la lluvia, sino que la recomendación asume que hay lluvia hasta la isoterma -2 °C.

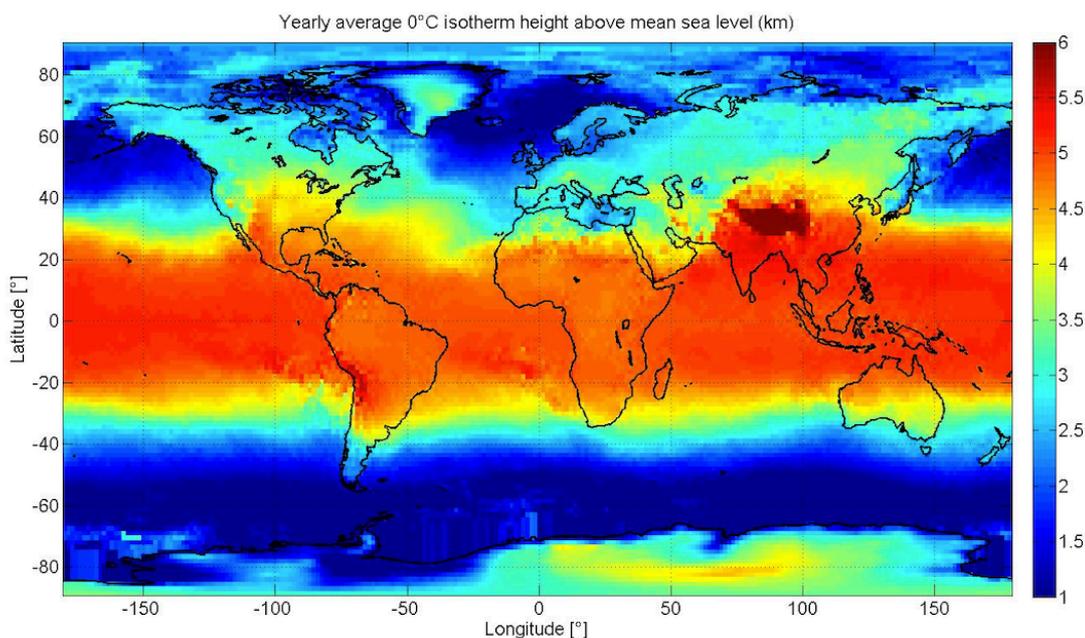


Figura 4 Media Anual de la isoterma a 0°C por encima del nivel del mar (km). Fuente: (ITU\_R, 2013)

### 2.3.3 Tamaño, forma y orientación de las gotas

Los tamaños, formas y orientaciones de las gotas de agua pueden variar dentro de una lluvia. Sin embargo, la distribución del tamaño de las gotas resulta ser bastante estable, siendo lo que varía la intensidad de la precipitación. Cuanto más se aproxime el tamaño de la gota a la longitud de onda, mayor será la atenuación producida por la gota. Por debajo de 10 GHz el efecto de gotas pequeñas es despreciable, cosa que no ocurre a partir de esta frecuencia.

La mejora de la instrumentación en las últimas décadas ha permitido determinar con mayor exactitud el tamaño de las gotas. Se han propuestos varios modelos basados en distribuciones log-normal, gamma o Weibull. Pero las medidas de distribución de gotas por encima de 0.5 mm parecen confirmar la precisión de la distribución de Laws y Parson que fueron los que realizaron la primera determinación experimental del tamaño de las gotas. Además, cuanto más grandes sean las gotas mayor atenuación producirán en la señal.

La forma de las gotas de lluvia está determinada por las fuerzas aerodinámicas que actúan sobre ellas a medida que caen hacia la tierra. Estas adquieren una forma esferoidal debido al efecto de la gravedad y la tensión superficial del agua. Suelen tener un eje de simetría casi vertical, aunque pueden inclinarse unos grados debido a la fuerza horizontal del viento.

### 2.3.4 Velocidad de caída

Este parámetro es importante pues determina la intensidad de la lluvia y se relaciona con el tamaño de las gotas. Se llama velocidad terminal a la velocidad límite que puede alcanzar una gota de agua.

El modelo de Gunn y Kinzer (Gunn, 1949) relaciona la velocidad terminal de las gotas con su tamaño con el aire en calma y manifiesta que la velocidad terminal aumenta con el radio de la gota hasta un máximo de 9 m/s cuando mide 2,5 mm aproximadamente.

### 2.3.5 Temperatura

En general, la temperatura de las gotas está en torno a la del aire que las rodea. Pueden existir mezclas de hielos y agua por debajo de la altura de la isoterma 0°C en regiones de corrientes descendentes y pequeñas gotas de lluvia por debajo de 0°C en regiones de corrientes ascendientes.

Esta temperatura influye en la parte compleja del índice de refracción del agua y varía con la frecuencia siendo máxima para frecuencias entre 10 y 100 GHz. Por debajo de 20 GHz la atenuación producida en la onda será más importante, pudiendo ser causa de divergencias de hasta un 20% en la atenuación específica [Velasco].

### 2.3.6 Fórmula de la atenuación específica de la lluvia

La atenuación específica  $\gamma_R$  (dB/km) debida a la lluvia puede calcularse a partir de la Recomendación UIT-R P.838-3 (ITU, 2005). Dicha atenuación se relaciona con la intensidad de lluvia  $R$  (mm/h) mediante la siguiente expresión exponencial:

$$\gamma_R = kR^\alpha \text{ dB/km} \quad (2.2)$$

donde  $k$  y  $\alpha$  son variables que dependen de la frecuencia, de la polarización de la onda electromagnética y del tamaño de la gota de agua, su forma y su densidad volumétrica y se obtienen a partir de las fórmulas (2.3) y (2.4)

$$\log_{10} K = \sum_{j=1}^4 a_j \exp \left[ - \left( \frac{\log_{10} f - b_j}{c_j} \right)^2 \right] + m_k \log_{10} f + c_k \quad (2.3)$$

$$\alpha = \sum_{j=1}^5 a_j \exp \left[ - \left( \frac{\log_{10} f - b_j}{c_j} \right)^2 \right] + m_\alpha \log_{10} f + c_\alpha \quad (2.4)$$

Donde:

$f$ : frecuencia (GHz)

$k$ : puede ser  $k_H$  o  $k_V$

$\alpha$ : puede ser  $\alpha_H$  o  $\alpha_V$

Los restantes coeficientes necesarios para estas fórmulas los podemos encontrar en las Tablas 2 y 3, para polarización horizontal y vertical.

<b>j</b>	<b>a<sub>i</sub></b>	<b>b<sub>i</sub></b>	<b>c<sub>i</sub></b>	<b>m<sub>k</sub></b>	<b>c<sub>k</sub></b>	<b>m<sub>α</sub></b>	<b>c<sub>α</sub></b>
1	-5,33980	-0,10008	1,13098	-0,18961	0,71147	-----	-----
2	-0,35351	1,26970	0,45400				
3	-0,23789	0,86036	0,15354				
4	-0,94158	0,64552	0,16817				
1	-0,14318	1,82442	-0,55187	-----	-----	0,67849	-1,95537
2	0,29591	0,77564	0,19822				
3	0,32177	0,63773	0,13164				
4	-5,37610	-0,96230	1,47828				
5	16,17210	-3,29980	3,43990				

Tabla 2 Coeficientes de las fórmulas (2.3) y (2.4) para polarización horizontal:

<b>j</b>	<b>a<sub>i</sub></b>	<b>b<sub>i</sub></b>	<b>c<sub>i</sub></b>	<b>m<sub>k</sub></b>	<b>c<sub>k</sub></b>	<b>m<sub>α</sub></b>	<b>c<sub>α</sub></b>
1	-3,80595	0,56934	0,81061	-0,16398	0,63297	-----	-----
2	-3,44965	-0,22911	0,51059				
3	-0,39902	0,73042	0,11899				
4	0,50167	1,07319	0,27195				
1	-0,07771	2,33840	-0,76284	-----	-----	-0,05374	0,83433
2	0,56727	0,95545	0,54039				
3	-0,20238	1,14520	0,26809				
4	-48,29910	0,79167	0,11623				
5	48,58330	0,79146	0,11648				

Tabla 3 Coeficientes de las fórmulas (2.3) y (2.4) para polarización vertical:

En la recomendación también podemos encontrar unas gráficas que representan visualmente la variación de los parámetros  $k$  y  $\alpha$  para frecuencias entre 1 y 1000 GHz con polarización horizontal y vertical.

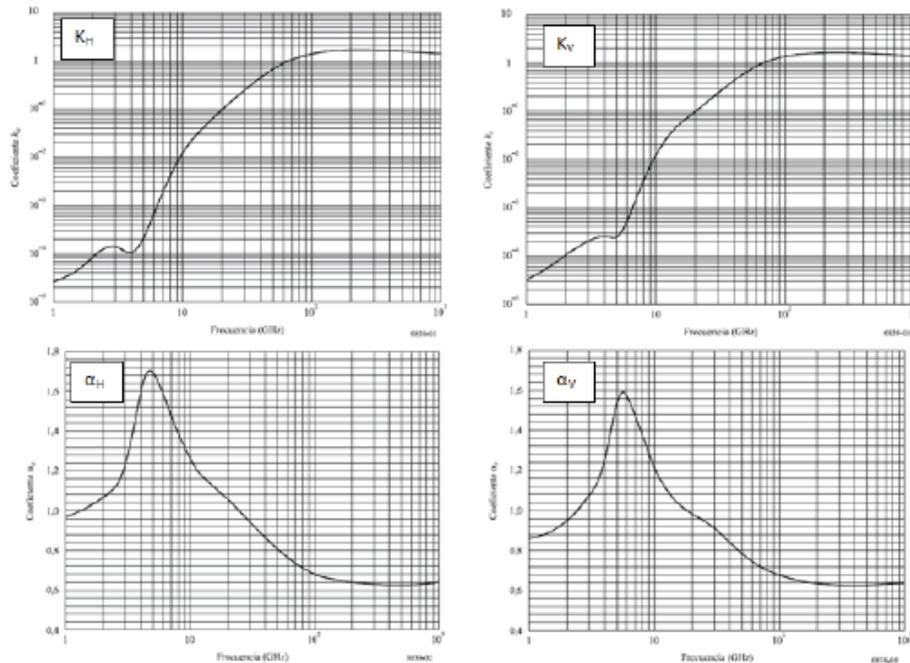


Figura 5 Variación de los parámetros  $k$  y  $\alpha$ . Fuente: (ITU, 2005)

Cabe destacar que los coeficientes  $k$  y  $\alpha$  anteriormente dados son solo para polarizaciones verticales u horizontales, pero si se quiere calcular para casos con una polarización lineal y circular, y para cualquier geometría del trayecto, los coeficientes de la ecuación (2.2) se calculan mediante las siguientes expresiones haciendo uso de las ecuaciones (2.3) y (2.4):

$$k = [k_H + k_V + (k_H - k_V) \cos^2 \theta \cos 2\tau] / 2 \quad (2.5)$$

$$\alpha = [k_H \alpha_H + k_V \alpha_V + (k_H \alpha_H - k_V \alpha_V) \cos^2 \theta \cos 2\tau] / 2k \quad (2.6)$$

donde  $\theta$  es el ángulo de elevación del trayecto y  $\tau$  es el ángulo de inclinación de la polarización con respecto a la horizontal (por ejemplo,  $\tau = 45^\circ$  para la polarización circular).



## Capítulo 3 - Adquisición de datos y Preprocesado

En este capítulo, describimos las distintas fuentes de datos al igual que la adquisición de ellos. Asimismo, se describen los pasos realizados en el tratamiento previo a su utilización.

### 3.1 Fuentes de datos

Trabajaremos con dos tipos de datos. Por un lado, los datos meteorológicos, son los recogidos por las estaciones meteorológicas, concretamente en este TFG trabajaremos con los datos recogidos de la Agencia Estatal de Meteorología (AEMET, 2020) como la que podemos observar en la Figura 6. Y, por otro lado, los datos de radioenlaces son los recogidos por los equipos de las antenas emisoras, en este TFG específicamente proporcionados por el Gobierno de Canarias, Dirección General de Telecomunicaciones, obtenidos dentro de su plan de Red de Emergencias y Seguridad de Canarias, también conocido como RESCAN.

Es muy importante tener en cuenta que los datos de meteorología son tomados cada hora en punto, es decir, se da un dato por hora, mientras que los datos de radioenlaces son muestras tomadas cada quince minutos empezando por la hora en punto, cuatro muestras cada hora. Dado que esas cuatro muestras equivalen a una muestra de meteorología será necesario juntar las cuatro muestras de radioenlaces en una sola o extrapolar y dividir en cuatro la muestra horaria de meteorología, según proceda. En cualquier caso, en nuestro sistema de aprendizaje la referencia es horaria porque esa es la periodicidad de los datos meteorológicos.



Figura 6 Estación meteorológica de la AEMET situada en Niembro (Asturias). Fuente: AEMET

Es importante resaltar que los datos meteorológicos de los que disponemos constituyen un gold standard, eso es, una fuente de referencia que es la mejor disponible en condiciones razonables. Esto viene del hecho de que dichas estaciones no están situadas exactamente en el vano de cada radioenlace, sino en su cercanía, ya que el emplazamiento de las estaciones y de los vanos se decidieron en contextos diferentes. Esta situación no es nueva y está referenciada en la literatura (Ho, 2005) (Hagit Messer, 2006) .

### 3.1.1 Región de Trabajo

Aunque poseemos datos suministrados por radioenlaces y estaciones meteorológicas situados en las islas de la Gomera, el Hierro, Tenerife, Fuerteventura y Gran Canaria, como se muestra en la Figura 7, en este TFG solo trabajaremos con datos obtenidos en la isla de Gran Canaria y Tenerife.



Figura 7 Estaciones meteorológicas de la AEMET en el archipiélago canario. Fuente: AEMET

Hemos seleccionado cuatro radioenlaces de cada isla como se puede observar en la Tabla 4. Estos radioenlaces están localizados en zonas con densidad de población alta debido al contexto (propósito) del proyecto VIMETRI-MAC.

Isla	Extremo A	Extremo B
Gran Canaria	Agate	Montaña_Gáldar
Gran Canaria	Agüimes	Tabaiba
Gran Canaria	Monteluz	Hospital
Gran Canaria	Moriscos	Teror
Tenerife	Anocheza	Birmagen
Tenerife	Anocheza	El_Reverón
Tenerife	Suculum	Bailadero
Tenerife	Suculum	Casas_Cumbre

Tabla 4 Radioenlaces seleccionados para Gran Canaria y Tenerife.

### 3.3 Radioenlaces

En la Figura 8 se ve representada la señal de potencia del radioenlace Monteluz-Hospital Militar desde el 8 de octubre de 2015 a las 22:00 hasta el 1 de noviembre de 2015 a las 21:00. Se puede observar que el día 22 de octubre de 2015 pasadas las 8:00 horas el nivel de potencia disminuyó desde unos -44 dBm hasta -56,68 dBm, esto es, una atenuación de 12'68 dBm. Esta atenuación puede ser interpretada como indicadora de un posible evento de lluvia. Retomaremos este hecho más adelante en el apartado (3.4).

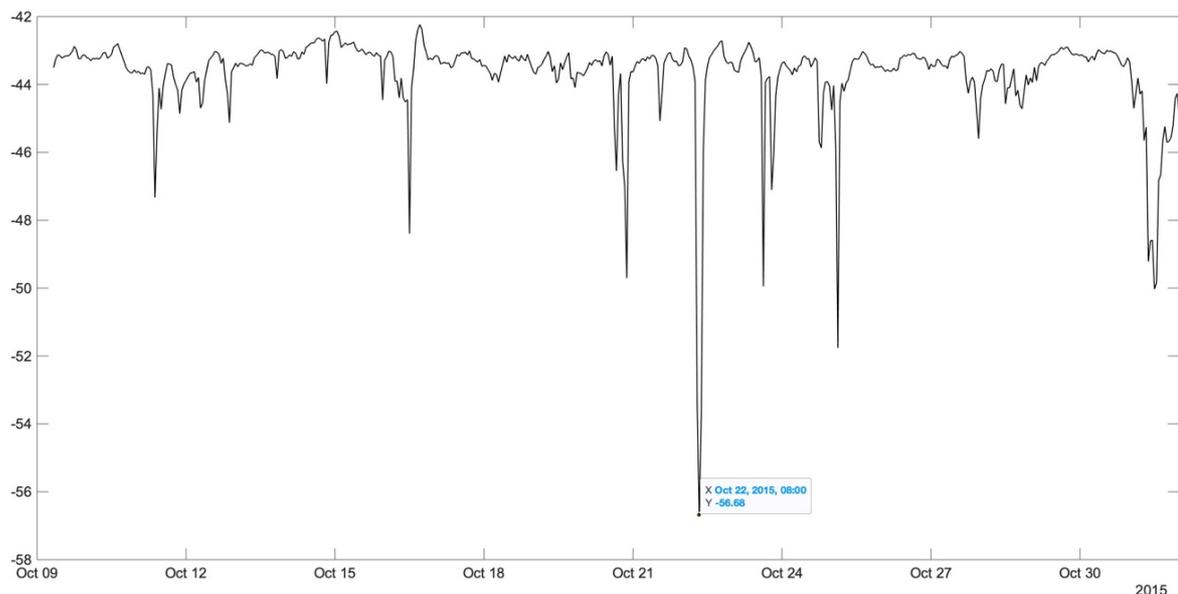


Figura 8 Señal de potencia recibida en el radioenlace Monteluz-Hospital Militar.

### 3.3.1 Características específicas

Los radioenlaces con los que trabajamos pueden tener una sola antena o varias antenas para prevenir desvanecimientos que pongan el radioenlace en corte. Actualmente, tenemos datos de 134 vanos de radioenlaces diferentes.

A continuación, presentamos las características generales de los radioenlaces de los que hemos obtenido los datos.

Trabajan todos a frecuencias de gigahercios, en las bandas de frecuencia de 6 a 38 GHz, que son frecuencias muy sensibles a los fenómenos meteorológicos que nos interesan ya que se producen atenuaciones vinculadas a estos (por ejemplo, atenuación por lluvia). Además, la mayoría están por encima de los 10 GHz, que es donde se presenta mayor sensibilidad. Estos radioenlaces muestran una distribución de polarizaciones horizontales y verticales bastante igualitaria.

Otra característica muy importante es la longitud de los vanos puesto que las atenuaciones debidas a las precipitaciones se expresan en dB/Km. La longitud de los vanos va desde los 2 kilómetros hasta los 64 kilómetros, aunque la mayoría se centran entre los 4 y 12 kilómetros de longitud.

La mayoría de las antenas presentan un diámetro menor a 70 centímetros, aunque hay algunos con un diámetro superior hasta alcanzar los 1.2 metros. Además, debido a las diferentes altitudes con las que nos encontramos debemos tener presente la altura a la cual está instalada cada antena, aunque la mayoría se encuentran entre los 0 y los 1400 metros, hay excepciones como algunos situados a 3500 metros, en los que se puede dar que las capas nubosas se sitúen en la horizontal respecto a la antena, o a 0 metros, es decir con cota a nivel del mar, que pueden presentar problemas como brumas, partículas de sal en el aire, etc.

La diferencia de alturas transmisor-emisor es importante pues afecta a las inclinaciones de las antenas y su relación respecto al sol incidente, a la acumulación de humedad en la antena en mayor o menor proporción o las distintas capas de aire que se han de atravesar con sus respectivas densidades (Guerra Moreno, 2017).

En la Tabla 5 se presentan las características técnicas de los 8 radioenlaces de estudio.

<b>Extremo A</b>	<b>Extremo B</b>	<b>GHz</b>	<b>Polarización</b>	<b>Antena diam. (m)</b>	<b>Longitud vano (km)</b>
Agate	Montaña_Gáldar	22,0708	Horizontal	0,6	64,195
Agüimes	Tabaiba	23,2800	Vertical	0,8	12,832
Monteluz	Hospital	22,0180	Vertical	0,3	5,696
Moriscos	Teror	23,1140	Vertical	0,3	2,733
Anocheza	Birmagen	22,0780	Vertical	0,8	17,976
Anocheza	El_Reverón	23,0860	Vertical	0,6	15,978
Suculum	Bailadero	22,0708	Horizontal	0,3	4,565
Suculum	Casas_Cumbre_B	23,0580	Horizontal	0,6	5,434

*Tabla 5 Características técnicas de los radioenlaces de estudio en este TFG*

### 3.3.2 Adquisición de los datos de los radioenlaces

Los datos de radioenlaces que utilizaremos pertenecen a la red de emergencias y seguridad de Canarias (RESCAN). Son archivos .txt que contienen una serie de parámetros:

- Fecha y hora: Instante de la captura de datos, cada cuarto de hora.
- SNR: Es la relación señal a ruido dada en dB.
- RSL\_mean: partiendo de la información de las hojas de características, podemos apoyarnos en que cada segundo se recibe 10 RSL. El ODU dará un valor de RSL cada 15 minutos,  $10(\text{muestras/s}) * 60(\text{s}) * 15(\text{minutos}) = 9000$  muestras de RSL. Se puede obtener el valor de RSL\_mean de dos formas principalmente: un valor intermedio entre el máximo y el mínimo nivel entre las 9000 muestras o un valor promedio obtenido a partir de las 9000 muestras. En los equipos con los que hemos trabajado la forma de obtener la RSL correspondiente es la segunda, es decir, el promedio de las 9000 muestras.
- ODU\_Temp: Es la temperatura, en °C.
- ODU Input Voltage: Es el voltaje interior, en dB.

De todos ellos los que vamos a utilizar son: fecha, hora y 15 min RSL mean (datos cada 15 minutos a lo largo de los cuales se realizan mediciones cada 100ms). Sin embargo, se introducirán todos los valores en la base de datos. Estos radioenlaces se basan en el estándar Terrestrial Trunked Radio (TETRA) el cual explicaremos a continuación.

## TETRA

El estándar TETRA es el estándar de radio digital aceptado para las comunicaciones críticas. Es un estándar abierto, centrado en satisfacer las necesidades de comunicaciones críticas en el mercado de la Seguridad Pública y agencias gubernamentales de Seguridad (Valencia, 2020). Es un estándar del Instituto Europeo de Estándares de Telecomunicación (ETSI) y nace por decisión de la Unión Europea con el objeto de poder utilizarse para garantizar comunicaciones críticas seguras, confiables y robustas en sectores críticos como lo son servicios de emergencias (policía, bomberos, ambulancias, etc.).

### 3.3.3 Localización de vacíos

Para poder obtener la señal de niveles de potencia, señal cuyos valores corresponden a niveles de potencia dados en dBm, o mW cuando proceda, realizaremos un tratamiento sobre los datos de los radioenlaces comprendido en varias etapas.

#### Localización de GAPs

Los GAPs son vacíos de datos, esto es, intervalos en los que no hay disponible información acerca de los niveles de RSL. Para la localización de estos guardamos en un vector la fecha de inicio y fin de la interrupción temporal. También guardamos la fecha de inicio y fin de nuestros datos y el número de interrupciones totales (González Gómez, 2017). En la Figura 9 se pueden apreciar tres periodos de tiempo en los que se han producido GAPs, uno claramente diferenciable ocurre entre el 10 y el 18 de marzo de 2016 donde no se registra ningún tipo de señal. Además, hay uno entre el 3 de febrero y el 4 de febrero, y otro puntual el 27 de marzo. Estas tipologías son bastante comunes.

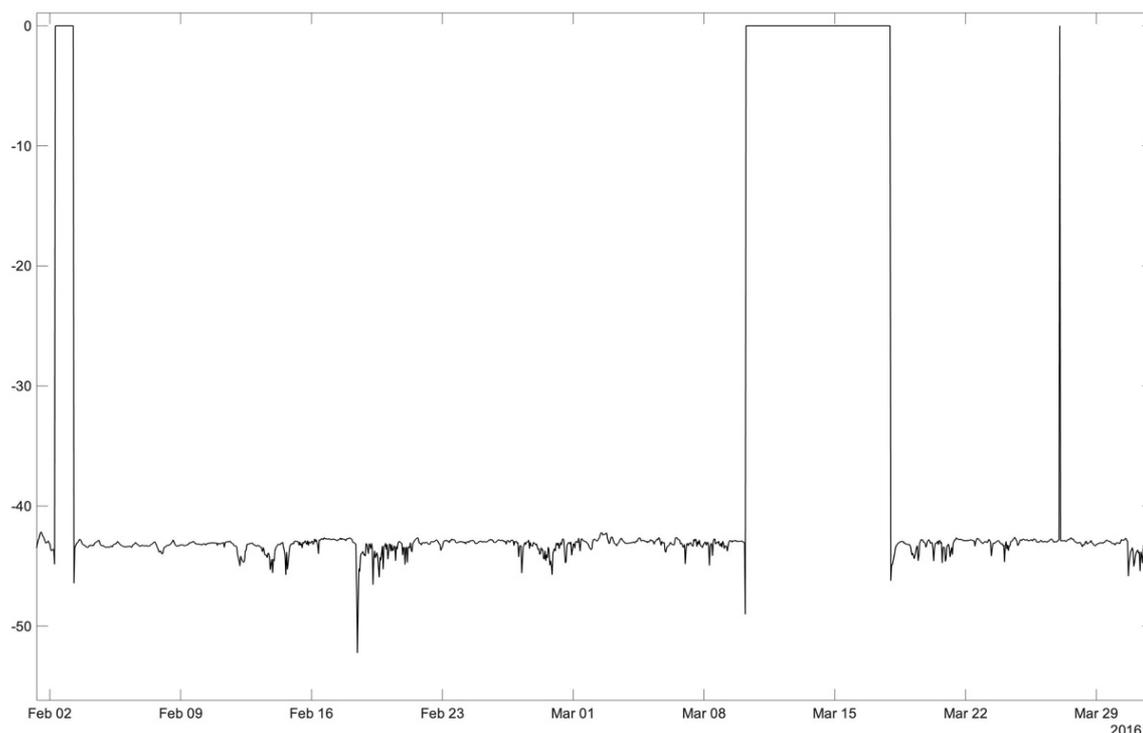


Figura 9 Señal de potencia en la que se han producido GAPs

### 3.4 Estaciones meteorológicas

La Figura 10 muestra los niveles de precipitación recogidos en la estación meteorológica de Las Palmas de Gran Canaria Tafira-Zurbarán entre el 8 de octubre de 2015 a las 22:00 y el 1 de noviembre de 2015 a las 21:00. Se puede observar que el día 22 de octubre de 2015 sobre las 9:00 horas el nivel de lluvia acumulado durante la hora anterior fue de 29 mm/h. Retomamos aquí lo expuesto al comienzo de (3.3) donde indicábamos que una atenuación de 12'68 dBm podría ser indicadora de un evento de lluvia.

Vamos a entrar en un poco más de detalle sobre las Figuras 8 y la 10. En los siguientes instantes se producen atenuaciones significativamente diferenciables a la vez que se recoge nivel de precipitación. El día 16 de octubre de 2015 a las 12:00 se produce una atenuación que llega hasta los -48,39 dBm, y el mismo día a la misma hora se recoge un nivel de precipitación de 6,2 mm/h. Otro caso es el del 20 de octubre del mismo año en el que a las 21:00 se registra una atenuación que llega a los -49,71 dBm y a la misma hora el nivel de lluvia acumulado marcado por el pluviómetro es de 5,4 mm/h. De lo anterior, bien podríamos inferir que hay una correlación entre eventos de lluvia y atenuaciones en su radioenlace asociado,

hecho este ya está recogido en las normas ITU (ITU, 2005) donde se describe como es la atenuación debida a lluvia. En el apartado 3.5 lo veremos con más detalle.

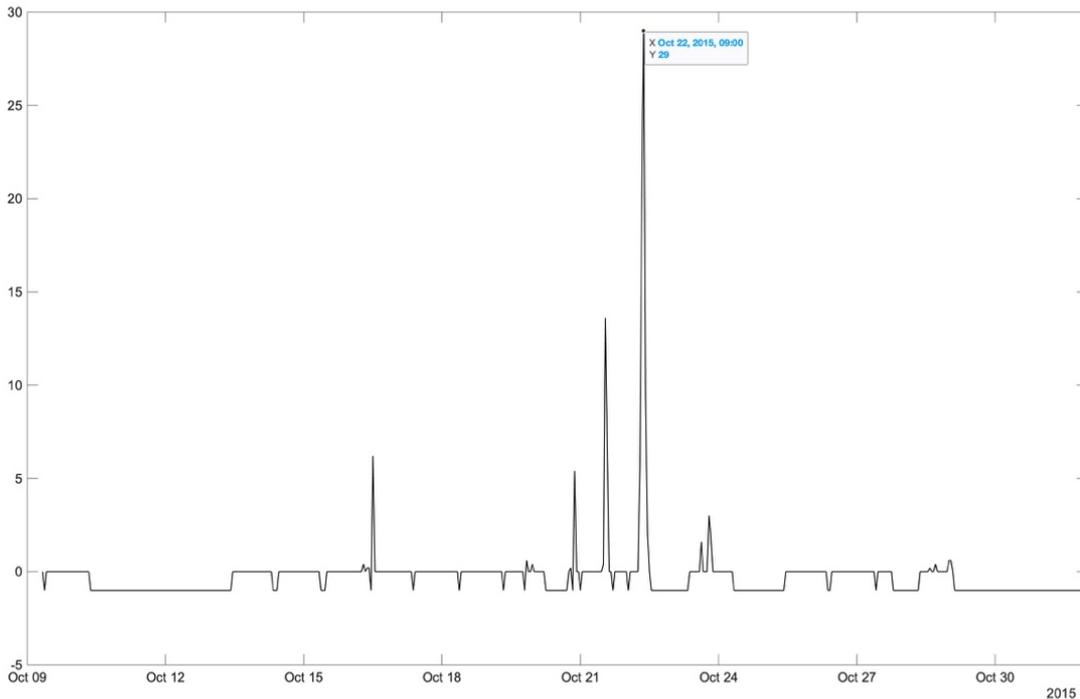


Figura 10 Nivel de lluvia acumulada en la estación Las Palmas de GC Tafira-Zurbarán

### 3.4.1 Características específicas

Los instrumentos utilizados para la obtención de datos suministrados por cada una de las estaciones meteorológicas son distintos de unas estaciones a otras debido a que utilizan modelos de aparatos distintos. Dejando a un lado estas diferencias todas poseen sensores de temperatura y humedad relativa, anemómetro, veleta y pluviómetro.

A continuación, haremos una breve descripción de cada uno de ellos. El pluviómetro, como el Vaisala Rain Gauge RG13 que podemos observar en la Figura 11, utiliza un mecanismo de balancín para producir un cierre de contacto cada vez que recibe una cantidad pequeña predeterminada de lluvia (0.1 mm / 0.2 mm / 0.5 mm). El mecanismo consiste en un cubo dividido pivotado a su centro, la lluvia se almacena en la mitad superior. Cuando está lleno, el mecanismo se inclina y descarga el agua almacenada, lo que permite a la otra mitad del cubo comenzar a llenarse (RG13/RG13H, 2020).



Figura 11 Pluviómetro: Vaisala Rain Gauge RG13. Fuente: <https://www.vaisala.com>

Los sensores de temperatura como por ejemplo PT 100 utilizado en la estación C659H San Cristóbal, estación de la Sociedad Española de Aplicaciones Cibernéticas (SEAC) (Cibernéticas, 2020), son un tipo específico de detector de temperatura RTD (detector de temperatura por resistencia). Están fabricados con platino con una resistencia eléctrica de 100 ohmios a una temperatura de 0 °C.

El sensor de humedad relativa (RHS) mide la humedad relativa en el rango de 0 a 100% y una resolución de 0,1% de humedad relativa.

El anemómetro es un instrumento de medida compuesto por 3 cazoletas de forma cónica inyectadas sobre un eje de acero inoxidable. El transductor consiste en un par de led y fototransistor que proporciona una salida de pulso con frecuencia proporcional a la velocidad del viento. Entre más rápida sea la respuesta y más bajo sea el momento de inercia y el umbral más precisa será la medida. Precisión de 0.1 m/s.

La veleta suele ser de aluminio y de acero, unida a un eje que incorpora un imán diametral que actúa sobre un sensor de posición giratorio dando una salida equivalente a la dirección indicada por la veleta. Suelen poseer una respuesta rápida y un bajo umbral. Existen distintos tipos de salidas.

### 3.4.2 Adquisición de los datos meteorológicos

Durante el periodo bajo estudio, los datos meteorológicos han sido adquiridos del portal web de la Agencia Estatal de Meteorología (AEMET, 2020). Como la obtención de esos datos solo es posible a través de un navegador, el cual

no ofrece la posibilidad de guardarlos, tendríamos que estar copiando y pegando esa información de manera manual. Por ello, se ha utilizado una técnica denominada Web scraping, un algoritmo que simula el uso del navegador por un ser humano, encargado de visitar la página, localizar los datos deseados y enviarlos a una dirección donde se almacena en forma de archivo de texto para su procesamiento y carga en nuestra base de datos meteorológica. Una opción de las más utilizadas es analizar sintácticamente el código fuente de la web a la que deseamos hacer scraping para que nuestro software sepa qué buscar (González Gómez, 2017).

Si quisiéramos obtener los datos de las últimas 24 horas de la estación meteorológica de Tafira solo tendríamos que acceder a la siguiente dirección web (AEMET, 2020b).

En la Tabla 6 encontramos un ejemplo de lo que encontraríamos en la página web, diversos parámetros de los datos recogidos en la estación meteorológica elegida:

- Fecha y hora: Instante de la captura de datos, cada hora en punto.
- Temperatura (en grados Celsius, °C): Es la temperatura del aire.
- Velocidad del viento (km/h): Velocidad media del viento en los 10 minutos anteriores a la hora indicada.
- Dirección del viento (Wdir): Dirección media del viento, en los diez minutos anteriores a la hora indicada. Los ocho rumbos en los que se da la dirección son: norte, sur, este, oeste, noreste, noroeste, sureste y suroeste.
- Velocidad racha (km/h): Velocidad máxima del viento, en los 60 minutos anteriores a la hora indicada.
- Dirección de racha (ocho rumbos): Dirección de donde viene el viento, en el instante de la racha.
- Precipitación de lluvia (mm): Precipitación acumulada en la última hora.
- Humedad relativa del aire (%): Expresada en porcentaje.

Fecha y hora	Temp (°C)	Vel. viento (km/h)	Dirección viento	Racha (km/h)	Dirección de racha	Precipitación (mm)	Humedad (%)
08/07/2020 17:00	26.9	10	Noroeste	26	Noroeste	0.0	53
08/07/2020 16:00	24.9	13	Noroeste	33	Noroeste	0.0	67
08/07/2020 15:00	23.7	16	Noroeste	28	Noroeste	0.0	75
08/07/2020 14:00	23.7	15	Noroeste	27	Noroeste	0.0	76
08/07/2020 13:00	23.8	15	Noroeste	30	Oeste	0.0	74
08/07/2020 12:00	23.8	13	Noroeste	28	Noroeste	0.0	74
08/07/2020 11:00	22.0	16	Noroeste	28	Noroeste	0.0	81
08/07/2020 10:00	21.1	14	Noroeste	25	Noroeste	0.0	87
08/07/2020 09:00	19.9	15	Noroeste	30	Noroeste	0.0	93
08/07/2020 08:00	19.4	16	Oeste	32	Oeste	0.0	95
08/07/2020 07:00	19.3	18	Noroeste	30	Oeste	0.0	95
08/07/2020 06:00	19.2	8	Oeste	24	Oeste	0.0	95
08/07/2020 05:00	19.2	11	Oeste	21	Oeste	0.0	95
08/07/2020 04:00	19.4	6	Oeste	19	Oeste	0.0	95
08/07/2020 03:00	19.5	9	Oeste	24	Oeste	0.0	95
08/07/2020 02:00	19.7	12	Oeste	24	Oeste	0.0	95
08/07/2020 01:00	20.0	9	Oeste	26	Oeste	0.0	95
08/07/2020 00:00	20.1	11	Oeste	24	Oeste	0.0	95
07/07/2020 23:00	20.2	16	Noroeste	29	Oeste	0.0	94
07/07/2020 22:00	20.5	14	Noroeste	30	Noroeste	0.0	93
07/07/2020 21:00	20.9	14	Noroeste	30	Oeste	0.0	90
07/07/2020 20:00	21.6	14	Noroeste	31	Noroeste	0.0	88
07/07/2020 19:00	22.8	19	Noroeste	31	Noroeste	0.0	82
07/07/2020 18:00	24.1	18	Noroeste	31	Noroeste	0.0	78

Tabla 6 Datos de la estación meteorológica de Tafira.

Los datos que utilizaremos en este TFG son: fecha, hora y precipitación de lluvia, no obstante, se dispone de todos en la base de datos.

### 3.4.3 Tratamiento de los datos meteorológicos

Para poder utilizar los datos adquiridos realizaremos la lectura desde la base de datos y posteriormente un estudio para localizar los espacios de tiempo en los que no los hay. Al igual que ocurría en los radioenlaces, los hemos denominado

GAPs. Un ejemplo de esto se puede observar en la Figura 12 que corresponde a la estación meteorológica de Las Palmas de Gran Canaria Tafira-Zurbarán entre los días 2 y 29 de febrero de 2016, y en el eje vertical se dan los niveles de lluvia medida en mm/h. Vemos que hay diversas interrupciones en las que no se registra nivel de precipitación alguno. Esto viene señalado con el nivel -1 (es un valor que ninguna estación nos da) que es una convención que hemos adoptado para que sea más sencillo identificarlos en nuestros programas y experimentos. Podemos verificar en la Figura 12 cómo estos GAPs se producen con bastante frecuencia y duración variable hasta el punto de que algunos vacíos de datos duran mucho tiempo llegando a ser de más de 24 horas. Por ejemplo, el día 8 de marzo de 2016 se producen varios intervalos de tiempo en los que no se recogen datos posiblemente varios días.

Aprovechamos lo expuesto en el párrafo anterior para resaltar algo que es un valor que añade este Proyecto. Nos referimos al hecho de que un sistema basado en radioenlaces complementará y enriquecerá la información meteorológica aportada por medios convencionales como es el caso de las estaciones meteorológicas.

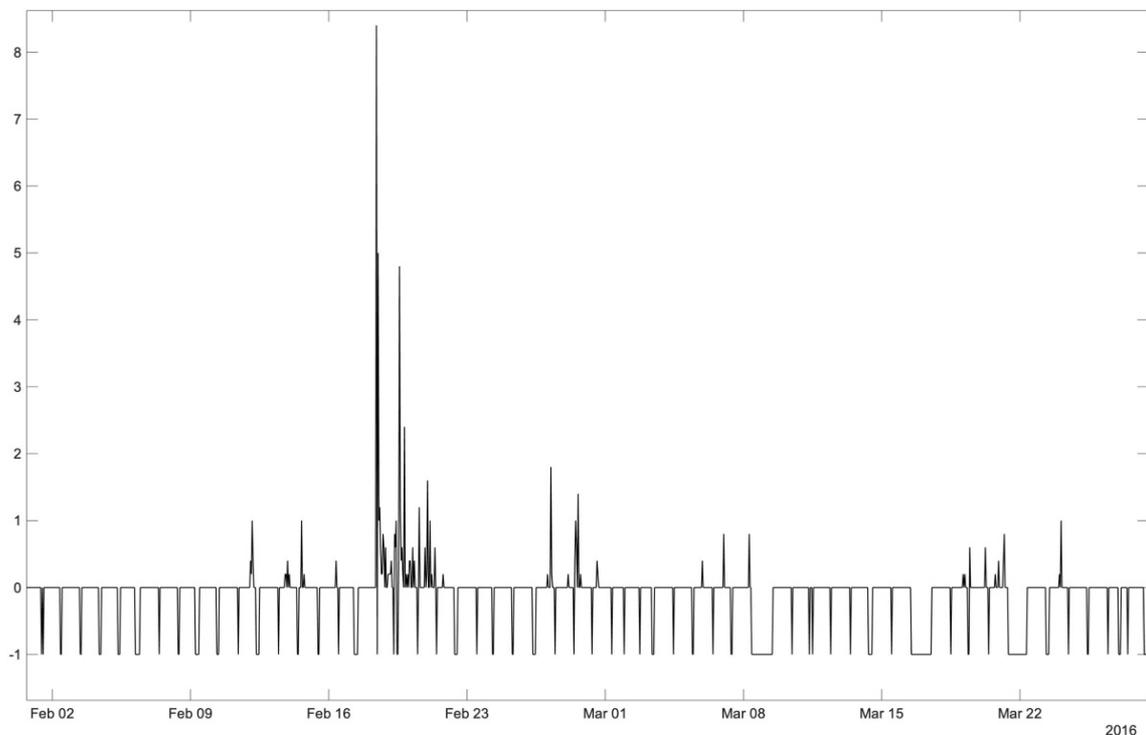


Figura 12 Datos de niveles de precipitación en los que se han sucedido GAPs

## Tipos de GAPS

Se presentan dos tipos de GAPS, estos son:

- Ausencia de datos. La base temporal es correcta pero no tenemos un dato real de precipitación. Para controlar esta situación introducimos un valor aberrante, con el fin de saber que no es un dato real y conseguir una continuidad temporal.
- Error de fecha. Se produce cuando faltan datos de periodos temporales completos, ya sea de una hora, un día o incluso semanas. Como solución localizamos la fecha de inicio de la interrupción temporal y la fecha de finalización de esta guardando estos datos en un vector. También guardamos la fecha de inicio de nuestros datos y su fecha de final, para poder así al trabajar con los datos de los radioenlaces descartar las fechas contenidas en ese vector.

## Causas de GAPS

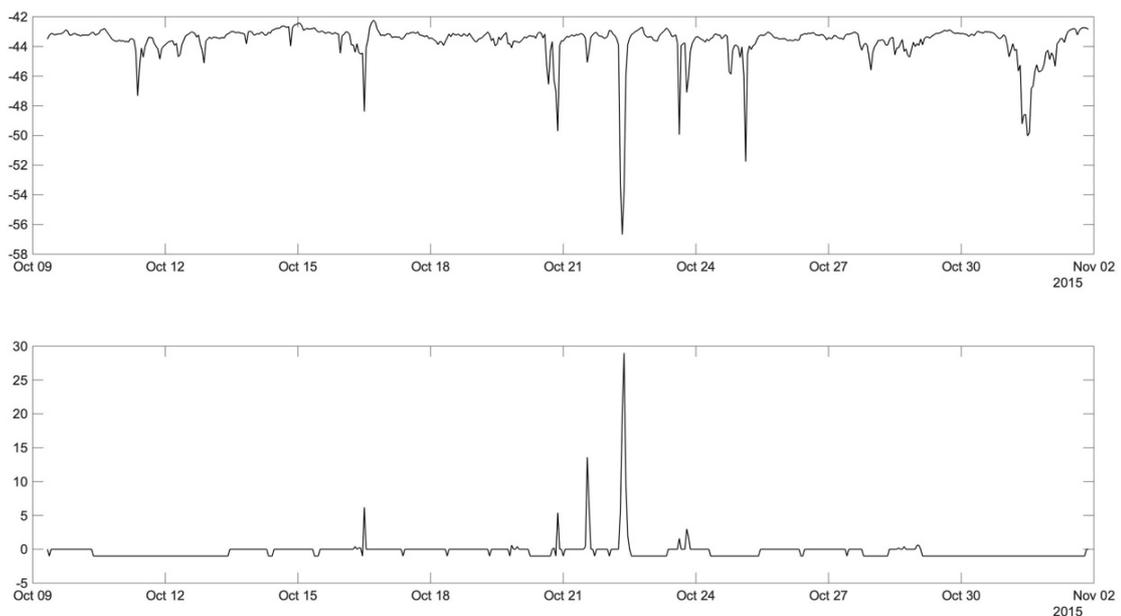
Los tipos de gap mencionados en el apartado anterior se pueden producir por diferentes factores, estos son:

- Vida útil de la batería. Todas las estaciones cuentan con una batería para que se puedan capturar y transmitir los datos. El problema reside en la antigüedad de ésta, puesto que los ciclos de carga y descarga serán menores cuanto más antigua sea y se producirán periodos temporales en los que la estación no transmita datos.
- Placas solares muy pequeñas. Cuanto más pequeña es una placa solar, menos luminosidad capta y antes se ensucian. Muchas veces solo se limpian cuando llueve.
- Lluvias torrenciales. La ganancia de la antena con la que se transmiten los datos es muy baja y si llueve demasiado dicha señal no llega al receptor.
- Interferencias. Las estaciones transmiten en banda libre a frecuencia de MHz por lo que se pueden producir interferencias, y que la señal transmitida no llegue al receptor.

En nuestra base de datos el tipo que se da es de ausencia de datos y las causas son diversas, por ejemplo, por fallo de la estación, fallos en la comunicación de datos, interrupción de la alimentación eléctrica, etc.

### 3.5 Sincronización de los datos

Tal como vimos en los apartados 3.3 y 3.4, existe la posibilidad de que haya correlación entre las atenuaciones de potencia y los eventos de lluvia. Dicho de otra forma, nos planteamos si hay una relación causa (lluvia) y efecto (atenuación). Para confirmar esta posibilidad, analizaremos simultáneamente ambas señales. Podemos ver y observar en la Figura 13 que hay una correspondencia evidente entre el hecho de que haya llovido y su manifestación en forma de atenuación de potencia recibida.



*Figura 13 Señal de potencia y nivel acumulado de lluvia en el mismo intervalo temporal*

Además, se ha comprobado por diversas fuentes que en los días previos y posteriores al día 22 de octubre cayeron grandes lluvias sobre la isla de Gran Canaria. En concreto en ese día, se suspendieron las clases y las actividades extraescolares, por si fuera poco, también hubo inundaciones y otros daños. En la Figura 14 se adjunta la portada del periódico Canarias7 el jueves 22 de octubre de 2015 en la que se muestra la emergencia por lluvias que sufría la isla en ese periodo de tiempo.



Figura 14 Portada del periódico Cararias7 del día 22 de octubre de 2015. Fuente: Canarias7

Hay que destacar que a veces hay vacíos (GAPs) de datos bien en el radioenlace, bien en la estación meteorológica. Esto nos hace tomar la decisión de excluir de nuestros experimentos posteriores los intervalos donde falta alguna de las fuentes de datos. Estos GAPs son claramente apreciables en la Figura 12 donde podemos ver en los niveles de la señal meteorológica que se producen espacios de tiempo donde no hay registro de señal como, por ejemplo, en el periodo del 10 al 13 de octubre. Por otro lado, en la misma figura se puede ver cómo a veces se producen atenuaciones en la señal de potencia, pero no se registra lluvia o no hay datos en el registro meteorológico. Un caso de estos sucesos podría ser el que se da el 30 de octubre donde se produce una atenuación de 6 dBm en la señal de potencia y en los datos meteorológicos no se produce ningún registro tanto de lluvia como de no lluvia, es decir, no hay datos. Esto es un problema porque nunca podremos saber si realmente esa atenuación se produjo por lluvia. Por otro lado, a veces ocurre que hay atenuaciones en nuestra señal, pero se registran niveles nulos de lluvia porque en el punto donde está la estación meteorológica no cayó agua en la hora anterior. Esto nos crea una pregunta, ¿esa atenuación fue producida por lluvia o por otros fenómenos (ruido)? En los experimentos demostraremos que nuestro sistema de detección es capaz de detectar lluvia con una fiabilidad tan alta que podemos inferir que la ha habido, aunque la estación meteorológica registre cero niveles de precipitación.

### 3.5.1 Eliminación de GAPS

Un paso muy importante es identificar todos los GAPS, tanto de los datos de radioenlace como de los meteorológicos. Por ello, hemos sacado de nuestros experimentos todos los GAPS de ambos datos, intervalos de señal de potencia en los casos en los que no teníamos datos meteorológicos e intervalos de datos de meteorología cuando no poseíamos señal de potencia. De esta forma, una ausencia de información motiva que prescindamos de ambos datos en los instantes de muestreo afectados. Conseguimos así tener el mismo número de muestras de la señal de potencia como de los datos meteorológicos. De esta forma estamos en disposición de aplicar técnicas de aprendizaje máquina (Machine Learning) con entrenamiento supervisado, tal como veremos en los capítulos 4 y 5.



## Capítulo 4 - Marco metodológico

En este capítulo realizamos una breve introducción al aprendizaje máquina y sus diferentes formas. También, explicamos brevemente los distintos métodos de clasificación que utilizaremos en este trabajo de fin de grado.

### 4.1 Introducción

Cuando en nuestro día a día mencionamos el dicho de que una imagen vale más que mil palabras nos referimos a que nunca seremos capaces de describir con palabras las características de algo con tal exactitud y precisión como lo puede hacer una imagen. Esta incapacidad de definir ciertos objetos o conceptos con la precisión adecuada es lo que nos hace proponernos transmitir a los ordenadores ciertos conocimientos mediante ejemplos. Sin embargo, para que funcione, el ordenador debe poder convertir los ejemplos en conocimiento. De ahí el interés en los algoritmos y las técnicas de aprendizaje automático.

En el caso particular de este Proyecto, nos encontramos con que una mera aplicación de conceptos como los vistos, por ejemplo, en el apartado 2.3.6 no es suficiente para detectar y cuantificar lluvia. Esto se debe a una serie de circunstancias como las ya expuestas en los Capítulos 2 y 3. Particularmente, una de las situaciones más claras y que complican más el uso de ecuaciones como la 2.2 es la variabilidad de las atenuaciones en función del nivel de lluvia y, sobre todo, el hecho de que no se disponga de estaciones meteorológicas justo en el vano donde está el radioenlace. Esto abre la vía del uso de técnicas de Machine Learning que nos permitan hacer modelos, ver relaciones y estimar lluvia. De ahí que hagamos de estas técnicas un eje fundamental de nuestro TFG. A continuación, daremos una visión general de las técnicas de aprendizaje máquina y de su aplicación específica en el marco en que nos movemos.

Podemos comenzar por el hecho de que es necesario un conjunto de ejemplos de entrenamiento preclasificados. En nuestro caso tendremos muestras de lluvias y de no lluvia. Estos ejemplos constituyen un conjunto de entrenamiento a partir del cual la máquina debe crear un clasificador, es decir, un algoritmo capaz de clasificar cualquier muestra futura en una de las dos clases: positivo (lluvia) y negativo (no lluvia).

Para poder comunicar los ejemplos de entrenamiento a la máquina, tenemos que describirlos de manera adecuada. El mecanismo más común se basa en los llamados atributos también conocidos como vectores de atributos o de características.

El conjunto de datos de entrenamiento constituye la entrada a partir de la cual se creará un modelo. Pero ¿qué modelo? Supongamos que lo queremos en forma de una función booleana que sea verdadera para los ejemplos positivos y falsa para los negativos. Sin pérdida de generalidad, y centrándonos en el objeto de este TFG, pongamos que debemos verificar la expresión  $[(\text{nivel de lluvia} = 0)]$  contra el conjunto de entrenamiento. Podremos ver que su valor es verdadero para todos los ejemplos negativos (no lluvia). En cuanto a los ejemplos positivos (lluvia) la expresión será falsa.

Ante el desconocimiento de cómo inventar un algoritmo de creación de modelos, podemos tratar de obtener algo de inspiración en un intento de crear un modelo a partir de un clasificador manualmente. De esta forma, el clasificador se podría crear mediante una secuencia de pasos de especificación y generalización que modifican gradualmente una versión dada del clasificador hasta cumplir ciertos requisitos predefinidos.

## 4.2 Generalidades sobre el aprendizaje máquina

El aprendizaje máquina trata de encontrar, a partir de un estado inicial, una secuencia de pasos que, después de pasar a través de un conjunto de estados de búsqueda provisionales, nos conducen a un estado final predefinido.

Esta búsqueda se realiza mediante pasos individuales que son llevados a cabo por operadores de búsqueda que han sido especificados previamente por el programador.

Nuestro objetivo no es reclasificar objetos cuyas clases ya conocemos, lo que realmente queremos es etiquetar ejemplos futuros, aquellos cuyas clases aún ignoramos. Por ello, no podemos medir la tasa de error comparando las clases conocidas de los ejemplos de entrenamiento con las recomendadas por el clasificador. El funcionamiento del clasificador se predice experimentalmente.

En un caso sencillo los ejemplos preclasificados disponibles serán trabajados por parejas: un conjunto de entrenamiento a partir del cual se crea el clasificador, y un conjunto de prueba en el que se evalúa el modelo entrenado.

Ejemplos disponibles	
Conjunto de entrenamiento	Conjunto de prueba

Una división diferente de entrenamiento/prueba dará lugar a un clasificador diferente, y también a una predicción diferente del funcionamiento futuro. Esto podría dar lugar a problemas si el número de ejemplos preclasificados es pequeño.

Una forma de evitar problemas es repetir el proceso de división aleatoria en las parejas de entrenamiento y prueba varias veces, siempre produciendo un clasificador del conjunto de entrenamiento  $i$ -ésimo, y luego midiendo la tasa de acierto,  $A_i$ , su correspondiente conjunto de prueba. En este punto, hemos de resaltar que las tasas de acierto son medidas de calidad del clasificador y hay variantes que utilizaremos convenientemente. Por ejemplo, si importante es exactitud, también lo es la sensibilidad a la presencia de lluvia. Entraremos más en detalle a lo largo del capítulo 6 cuando evaluemos los experimentos.

En algunos casos, el conjunto de entrenamiento se crea manualmente: un experto prepara los ejemplos, los etiqueta con etiquetas de clase, elige los atributos y especifica el valor de cada atributo en cada ejemplo. En otros, el proceso está automatizado.

A veces, los vectores de atributos se extraen automáticamente de una base de datos y son etiquetados por un experto. También, se pueden obtener algunos ejemplos de una base de datos, y otros añadirlos manualmente o combinar dos o más bases de datos. El número de variaciones es prácticamente ilimitado, pero sea cual sea la fuente de los ejemplos, es probable que haya errores. Algunos problemas que se suelen presentar son:

- **Atributos irrelevantes.** Algunos atributos son importantes, mientras que otros no lo son.
- **Falta de Atributos.** A veces pueden faltar algunos atributos críticos.

- **Atributos redundantes.** No son especialmente dañinos, son los atributos que son redundantes en el sentido de que sus valores pueden obtenerse de otros atributos.
- **Valores de atributos faltantes.** El valor de algunos atributos no se conoce. Este es el caso que se nos ha dado en este proyecto.

Además, los valores de los atributos y las etiquetas de clase a menudo no son verídicos debido a que las fuentes de información son poco fiables, los dispositivos de medición son deficientes, errores tipográficos y muchas otras razones. Afortunadamente, este no es el caso de este proyecto.

A veces, establecer la clase de cada ejemplo no es suficiente en algunas aplicaciones, también se requiere conocer las razones detrás de dicha clasificación.

#### 4.2.1 Formas de aprendizaje.

Hay dos tipos comunes de aprendizaje que explicaremos a continuación en este apartado. Son conocidos como aprendizaje supervisado y aprendizaje no supervisado.

El aprendizaje supervisado se utiliza para estimar un mapeo desconocido (entrada, salida) a partir de muestras conocidas (entrada, salida). Las tareas de clasificación y regresión se incluyen en este grupo. El término supervisado denota el hecho de que se conocen valores de salida para muestras de entrenamiento (es decir, proporcionados por un maestro o un sistema que se está modelando). Esta es la forma de aprendizaje elegida para este Proyecto pues disponemos de la información que lo permite.

En el aprendizaje no supervisado, solo se dan muestras de entrada a un sistema de aprendizaje, y no existe valores conocidos de la salida durante el aprendizaje. El objetivo del aprendizaje no supervisado puede ser aproximar la distribución de probabilidad de las entradas o descubrir una estructura natural (es decir, grupos) en los datos de entrada.

En los sistemas biológicos, las tareas de reconocimiento y percepción de bajo nivel se aprenden a través del aprendizaje no supervisado, mientras que las capacidades de nivel superior generalmente se adquieren a través del aprendizaje supervisado.

Por ejemplo, los bebés aprenden a reconocer rostros familiares mucho antes de que puedan entender el habla humana. Por el contrario, las habilidades de lectura y escritura no se pueden adquirir de manera no supervisada; necesitan ser enseñados. Esta observación sugiere que los esquemas biológicos de aprendizaje no supervisado se basan en estructuras internas potentes (para una representación y procesamiento óptimos de datos sensoriales) desarrollados a través de los años de evolución, en el proceso de adaptación a la naturaleza estadística del entorno. Por lo tanto, puede ser beneficioso utilizar estructuras inspiradas biológicamente para el aprendizaje no supervisado en sistemas de aprendizaje artificial (Cherkassky & Murier, 2007).

#### **4.2.2 Bases de datos desbalanceadas y estrategias de balanceo.**

El aprendizaje a través de datos desbalanceados es un nuevo reto en muchas aplicaciones de minería de datos, aprendizaje máquina, etc. A veces, se tienen un número pequeño de muestras pero que poseen un gran interés. Por ejemplo, en el ámbito económico, es importante para detectar casos fraudulentos entre muchas transacciones. En otras palabras, el aprendizaje con datos desbalanceados ocurre siempre que una distribución (mayoritaria) de datos domine significativamente a las otras (minoritarias). Un caso claro de este se da en nuestro TFG porque disponemos de muchos instantes de muestreo, pero muy pocos de ellos corresponden a lluvia, que es el fenómeno que buscamos detectar y cuantificar.

Lo descrito en el párrafo anterior presenta varios problemas:

- Las clases minoritarias pueden quedar mal representadas en los experimentos.
- Los datos de calidad de (tasas de acierto, sensibilidad, etc.) se podrían ver excesivamente dominadas por la clase mayoritaria.
- Un análisis de los resultados se hace más difícil que con una base balanceada.

Hay distintos métodos para abordar este problema, los cuales veremos a continuación:

- **Estrategias de muestreo.** Tiene como objetivo desarrollar varias técnicas de sobremuestreo o submuestreo para compensar distribuciones desequilibradas en los conjuntos de datos originales. Hay la posibilidad de hacer un híbrido entre ambas.
- **Generación de datos artificiales.** Tiene como objetivo superar el desequilibrio en los conjuntos de datos originales generando muestras de datos artificialmente, aunque buenas desde el punto de vista de que las réplicas son semejantes a las originales.
- **Matriz de coste.** Utiliza una matriz de costos para diferentes tipos de errores para facilitar el aprendizaje de conjuntos de datos desequilibrados.
- **Aprendizaje activo.** Normalmente se utiliza para problemas con datos de entrenamientos sin etiquetar. Este método selecciona eficazmente un conjunto aleatorio de conjunto de datos de entrenamiento, reduciendo significativamente el costo computacional para grandes conjuntos de datos desequilibrados.

Un algoritmo de aprendizaje que aborda el problema de clasificación de datos desequilibrados es ADASYN. Basándose en la distribución de datos original, este algoritmo puede generar muestras de datos artificiales de forma adaptativa para la clase minoritaria y así reducir el sesgo introducido por la distribución de datos desequilibrados. Además, también puede cambiar el límite de decisión del clasificador para centrarse más en los ejemplos difíciles de aprender mejorando así el aprendizaje. Estos dos objetivos se logran mediante un ajuste dinámico de ponderaciones y un procedimiento de aprendizaje adaptativo acorde a las distribuciones de datos (He, Bai, Garcia, & Li, 2008).

### 4.3 El problema de regresión. Descripción general.

Un modelo de regresión es un patrón que permite describir cómo influye una variable X sobre otra variable Y. Según Cook y Weisberg, el objetivo de la regresión es comprender en la medida de lo posible con los datos disponibles, cómo la distribución condicional de alguna respuesta varía entre las subpoblaciones determinadas por los posibles valores del predictor o predictores (Cook & Weisberg, 1999).

La regresión es un proceso estadístico para estimar las relaciones entre variables. Ayuda a entender cómo el valor de una variable dependiente varía al cambiar el valor de una de las variables independientes, manteniendo el valor de las variables independientes fijas. El interés se centra en la distribución de la variable respuesta  $Y$  condicionada en uno o más predictores  $X$ .

El análisis de regresión estima la esperanza condicional de la variable dependiente dadas las variables independientes. El objetivo es una función de las variables independientes llamada función de regresión. En el análisis de regresión, también es importante la distribución de probabilidad, es decir, la variación de la variable dependiente en torno a la función de regresión.

Muchas veces se equipara el análisis de regresión con algún caso especial del modelo lineal generalizado: regresión normal (lineal), regresión binomial, regresión de Poisson u otras formas que son menos comunes. Sin embargo, a veces hay más de una de esas ecuaciones como en los modelos jerárquicos cuando los coeficientes de regresión en una ecuación se pueden expresar como respuestas dentro de otras ecuaciones, o cuando un conjunto de las ecuaciones está vinculadas a través de sus variables de respuesta (Berk, 2017).

En la práctica el funcionamiento de los diferentes métodos de regresión depende de la forma de generación de datos. Dado que el proceso de generación de datos generalmente no se conoce, el análisis de regresión depende a menudo de hacer suposiciones acerca de este proceso. Estos supuestos son a veces comprobables si disponemos de una cantidad suficiente de datos.

En resumen, este método se utiliza para la predicción donde su uso es importante en el aprendizaje máquina. Se requiere comprender cuáles de las variables independientes están relacionadas con la variable dependiente, y explorar las formas de estas relaciones.

#### 4.3.1 Regresión lineal

Tal como explica Richard A. Berk en el libro *Statistical Learning from a Regression Perspective* (Berk, 2017) la regresión lineal se utiliza para aproximar la relación de dependencia entre una variable dependiente  $Y$ , las variables independientes  $X_i$  y un término aleatorio  $\varepsilon$ .

Un modelo de regresión lineal supone que la función de regresión  $E(Y | X)$  es lineal en las entradas  $X_1, \dots, X_p$ .

El proceso por el cual se calculan los valores de  $Y$  se puede expresar como se muestra en la fórmula (4.1) o de forma compacta en la (4.2):

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_p X_{pi} + \varepsilon_i \quad (4.1)$$

$$Y_i = \beta_0 + \sum \beta_p X_{pi} + \varepsilon_i \quad (4.2)$$

donde:

$Y_i$ : variable dependiente.

$X_1, X_2, \dots, X_p$ : variables independientes.

$\beta_{i (i > 0)}$ : son los parámetros respectivos a cada variable independiente.

$\beta_0$ : es la intersección o término constante.

$p$ : número de parámetros independiente para tener en cuenta en la regresión.

$\varepsilon$ : es el error asociado a la medición  $i$  del valor  $X_{pi}$  y siguen los supuestos de modo que  $\varepsilon_i \sim N(0, \sigma^2)$ .

Es preciso destacar que las rectas de regresión pueden ser las que mejor se ajustan a la nube de puntos generada por una distribución conjunta. La correlación 'r' de las rectas frente a los datos determinará la calidad del ajuste. Si 'r' es cercano o igual a 1, el ajuste será bueno y las predicciones realizadas a partir del modelo obtenido serán muy fiables, esto es, el modelo obtenido resulta verdaderamente representativo. Por el contrario, si r es cercano o igual a 0, se tratará de un ajuste malo en el que las predicciones que se realicen a partir del modelo obtenido no serán fiables.

### 4.3.2 Regresión logística

La regresión logística permite estimar la probabilidad de una variable cualitativa binaria en función de una variable cuantitativa (Rodrigo, 2016). Si una variable cualitativa con dos niveles se codifica como 1 y 0, matemáticamente es

posible ajustar un modelo de regresión lineal por mínimos cuadrados  $\beta_0 + \beta_1 X$ . El problema es que, al tratarse de una recta, para valores extremos del predictor, se obtienen valores de  $Y$  menores que 0 o mayores que 1, lo que entra en contradicción con el hecho de que las probabilidades siempre están dentro del rango  $[0,1]$ . Para evitar esto, como se puede observar en la Figura 15, la regresión logística emplea una función cuyo resultado está siempre comprendido entre 0 y 1.

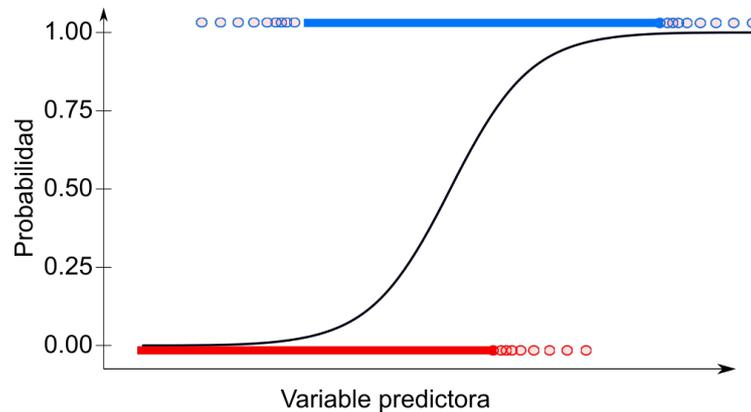


Figura 15 Ejemplo de función de regresión logística. Fuente: elaboración propia

Existen varias funciones que lo cumplen. Una de las más utilizadas es la (4.3), la función logística o sigmoide:

$$\text{Función sigmoide} = \sigma(x) = \frac{1}{1 + e^{-x}} \quad (4.3)$$

Como se puede observar en la ecuación (4.3) para valores de  $x$  muy grandes positivos, el valor de  $e^{-x}$  es aproximadamente 0 por lo que el valor de la función sigmoide es 1. Por el contrario, para valores de  $x$  muy grandes negativos, el valor  $e^{-x}$  tiende a infinito por lo que el valor de la función sigmoide es 0.

Sustituyendo la  $x$  de la ecuación (4.3) por la función lineal  $(\beta_0 + \beta_1 X)$  se obtiene que:

$$P(Y = k|X = x) = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}} \quad (4.4)$$

Donde  $P(Y=k|X=x)$  puede interpretarse como: la probabilidad de que la variable cualitativa  $Y$  adquiera el valor  $k$  (el nivel de referencia, codificado como 1), dado que el predictor  $X$  tiene el valor  $x$ .

La regresión logística múltiple se basa en los mismos principios que la regresión logística simple ampliando el número de predictores que pueden ser tanto continuos como categóricos. En la expresión (4.5) se muestra el valor de la probabilidad  $Y$  para múltiples predictores queda una expresión como la siguiente:

$$p(Y) = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_i x_i}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_i x_i}} \quad (4.5)$$

#### 4.4 Redes neuronales

Es una técnica de aprendizaje supervisado que emula el funcionamiento de una red neuronal biológica, para que a partir de unos datos de entrada obtengamos una determinada salida.

La unidad básica es el perceptrón y mediante la unión de varios de ellos se creará la red neuronal artificial. Cada uno representa una función en base a sus entradas (que representan las dendritas), que nos dará la señal de salida que enviará a otra unidad (que representa la información transmitida por el axón) (Núñez Vivero, 2018).

El perceptrón multicapa es una red neuronal artificial (RNA) formada por múltiples capas, de tal manera que tiene capacidad para resolver problemas que no son linealmente separables, lo cual es la principal limitación del perceptrón.

La función de una neurona, unidad básica de un perceptrón multicapa, es bastante sencilla. Se trata de una suma ponderada de señales que llegan a la entrada que están sujetas a una función de transferencia, también conocida como función de activación que explicaremos más adelante. En la Figura 16 se puede observar un esquema sencillo de una neurona artificial.

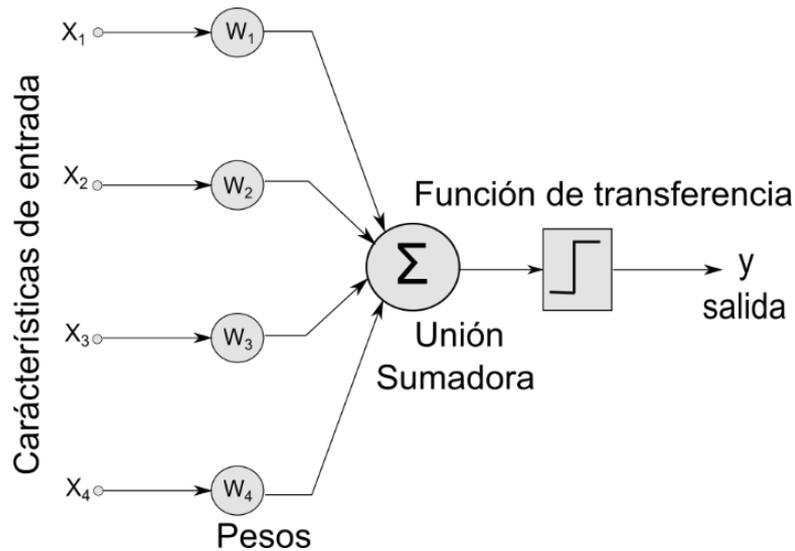


Figura 16 Esquema básico de una neurona artificial con cuatro características de entrada. Fuente: elaboración propia

El proceso de aprendizaje en las redes neuronales artificiales consiste en ajustar los diferentes pesos en todos los perceptrones para conseguir obtener la salida final que queremos a partir de unas determinadas entradas.

La función aplicada sobre el sumatorio (función de activación) se encarga de calcular la salida en función de todas las entradas y pesos, simulando el umbral del potencial de acción en una red biológica, pero en realidad en las redes artificiales no siempre es un umbral, sino que existen otras funciones que se aplican habitualmente a la salida del perceptrón, por lo que el tipo de función de activación que escojamos supone una elección muy importante sobre el funcionamiento de nuestra red (Núñez Vivero, 2018). En el caso de este TFG usamos una función de tipo sigmoide, que es muy habitual.

Cuando introducimos a la red un ejemplo  $x = (x_1, \dots, x_n)$ , los atributos pasan a las neuronas a través de los enlaces. Los valores  $x_k$  se multiplican por los pesos asociados con los enlaces, la  $j$ -ésima neurona oculta recibe como entrada la suma ponderada,  $\sum_k w_{kj}^{(2)} x_k$  y somete esta suma a una sigmoide,  $f\left(\sum_k w_{kj}^{(2)} x_k\right)$ . La neurona de salida  $i$ -ésima recibe la suma ponderada de los valores provienes de las neuronas ocultas y, nuevamente, lo somete a la función de transferencia. Así es cómo se obtiene la salida  $i$ -ésima. Este proceso de propagación de los atributos desde la entrada de la red hasta su salida se denomina propagación hacia adelante (Kubat, 2017)

Como se puede observar en la Figura 17 la arquitectura de la red neuronal consistirá en diversas capas de neuronas. Cuantas más capas y neuronas mayor complejidad y coste computacional. La obtención de la mejor arquitectura se consigue en un proceso de prueba y error hasta dar con la que más se adecúe.

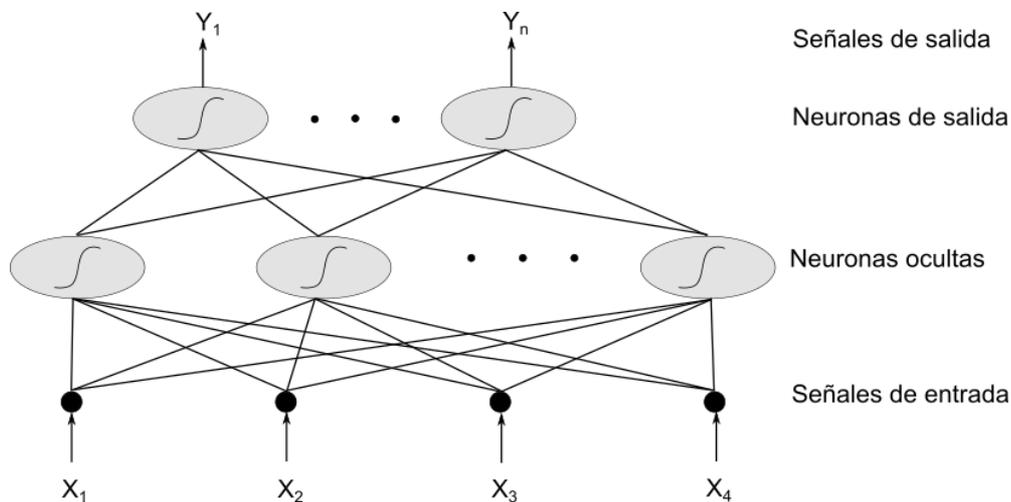


Figura 17 Representación gráfica de la estructura de un perceptrón multicapa de dos niveles.

Fuente: elaboración propia

Las redes neuronales se componen de tres tipos de capas según su posición:

- **Capa de entrada:** Es por la que entran los datos o características iniciales. Habrá tantas neuronas como datos de entrada queramos introducir, esto es,  $\{X_1, \dots, X_p\}$ , donde  $P$  es el número de características (por ejemplo,  $p=4$  en la Figura 11).
- **Capa de salida:** se encarga de devolver los datos finales, tendrá tantas neuronas como salidas queramos, esto es,  $\{Y_1, \dots, Y_n\}$ , donde 'n' es el número de salidas.
- **Capas ocultas:** se encuentran entre la capa de entrada y salida y se encargan de procesar los datos para conseguir los resultados deseados, puede haber muchas capas con tantas neuronas como deseemos en cada una de ellas.

#### 4.4.1 Propagación hacia atrás de errores o backpropagation.

En la fase de entrenamiento, lo complicado es encontrar los valores para los pesos que optimicen el rendimiento de clasificación de la red. Al igual que en el caso de los clasificadores lineales, esto se logra mediante el entrenamiento.

Al principio, los pesos se inicializan con valores de números pequeños aleatorios, generalmente se utiliza el intervalo  $(-0.1, 0.1)$ . Después, se presentan los ejemplos de entrenamiento, uno por uno, y cada uno de ellos se propaga hacia adelante a la salida de la red. La señal de salida se compara con la salida deseada y se calcula una señal de error para cada una de las salidas. Esto nos indica cómo modificar los pesos. Después de la modificación del peso, se presenta el siguiente ejemplo. Cuando se alcanza el último ejemplo de entrenamiento, se completa una iteración o epoch. En perceptrones multicapa, la cantidad de epoch necesarias para un entrenamiento exitoso es mucho mayor que en el caso de los clasificadores lineales: puede ser de miles, decenas de miles, incluso más.

Hay que tener en cuenta que la retropropagación de error es computacionalmente costosa. Tras la presentación de un ejemplo, se debe calcular la responsabilidad de cada neurona individual y luego modificar los pesos en consecuencia. Esto tiene que repetirse para todos los ejemplos de entrenamiento, generalmente para muchas epoch (Kubat, 2017).

#### 4.4.2 Redes neuronales como método de regresión

Los modelos de redes neuronales pueden ser considerados como nuevos paradigmas para el análisis estadístico de regresión lineal. Una de las razones del uso de las redes neuronales es que no necesitan el cumplimiento de supuestos teóricos como en los modelos estadísticos clásicos. El modelo del perceptrón multicapa es equivalente a un modelo de regresión lineal, debido a la similitud de la variable de salida (variable respuesta  $Y$ ) que se relaciona aplicando la función de activación (función identidad) sobre una combinación lineal de pesos (coeficientes) con las variables de entrada (variables predictoras) (Menacho, 2014).

## 4.5 Máquinas de vectores soportados.

El método de clasificación-regresión Máquinas de Vector Soporte (*Vector Support Machines, SVMs*) fue desarrollado en la década de los 90, dentro de campo de la ciencia computacional. Al principio se desarrolló como un método de clasificación binaria, pero su aplicación se ha extendido a problemas de clasificación múltiple y regresión. Las SVM han resultado ser uno de los mejores clasificadores para un amplio abanico de situaciones, por lo que se considera uno de los referentes dentro del ámbito de aprendizaje estadístico y *machine learning (ML)*.

Se fundamentan en el *Maximal Margin Classifier*, que, a su vez, se basa en el concepto de hiperplano. (Amat Rodrigo, 2017). Una máquina de vectores de soporte construye un hiperplano óptimo en forma de superficie de decisión, de modo que el margen de separación entre las dos clases en los datos se amplía al máximo. Los vectores de soporte hacen referencia a un pequeño subconjunto de las observaciones de entrenamiento que se utilizan como soporte para la ubicación óptima de la superficie de decisión (Daicy Cuenca, 2020). Su característica fundamental es la separación óptima. Este tipo de algoritmo busca el hiperplano que tenga la máxima distancia(margen) con los puntos que estén más cerca de él mismo.

### 4.5.1 El clasificador de vectores soportados

Supongamos que hay una variable de respuesta binaria codificada como 1 y -1. También hay una  $f(x)$ , donde  $x$  es un vector de uno o más predictores. La función se puede escribir de una manera lineal familiar como:

$$f(x) = \beta_0 + x^T \beta \quad (4.6)$$

La función 4.6 es principalmente una regresión lineal con un resultado binario de 1 o -1 sin restricciones sobre qué valores numéricos produce la función. Si  $f(x)$  es un número positivo, se asigna un 1 a una observación. Si  $f(x)$  es un número negativo, se asigna un -1 a una observación. Entonces se pueden comparar los 1 y -1 de la función con los 1 y -1 de la variable de respuesta. El problema es cómo

hacer que los dos conjuntos de 1 y -1 correspondan tanto como sea posible, no solo en los datos disponibles, sino con datos nuevos.

El clasificador de vectores soportados busca dos líneas paralelas (límites) que maximizan la distancia perpendicular al límite de decisión. Este es una línea recta y a medio camino entre las dos líneas paralelas. El límite de decisión se representa con una línea negra en la Figura 18.

Las observaciones que quedan por encima de estos límites se denominan vectores de soporte. La distancia entre el límite de decisión y cualquiera de las dos líneas paralelas se llama el margen. Cuanto más ancho es el margen, mayor es la separación entre las dos clases. Entre más grande sea el margen menor serán los errores.

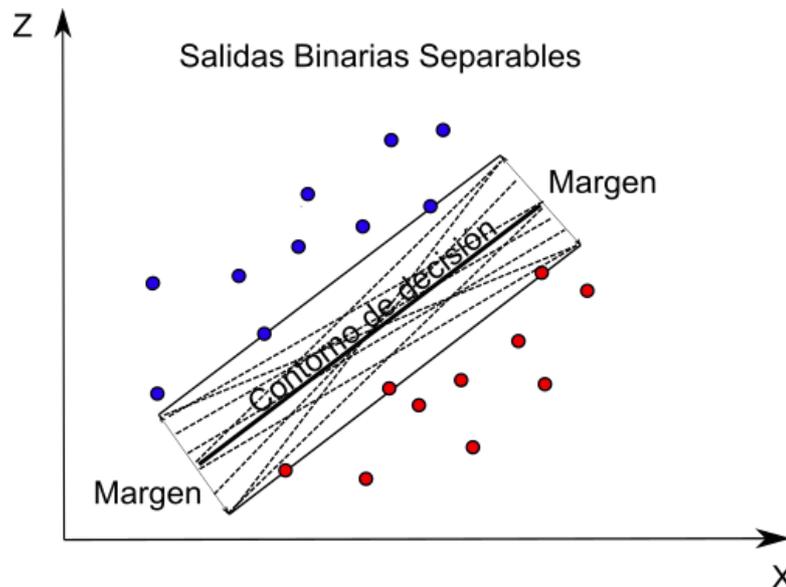


Figura 18 Ejemplo de SVM con dos predictores X y Z y dos clases linealmente separables representadas en rojo y azul. Fuente: elaboración propia

Los casos que se encuentran en un lado del límite de decisión se etiquetan como una clase, y los casos que se encuentran en el otro lado del límite de decisión se etiquetan como la otra clase.

A continuación, a todos los casos nuevos para los que no se conoce la clase final se les asignará la clase determinada por el lado del límite de decisión en el que caen. Esa ubicación será una función de X y Z. La regla de clasificación que se desprende del límite de decisión se denomina umbral duro, y el límite de decisión a menudo se denomina hiperplano de separación. Algunas veces los dos límites reciben el nombre de límite de margen.

Un cierto número de observaciones se situarán en el lado equivocado de su límite de margen. Estas se denominan variables de holgura. Algunas variables de holgura caerán justo a través de su límite de margen, y otras caerán muy lejos. En respuesta, se puede tener en cuenta la distancia entre la cerca relevante y la ubicación de la variable de holgura. La suma de tales distancias puede verse como una medida de cuán permisivo se ha sido.

Si se permite una suma mayor, es posible ubicar un hiperplano de separación dentro de un margen mayor. Una vez más, los márgenes más grandes son buenos. Pueden seguir clasificaciones más estables. Pero las soluciones más permisivas implican más sesgos porque se introducirán clasificaciones erróneas. Reaparece una forma de compensación de sesgo-varianza. Ajustar el clasificador de vectores de soporte con variables de holgura a veces se denomina umbral suave (Berk, 2017).

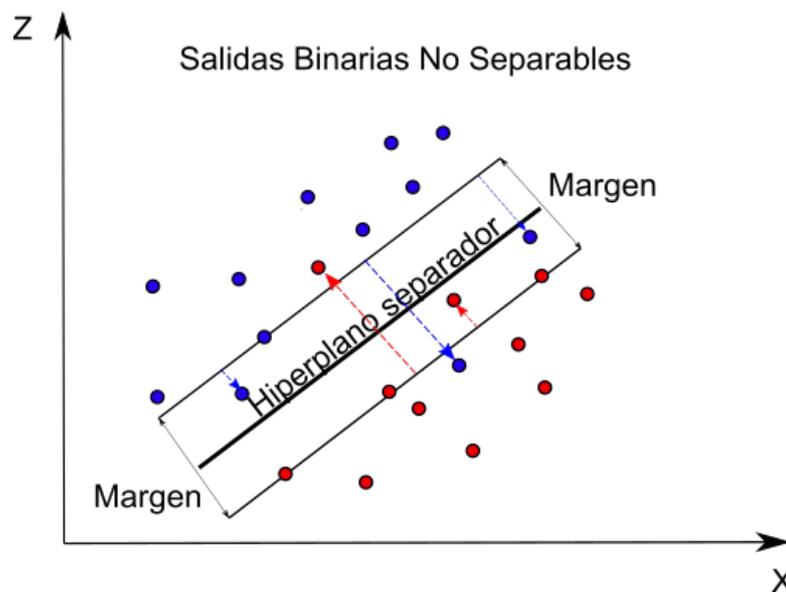


Figura 19 Resultados binarios no separables. Fuente: elaboración propia

#### 4.5.2 El concepto de Núcleo de la máquina

Los clasificadores de vectores soportados se convierten en máquinas de vectores de soporte cuando un núcleo reemplaza un conjunto convencional de predictores. Sin embargo, el uso de núcleos no es sencillo. Puede haber varios candidatos de kernel sin orientación formal sobre cuál elegir. Además, los núcleos vienen con parámetros de ajuste cuyos valores generalmente deben determinarse empíricamente. Los resultados del núcleo dependen de la escala con predictores categóricos, una complicación importante (Berk, 2017).

#### 4.5.3 Máquinas de vectores soportados para la regresión.

Las máquinas de vectores soportados se pueden modificar para aplicarlas a variables de respuesta cuantitativa. Una forma común es despreciar, en el proceso de ajuste, los residuos más pequeños en valor absoluto que una constante. Para los demás residuos, se aplica una función de pérdida lineal (Berk, 2017).



## Capítulo 5 - Algoritmos de detección y cuantificación de precipitación

En este capítulo describimos, paso a paso, cómo desarrollamos el marco metodológico y particularizamos hasta llegar para llegar a los algoritmos que usaremos.

### 5.1 Eliminación baja frecuencia de la señal

Centrándonos en los niveles de potencia, vemos que hay dos tipos de componentes que hemos de filtrar ya que no aportan nada en términos de detección o cuantificación de lluvia, y pudieran dar lugar a errores que, además podemos y debemos evitar. Por un lado, están los saltos bruscos de potencia recibida, que normalmente son debidos a cambios en los controles automáticos de ganancia, ajustes manuales en la potencia de transmisión, etc. Por otro lado, hay variabilidad debida a los cambios que en la atmósfera produce la evolución día-noche. Estos no son muy acusados, pero al no ser debidos a eventos meteorológicos es mejor evitar su efecto.

Para eliminar los saltos de potencia y la ciclicidad día-noche realizamos un filtrado morfológico, particularmente un cierre. En el análisis morfológico un concepto importante es el de elemento estructurante, visto este como el tamaño de la ventana de análisis que usamos cada en cada instante. En nuestro sistema, tomamos como referencia el instante de muestreo en que estamos, y tomamos un elemento estructurante unidimensional definido hacia atrás mediante muestras ya registradas. El cierre es una forma de filtrado que simplifica la función original suavizando los datos originales de forma no lineal, elimina así picos negativos que sean más estrechos que el elemento estructurante. En nuestro caso, hemos realizado pruebas con varios tipos de filtrado como apertura, cierre, top-hat, etc., y hemos comprobado experimentalmente que el cierre es le que se adapta mejor a nuestros propósitos pues permite dos cosas: (1) eliminar picos espurios no asociados a eventos meteorológicos, y (2) servir de referencia para determinar los desvanecimientos asociados a los eventos reales. Así, conseguimos dejar aisladas las componentes de alta frecuencia (variabilidad) que es donde se manifiestan las

atenuaciones y en consecuencia los eventos de lluvia. El tamaño de elemento que hemos tomado corresponde a 2 horas previa a cada instante de análisis.

En la Figura 20 vemos un ejemplo de señal RSL sobre la que hemos aplicado una apertura y un cierre a lo largo de este intervalo (50 horas). En el eje vertical se representan los niveles de potencia dados en dBm y en el eje horizontal las muestras (4 muestras corresponden a una hora). El elemento estructurante es de 8 muestras que abarcan 2 horas.

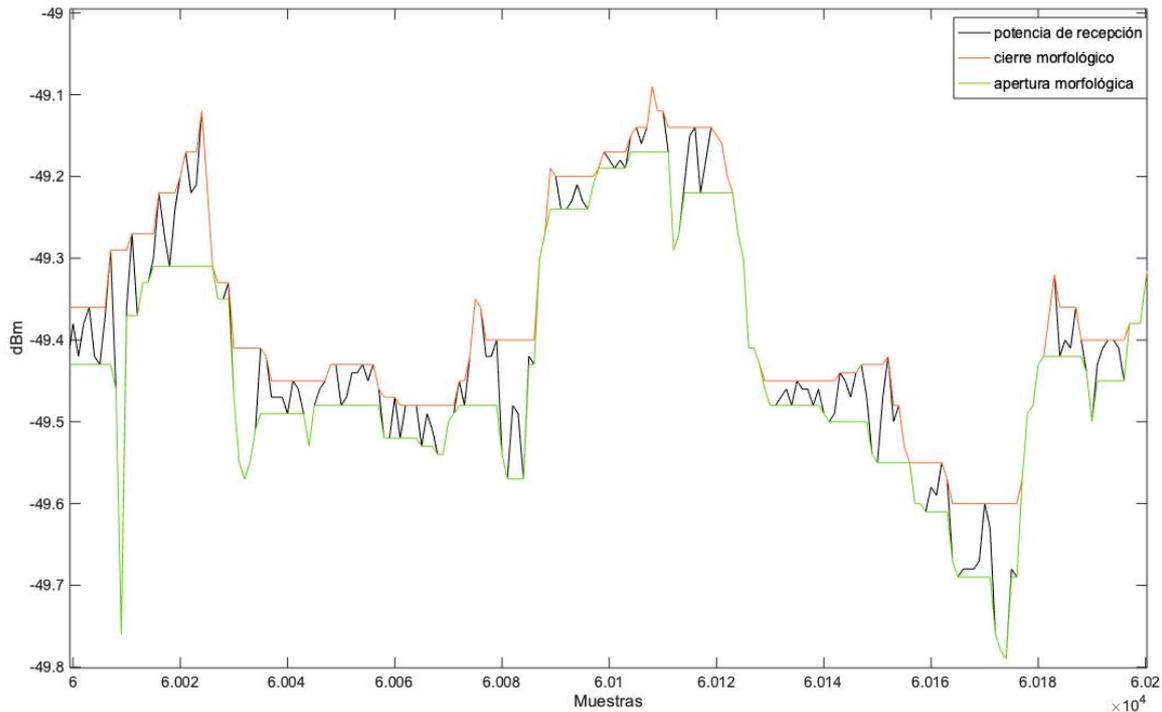


Figura 20 Señal RSL a la que se le ha aplicado una apertura y cierre morfológico

En la Figura 21 se representan 3 gráficos distintos. En el eje vertical se representan los niveles de potencia dados en dBm y en el eje horizontal las muestras (4 muestras corresponden a una hora). En el gráfico superior mostramos la señal del radioenlace Monteluz – Hospital Militar tal cual sacamos los datos de la base de datos. Seguidamente, en el gráfico intermedio se aprecia la baja frecuencia resultante de hacer el filtrado morfológico. Finalmente, en el gráfico inferior encontramos la alta frecuencia, que se obtiene restando la componente de baja frecuencia a la señal original.

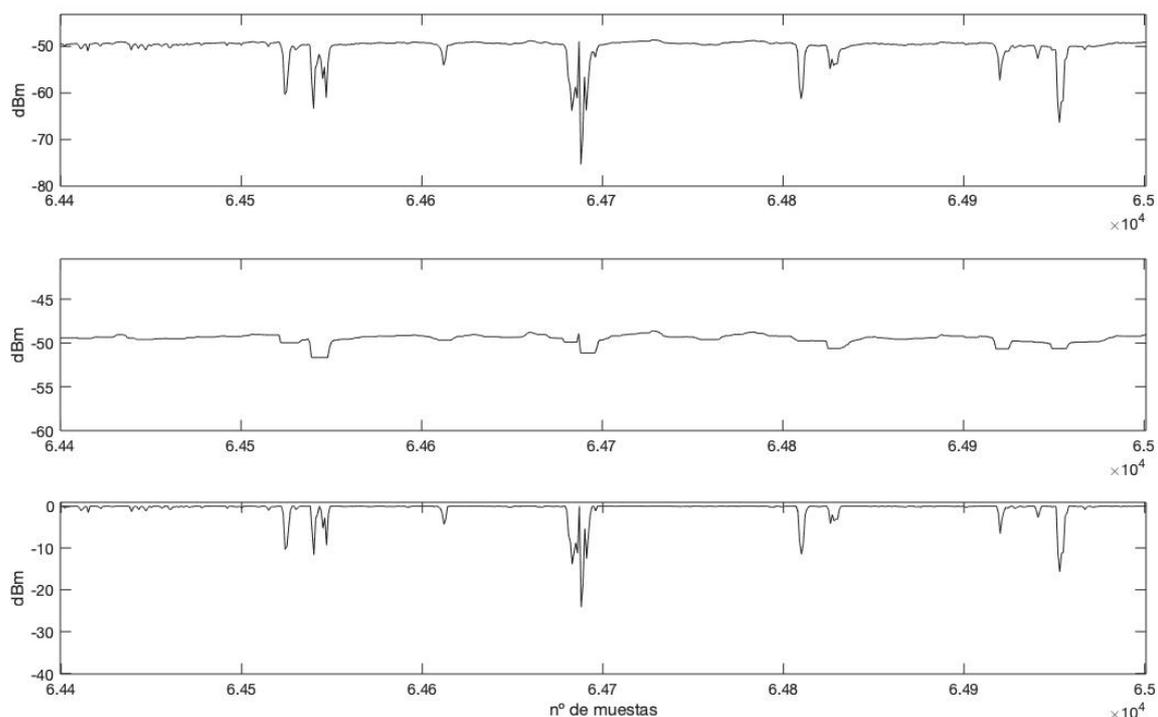


Figura 21 Eliminación de baja frecuencia de la señal mediante cierre morfológico

## 5.2 Extracción de características

Para extraer características realizaremos un filtrado paso bajo mediante la correlación de una señal ideal a través de la señal en estudio. Esta señal ideal es representativa de un día normal sin lluvia, y con ella buscaremos los momentos en que la señal se aparta de la normalidad. Comenzamos calculando una ventana o máscara que se correlará con la señal de estudio (Guerra Moreno, 2017). Para ello se seleccionaron varios días sin precipitación en los datos de los pluviómetros y con ausencias de atenuaciones destacadas en RSL (valores medios de varios días sin precipitación). En la Figura 22 se muestran los resultados obtenidos de la selección donde, por ejemplo, data1 es la máscara original calculada de  $n=23$  muestras (impar). Este valor de  $n$  corresponde con 6 horas de tal forma que con esta máscara estamos correlando con un periodo de esta longitud. Si queremos considerar periodos más largos podemos usar una máscara más larga según la fórmula  $L=(n \cdot (4 \cdot m) - 1)$ , donde vamos en periodos de 6 horas controlando 'm'. Así, por ejemplo, para  $m=4$  la máscara abarca un día, y para  $m=8$  serán dos días.

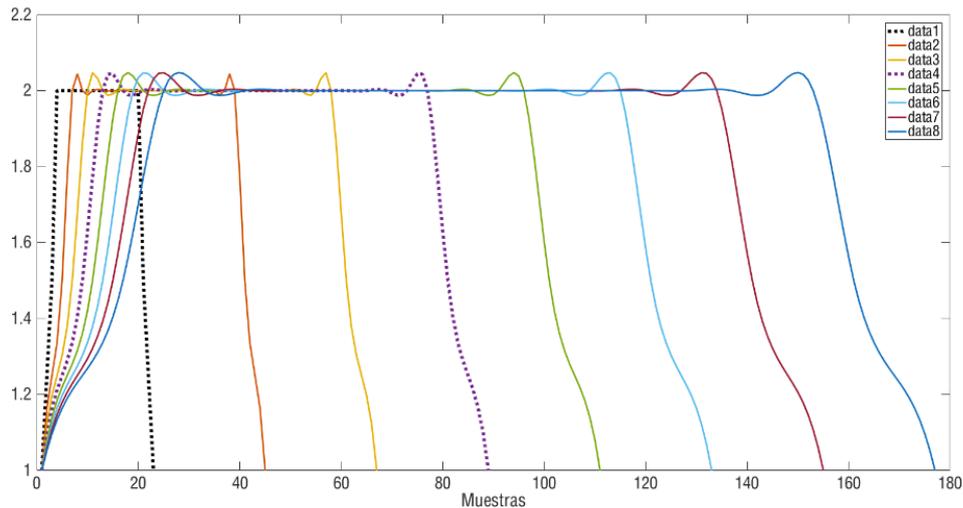


Figura 22 Ventanas de correlación de 8 máscaras Fuente: (Guerra Moreno, 2017)

Al correlar cada una de las máscaras obtenemos 8 señales filtradas en paralelo 15-minutales y, para completar la extracción de características, le añadimos la señal original invertida. Así obtenemos una matriz de señales que describen la forma de onda para cada instante de tiempo 't' y dada una serie de máscaras.

El siguiente paso consiste en realizar un proceso de transformación de muestreo 15-minutal a horario sin pérdida de información. Es imprescindible realizar este proceso así porque el pluviómetro ofrece datos de precipitación en una hora acumulada, es decir, no tiene por qué haber sucedido en el instante ese exactamente ni en toda la hora de integración de mm/h. De hecho, durante una hora determinada puede haber varios eventos de lluvia.

Por ejemplo, si se produce precipitación a las 16:15 de 10 minutos, hasta las 17:00 no obtendríamos un resultado de lluvia acumulada en esa hora. Por ello debemos respetar (en los datos de atenuación) esta información sin destruirla.

En la Figura 23 y 24 se ven representados a la izquierda ejemplos de vectores de características de no lluvia y a la derecha de lluvia. Además, de izquierda a derecha en cada uno de los gráficos las 9 primeras muestras corresponden a características en el primer cuarto de hora, las 9 siguientes al segundo cuarto y así hasta completar la hora. Las características en cada cuarto tienen diferenciación tanto más cuanto más lluvia hay. Por ejemplo, en las gráficas 23 y 24 mostramos vectores de características correspondientes a diferentes intervalos de lluvia (derecha) y no lluvia (izquierda). Podemos apreciar en la

izquierda que hay un margen dinámico muy plano comparado con el de la derecha en el que hay márgenes de amplitud diferentes en cada cuarto de hora. A veces se produce que el margen dinámico en el gráfico de la derecha puede ser confundido con características propias de no lluvia. Esto posiblemente se daba a momentos en los que en la estación meteorológica se estaba recogiendo lluvia (sí llovía), pero en el vano no (no llovía). Al revés podría pasar la situación inversa, esto es, que en la estación no se recogía lluvia (no llovía) mientras que en el radioenlace si se manifiesta lluvia (sí llueve). La causa de estas disparidades está en que las estaciones meteorológicas no están situadas en algún punto del vano, sino que hemos tomado las más cercanas.

En un caso ideal, esperamos que numéricamente a partir de un cierto nivel de lluvia (por ejemplo, 2 mm/h) se aprecien solo vectores de lluvia para que no haya solape. En este caso (ideal) se hablaría de una clasificación óptima. Sin embargo, esto no siempre se va a producir debido al hecho de que en los niveles muy bajos de lluvia los niveles de potencia difícilmente manifiesten la presencia de lluvia y se pueden confundir con ruido (por ejemplo,  $\pm 0.5$  dB de cuantificación de los RSL en valor continuo al pasarlos a discreto).

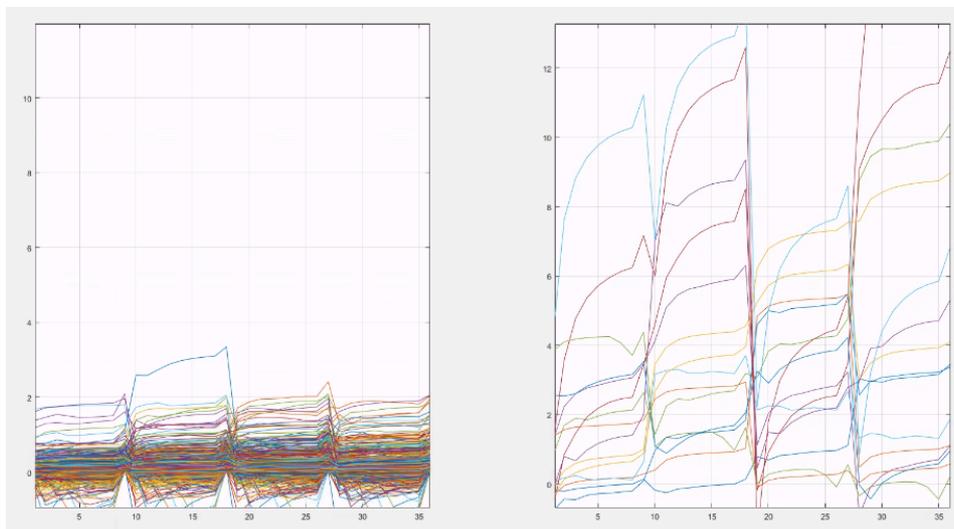


Figura 23 Representación de los vectores de características de no lluvia (izquierda) y lluvia(derecha) con 8 máscaras

Se puede apreciar en la Figura 24 que la forma de los vectores de características de no lluvia son planos mientras que los de lluvia no, es decir, las formas de onda son distintas. Centrémonos en dos casos en concreto, si observamos las componentes resaltadas en verde del gráfico de la derecha

apreciamos que en el primer y segundo cuarto de hora llovió con intensidades diferentes (mayor en el segundo cuarto de hora), en el tercero cayó el nivel de lluvia y en el cuarto volvió a subir. Así mismo, si nos fijamos en la curva violeta vemos como el nivel de intensidad de lluvia va disminuyendo progresivamente.

Con todo este proceso demostramos y verificamos que los vectores de características manifiestan lluvia, que podemos distinguir entre lluvia y no lluvia, y que verdaderamente hay diferencias en ellos en amplitud y en forma. A tal efecto, nuestro trabajo es una demostración de que efectivamente se puede detectar lluvia con el sistema planteado.

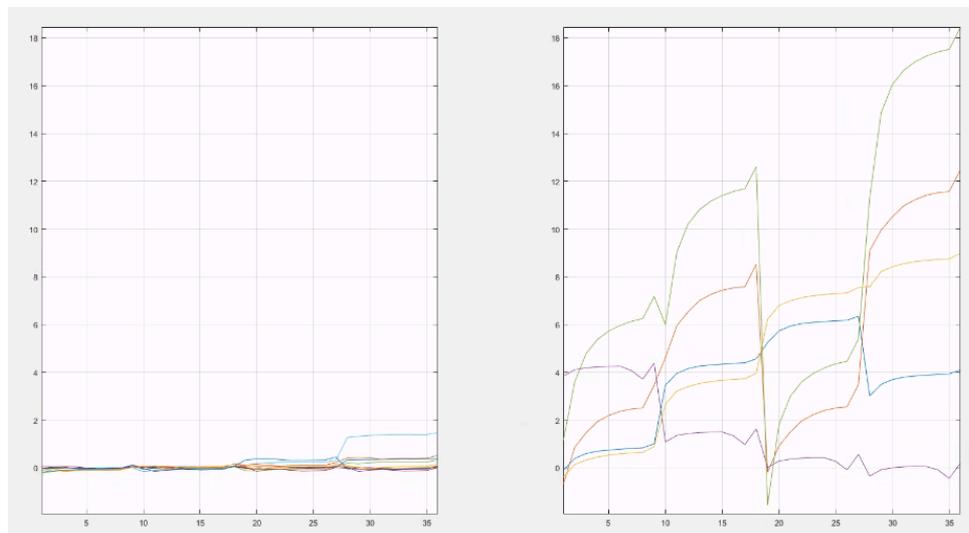


Figura 24 Representación de los 10 primeros vectores de características de no lluvia (izquierda) y lluvia(derecha)

### 5.2.1 Características transformadas.

Buscando aumentar el poder discriminante de los vectores de características hemos hecho una transformación vector a vector en la que obtenemos la media, la varianza, el sesgo (skewness), el apuntamiento (kurtosis) y la transformada de Fourier de los vectores de características.

El sesgo de un estimador es la diferencia entre la esperanza (o valor esperado) del estimador y el verdadero valor del parámetro a estimar. Interesa que un estimador sea centrado, es decir, que su sesgo sea nulo por ser su esperanza igual al parámetro que se desea estimar.

El coeficiente de curtosis (o apuntamiento) analiza el grado de concentración que presentan los valores alrededor de la zona central de la distribución.

### 5.3 Balanceo de la base de datos

En este proyecto, como se muestra en la Tabla 7, tenemos una base de datos desbalanceada en la que los eventos de no-lluvia son marcadamente mayoritarios frente a los eventos de lluvia de diferentes niveles. Es más, considerando conjuntamente todos los eventos de lluvia su número es muy inferior al de no-lluvia. Esto es coherente con el hecho de que en Canarias llueve poco.

Extremo A	Extremo B	Eventos de lluvia	Eventos de no lluvia	Porcentaje de eventos de lluvia sobre el total
Agaete	Montaña_Gáldar	104	2563	3,90%
Agüimes	Tabaiba	70	2597	2,62%
Monteluz	Hospital	150	2517	5,62%
Moriscos	Teror	390	2277	14,62%
Anocheza	Birmagen	97	2586	3,61%
Anocheza	El_Reverón	73	2610	2,72%
Suculum	Bailadero	153	2530	5,70%
Suculum	Casas_Cumbre	133	2550	4,96%

*Tabla 7 Eventos de lluvia y de no lluvia de los radioenlaces de estudio*

La estrategia que hemos seguido es balancear la base de datos submuestreando los eventos de no lluvia, y sobremuestreando los de lluvia mediante el algoritmo ADASYN. Esto se hace generando nuevas muestras de los vectores de características de tal forma que sean semejantes a los que se corresponden con lluvia real. Por tanto, podemos asumir que son características igualmente representativas de la realidad.

### 5.4 Definición de niveles de referencia

Como consecuencia de tener vectores de características vinculados a eventos de lluvia cuyos valores se pueden confundir con los de no lluvia, hemos decidido introducir el concepto de nivel de referencia a partir de la cual vamos comparando por etapas. El objetivo de esto es estudiar a partir de qué nivel de lluvia podemos hacer una detección óptima en términos de los parámetros de calidad del clasificador (por ejemplo, exactitud, sensibilidad, etc.).

Las etapas son como sigue:

1. Primera etapa: Se enfrentan los vectores de no lluvia frente a los de lluvia. Este es el primer nivel de referencia (no lluvia en estado puro).
2. Segunda etapa: Se crea un nuevo nivel de referencia agrupando los vectores de características no lluvia y los del nivel más bajo de lluvia. A su vez, de cara al entrenamiento y posterior test, se re-etiquetan a nivel '0' (no lluvia) todos estos vectores, y los demás mantienen su etiqueta con el nivel real de lluvia. Este nuevo nivel se compara a todos los demás.
3. Etapas sucesivas: Los niveles de referencia se crean acumulando los vectores de no lluvia con los de los siguientes niveles más bajos. Igualmente, se van añadiendo etiquetas a '0' cada vez que un nuevo nivel se añade al de referencia.
4. El proceso se acaba al llegar al penúltimo nivel, o a un nivel prefijado. Experimentalmente hemos comprobado que en los radioenlaces bajo estudio con repetir las etapas hasta un 25% de los niveles de lluvia es suficiente.

## 5.5 Calibrado de la regresión

Una vez que hemos obtenido los resultados de regresión sobre los distintos niveles de lluvia, hemos comprobado que los valores obtenidos tienen una serie de errores sistemáticos. Estos errores son un problema y el hecho de que sean sistemáticos nos lleva a plantear una corrección igualmente sistemática con la que concluiremos el proceso de aprendizaje. Así, siguiendo con los datos de entrenamiento, los cuales se muestran en la Figura 25 (gráfica superior), diseñamos una última etapa de calibrado que aproxime con la mayor precisión posible los valores reales de lluvia. Debe considerarse que esta es la última fase del aprendizaje propiamente dicha.

El proceso de calibrado se realiza en varias etapas sucesivas:

Etapa 1: Una vez generado el modelo correspondiente, aplicamos los vectores de características de entrenamiento sobre dicho modelo, y obtenemos así los valores de lluvia asociados a cada vector.

Etapa 2: Tomamos con referencia el nivel 0 de lluvia (no lluvia) de tal forma que los demás se calibrarán a partir de él. Con esto damos prioridad a la detección propiamente dicha, particularmente, a la discriminación entre no lluvia y lluvia. Para ello, estimamos el nivel de referencia que, tal como se explicó en el apartado 5.4, hará la función de cero. A tal efecto, realizamos una apertura y cierre sobre las estimaciones de no lluvia exclusivamente para obtener sus valores mínimo y máximo. El resultado se puede ver en la Figura 25 (gráfica intermedia). Una vez obtenidas la apertura y el cierre, usamos su valor intermedio como referencia de cero y lo restamos a los valores resultado de aplicar el modelo. Esto se puede ver en la Figura 25 (gráfica inferior) donde podemos comprobar el resultado.

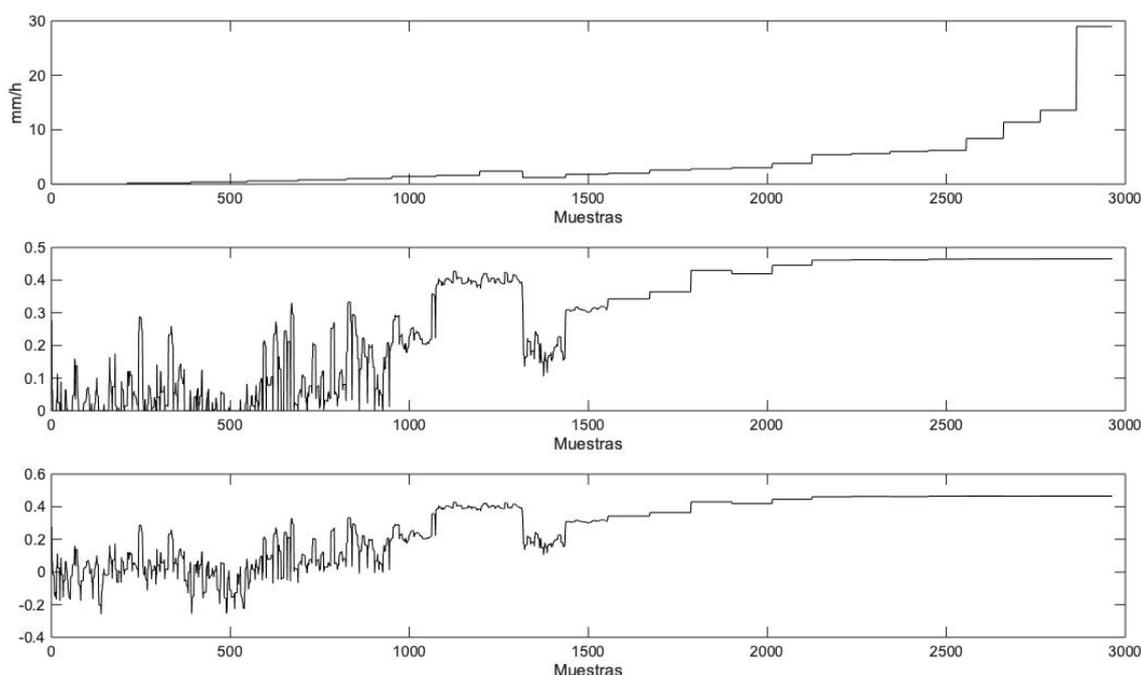


Figura 25 Gráfica superior: datos de entrenamiento; Gráfica intermedia:

Etapa 3: Una vez obtenida la referencia de cero, pasamos a los demás niveles para obtener sus valores de calibrado correspondientes. Nivel a nivel, el proceso de obtención de valor de calibrado es igual. El resultado final se muestra en la figura 26 donde mostramos superpuestos los valores reales y los obtenidos como resultado final del calibrado. Se puede observar cómo en los niveles altos coinciden los valores de referencia (negro) con los obtenidos (rojo). Sin embargo, para niveles bajos hay pequeñas desviaciones, por ejemplo, para no lluvia que corresponde con las primeras 207 muestras hay picos en los cuales se identificaría lluvia (error). Este tipo de error se debe a las posibles causas que hemos ido

comentando durante todo este proyecto (por ejemplo, que los vectores de características en niveles bajos son difícilmente distinguibles, que en la estación meteorológica no estuviese lloviendo, etc.).

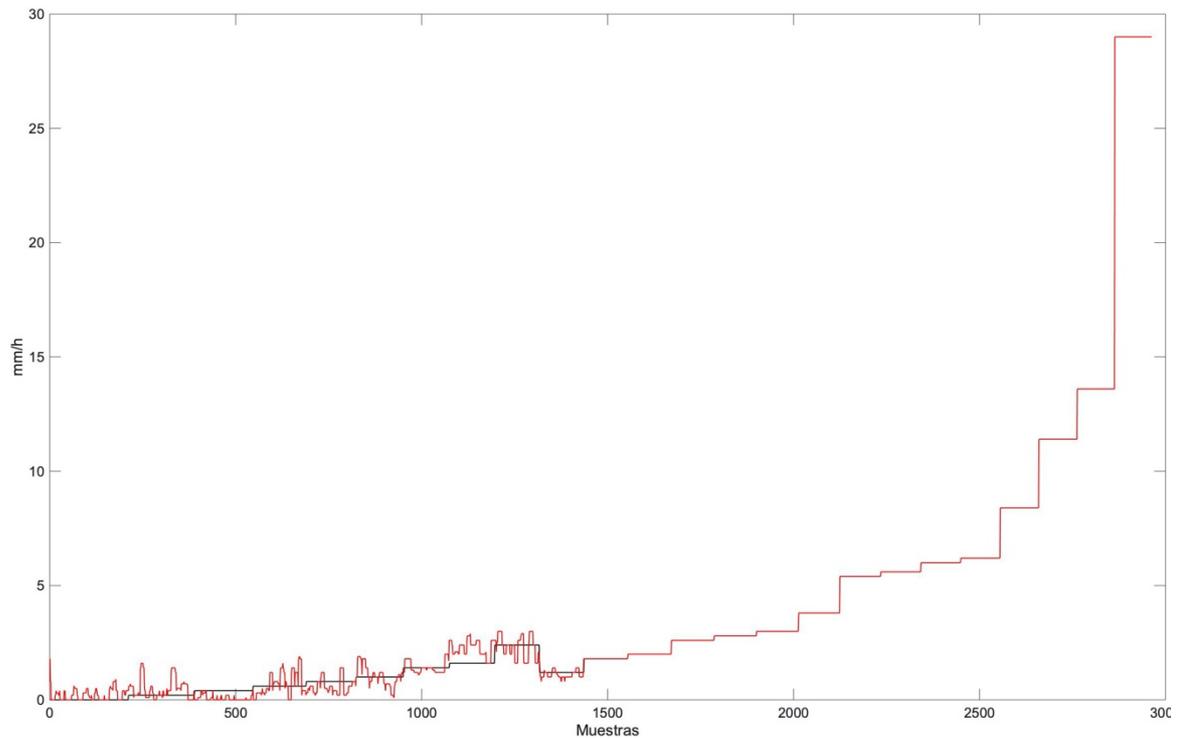


Figura 26 Señal calibrada del radioenlace a Monteluz - Hospital



## Capítulo 6 - Experimentos y Resultados

En este capítulo se presentan los parámetros que hemos utilizado para definir la calidad de los experimentos, se detallan los experimentos realizados y se muestran los resultados obtenidos.

### 6.1 Parámetros de calidad

Para definir los parámetros de calidad tendremos en cuenta dos cosas. Por un lado, la detección y, por otro, la cuantificación. En la detección, hacemos una definición de parámetros propiamente dicha a partir de las detecciones según el tipo de acierto o confusión. En la cuantificación damos el error cometido al cuantificar. En ambos casos, se hace un estudio de la fiabilidad de los modelos obtenidos.

#### 6.1.1 Definición de parámetros de detección

Para poder comprender los parámetros de calidad utilizados es necesario primero entender el significado de los siguientes parámetros previos: TP: True Positive, FN: False Negative, TN: True Negative y FP: False Positive, veamos qué significan:

- True Positive (verdadero positivo): el modelo predice correctamente la clase positiva.
- False Positive (falso positivo): el modelo predice incorrectamente la clase positiva.
- True Negative (verdadero negativo): el modelo predice correctamente la clase negativa.
- False Negative (falso negativo): el modelo predice incorrectamente la clase negativa.

En la Tabla 8 se muestran estos conceptos visualmente:

		Prediction	
		Positive	Negative
Observation	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

Tabla 8 Definición gráfica de True Positive, True Negative, False Positive y False Negative.

Lo anterior es una matriz de confusión y nos permite visualizar el desempeño de un algoritmo de clasificación. Cada columna de la matriz representa el número de predicciones de cada clase, mientras que cada fila representa a las instancias en la clase real. En la Tabla 9 se presenta nuestro modelo casos de lluvia con una matriz de confusión de 2x2 que muestra los cuatro resultados posibles.

		Valores señal Radioenlace	
		Predicción (1)	Predicción (0)
Valores de meteorología	Valor real (1)	Detectamos lluvia y hay precipitación recogida en los datos meteorológicos (TP).	Hay precipitación en los datos meteorológicos, pero no la detectamos en la señal (FN).
	Valor real (0)	Detectamos lluvia, pero no hay precipitación recogida en los datos meteorológicos (FP)	No detectamos lluvia en el radioenlace y tampoco hay en los datos meteorológicos (TN).

Tabla 9 Matriz de confusión del modelo

Para evaluar el rendimiento de la clasificación hemos utilizado los parámetros: sensibilidad, especificidad, precisión y F1-Score. Estudiamos así la calidad desde varios puntos de vista. La sensibilidad es la proporción de casos positivos que fueron correctamente identificados del total de positivos y la especificidad es la proporción de casos negativos que fueron correctamente

identificados del total de negativos, tal y como se especifica en las ecuaciones (6.1) y (6.2).

$$\text{Sensibilidad} = \frac{\text{n}^\circ \text{ eventos detectados}}{\text{n}^\circ \text{ eventos meteorología}} = \frac{TP}{TP + FN} \quad (6.1)$$

$$\text{Especificidad} = \frac{\text{n}^\circ \text{ falsas alarmas}}{\text{n}^\circ \text{ totales de muestras sin evento}} = \frac{TN}{TN + FP} \quad (6.2)$$

La exactitud mide la proporción del total de elementos clasificados correctamente, ecuación (6.3).

$$\text{Exactitud} = \frac{TP + TN}{TP + TN + FP + FN} \quad (6.3)$$

El valor F1 mide la precisión del modelo. Se calcula haciendo la media armónica entre la sensibilidad definida en la ecuación (6.1) y la exhaustividad, estos se definen en las siguientes ecuaciones:

$$\text{Exhaustividad} = \frac{TP}{TP + FN} \quad (6.4)$$

$$F_1 = 2 \cdot \frac{\text{Sensibilidad} \cdot \text{Exhaustividad}}{\text{Sensibilidad} + \text{Exhaustividad}} \quad (6.5)$$

### 6.1.2 Definición de parámetros de cuantificación

Para estudiar la calidad de la cuantificación utilizamos la Raíz del Error Cuadrático Medio (RECM, en inglés RMSD). Ésta es una medida de precisión para comparar errores de predicción. Este estimador mide la cantidad de error que hay entre dos conjuntos de datos, es decir, compara un valor predicho y un valor obtenido o conocido. El RECM da una estimación intuitiva, cuanto más pequeño es un valor, mas cercanos son los valores predichos y observados. El RECM siempre es positivo.

$$RECM = \sqrt{\frac{\sum_{i=1}^N (x_p - x_o)^2}{N}} \quad (6.6)$$

Donde:

$x_p$  = valor predicho

$x_o$  = valor observado

$N$  = número de valores que se usan en el cálculo

### 6.1.3 Estudio de la fiabilidad.

Es importante saber cuanto de reproducible son nuestros experimentos en situaciones diferentes a las de entrenamiento. Para esto hemos elegido hacer una repetición de experimentos de la siguiente forma:

1. Definir un número total de iteraciones, en nuestro caso 30.
2. En cada iteración hemos seleccionado de la base de datos un 70% para entrenar y un 30% para testear los modelos, en el bien entendido de que los ejemplos de test no participaron en el entrenamiento.
3. La selección entre entrenamiento y test se hace de forma aleatoria.
4. Se garantiza que para cada nivel de lluvia hay un 70% para entrenar y un 30% para test.

Una vez obtenidos los valores de detección y cuantificación en cada uno de los 30 experimentos se van acumulando de tal modo que calcularemos la media y la desviación para todos los parámetros de calidad. Estudiamos de esta forma la capacidad de generalización que tienen nuestros modelos. Como estimador de media aritmética usaremos la ecuación 6.7:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (6.7)$$

Asimismo, la desviación estándar se estima a partir de la ecuación 6.8:

$$s = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N - 1}} \quad (6.8)$$

## 6.2. Experimentos con regresión lineal

Se realizaron varios tipos de experimentos en los que se modificó el número de máscaras, las características utilizadas, el número de iteraciones, el tipo de balanceo, etc. Analizando los resultados se llegó a la conclusión que la mejor combinación en regresión lineal es usando las características originales expuestas en el apartado 5.2 y una máscara ya que todos los radioenlaces excepto uno (Suculum – BailaderoA) presentaron mejores valores medios y desviación estándar de sensibilidad, especificidad, precisión y F1-Score para este caso. El radioenlace Suculum – BailaderoA mostró mejores valores usando las características originales y transformadas y 8 máscaras, aunque los resultados no difieren sustancialmente del experimento calificado como óptimo (características originales y una máscara). Veamos cómo hemos llegado a estas conclusiones.

Presentaremos los resultados obtenidos con regresión lineal por medio de una serie de gráficas en las cuales el eje horizontal corresponde con los distintos niveles de lluvia de referencia (ver apartado 5.4) y el eje vertical con los valores medios (gráfica superior) y desviación estándar (gráfica inferior) de sensibilidad, especificidad, precisión y F1-Score dados en porcentaje. Estos parámetros corresponden a los colores azul, verde, negro y amarillo, respectivamente.

Antes de mostrar los resultados hay que explicar que en el eje horizontal los números enteros son los que corresponden con los distintos niveles de lluvia, es decir, como se explicó en el apartado 5.4, el 0 corresponde al nivel no lluvia contra todos los niveles de lluvia, el 1 corresponde a los vectores de características del nivel mas bajo de lluvia más los de no lluvia contra todos los demás y así sucesivamente. Además, en las gráficas solo se muestra el 25% de los niveles de lluvia obtenidos.

La Figura 27 muestra los resultados obtenidos para el radioenlace Agaete – Montaña de Gáldar. Destacamos que los porcentajes son altos ya que los valores medios de los parámetros de calidad rebasan siempre el 80%. Además, podemos

observar cómo todos los parámetros de calidad decrecen conforme cambia el nivel de referencia. Este decrecimiento resulta un tanto sorprendente porque la idea de los niveles de referencia es buscar el efecto contrario. Se debe a que sin usar los niveles de referencia los modelos ya son suficientemente buenos. En cuanto a la desviación los porcentajes son bajos ya que se mantienen siempre por debajo del 3%. Para este radioenlace podemos tomar como nivel de referencia tanto el nivel 0, como el 1 y el 2 ya que todos presentan valores muy buenos, aunque diferentes para cada parámetro de calidad.

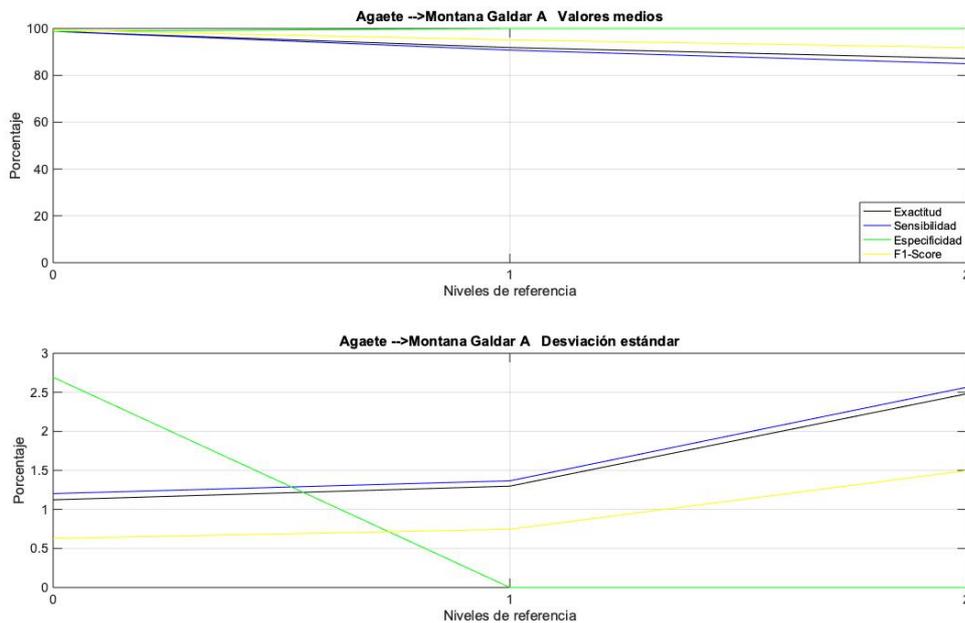


Figura 27 Resultados para el radioenlace Agaete – Montaña de Gáldar

La Figura 28 muestra los resultados obtenidos para el radioenlace Agüimes -Tabaiba. Para este radioenlace tomamos como nivel de referencia el 2. En este nivel es donde se presentan los mejores porcentajes de valores medios y desviación estándar para los diferentes parámetros de calidad llegando a niveles de valores medios que superan el 80% siempre y en muchas ocasiones el 90%. Asimismo, los porcentajes de la desviación estándar a partir del nivel 2 son muy bajos, por debajo del 4%.

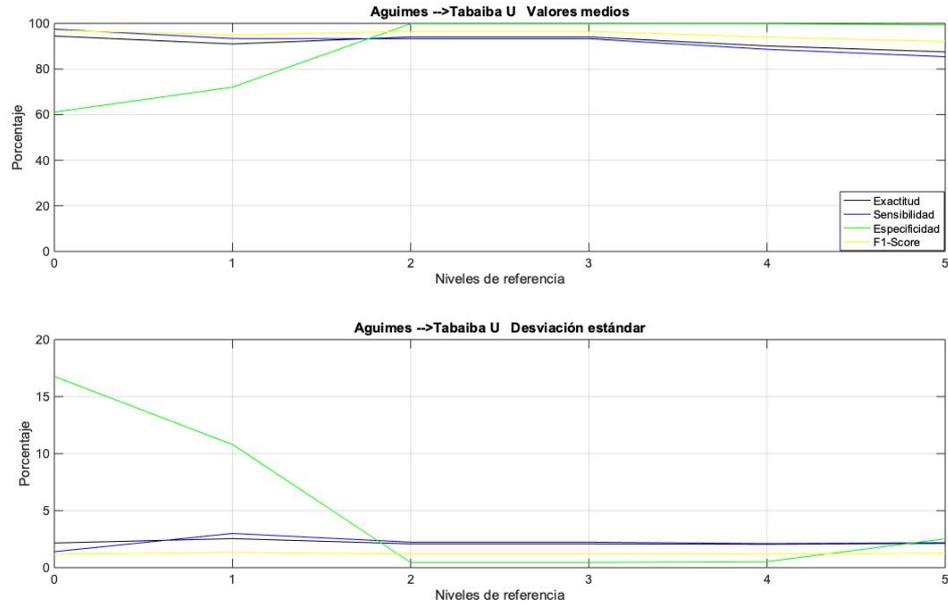


Figura 28 Resultados para el radioenlace Agüimes – Tabaiba

La Figura 29 muestra los resultados obtenidos para el radioenlace Monteluz – Hospital. Presenta unos porcentajes de valores medios muy buenos siendo estos valores del 80% o superior. En cuanto a la desviación estándar, los parámetros de calidad de exactitud, sensibilidad y F1-Score empeoran (aumentan) conforme cambia el nivel de referencia, sin embargo, se mantienen en buenos valores pues en el peor caso estos rondan el 4%. Por otro lado, se observa unos porcentajes de especificidad altos pues en niveles bajos llega a superar el 10%.

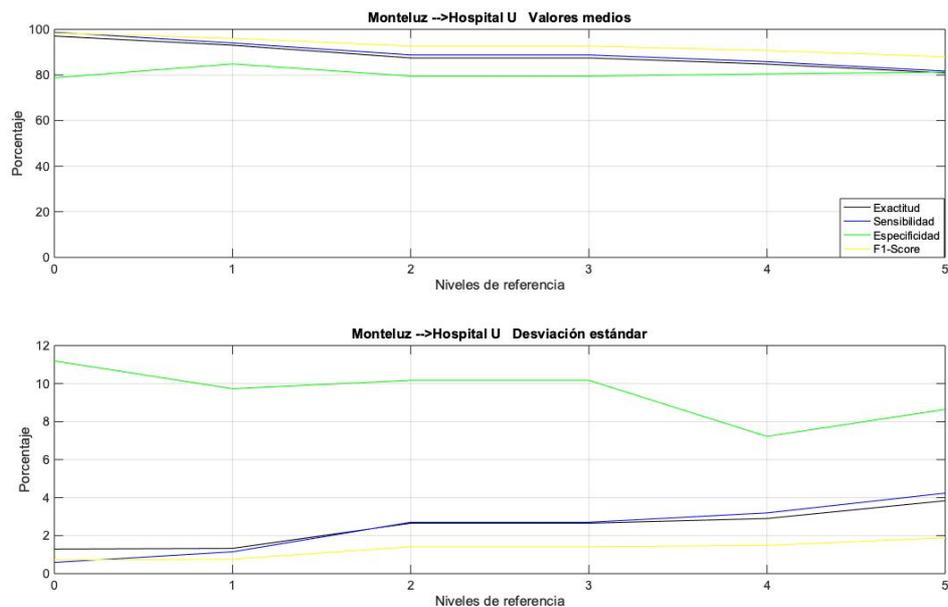


Figura 29 Resultados para el radioenlace Monteluz - Hospital

La Figura 30 muestra los resultados obtenidos para el radioenlace Moriscos – Teror. Destacamos que los porcentajes son altos ya que los valores medios de los parámetros de calidad rebasan siempre el 85%. En cuanto a la desviación estándar los porcentajes son buenos pues, aunque hay variaciones, nunca se rebasa el 4%. Además, concretamente para los parámetros de exactitud, sensibilidad y F1-Score estos valores siempre se mantienen por debajo del 1%.

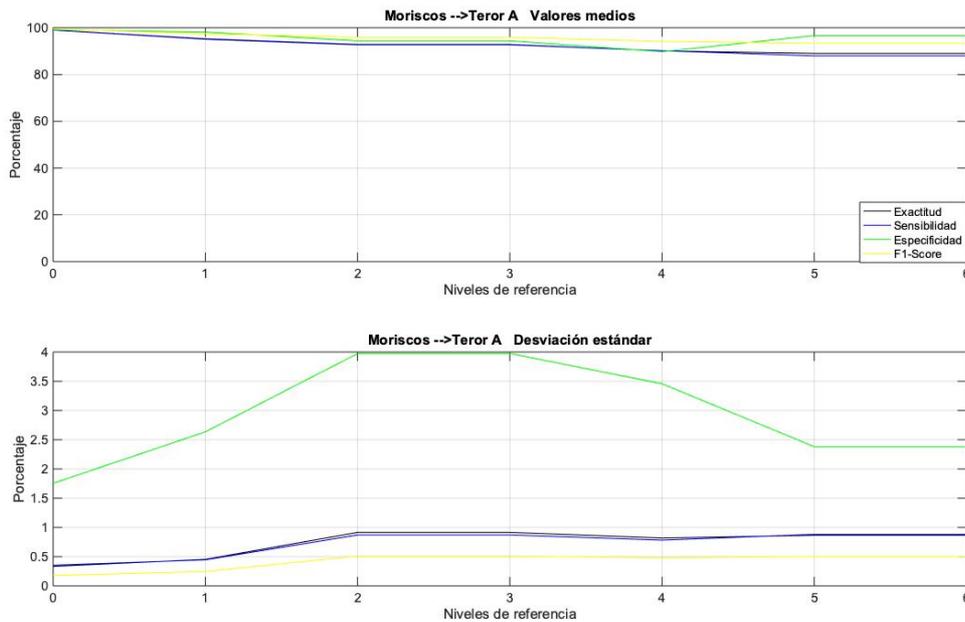


Figura 30 Resultados para el radioenlace Moriscos - Teror

La Figura 31 muestra los resultados obtenidos para el radioenlace Anocheza - Birmagen. Se observa que para todos los parámetros de calidad los porcentajes de valores medios disminuyen conforme cambia el nivel de referencia. Sin embargo, para la sensibilidad, exactitud y F1-Score se mantiene en porcentajes altos superiores al 79%. Asimismo, estos mismos parámetros presentan una desviación estándar muy buena, menor al 4%. En cuanto a la especificidad no se obtienen resultados tan buenos pues llega a valores medios del 60% y porcentajes de desviación estándar superiores al 10%. Por ello en este radioenlace tomamos como nivel de referencia el 2 para nunca superar este 10%.

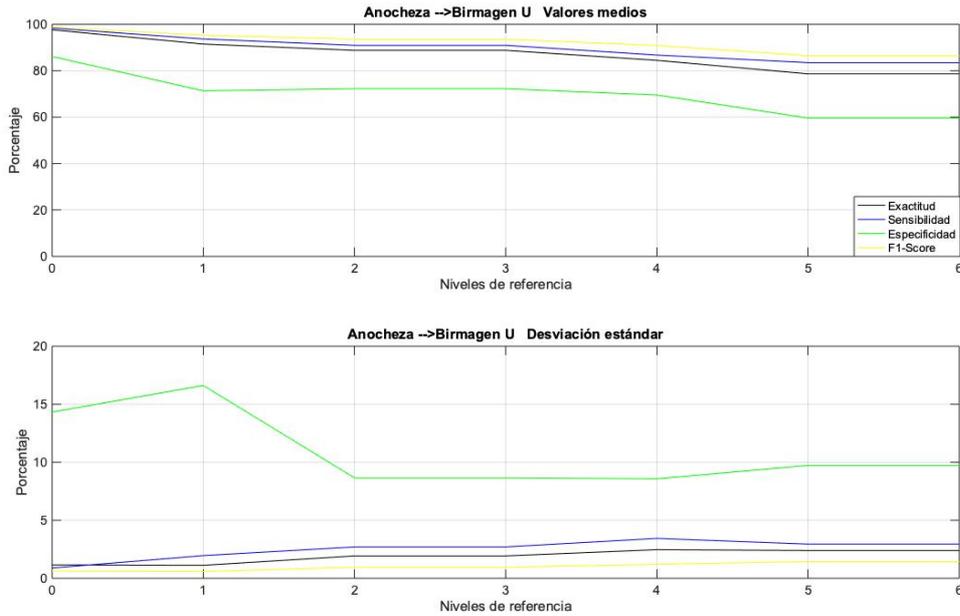


Figura 31 Resultados para el radioenlace Anocheza - Birmagen

La Figura 32 muestra los resultados obtenidos para el radioenlace Anocheza – El Reverón. Para este radioenlace tomamos como nivel de referencia el 1 pues es a partir de este nivel donde se presentan los mejores porcentajes para los diferentes parámetros de calidad. Estos porcentajes son muy buenos rozando el 90% o superiores para los valores medios y el 3% o inferiores para la desviación estándar.

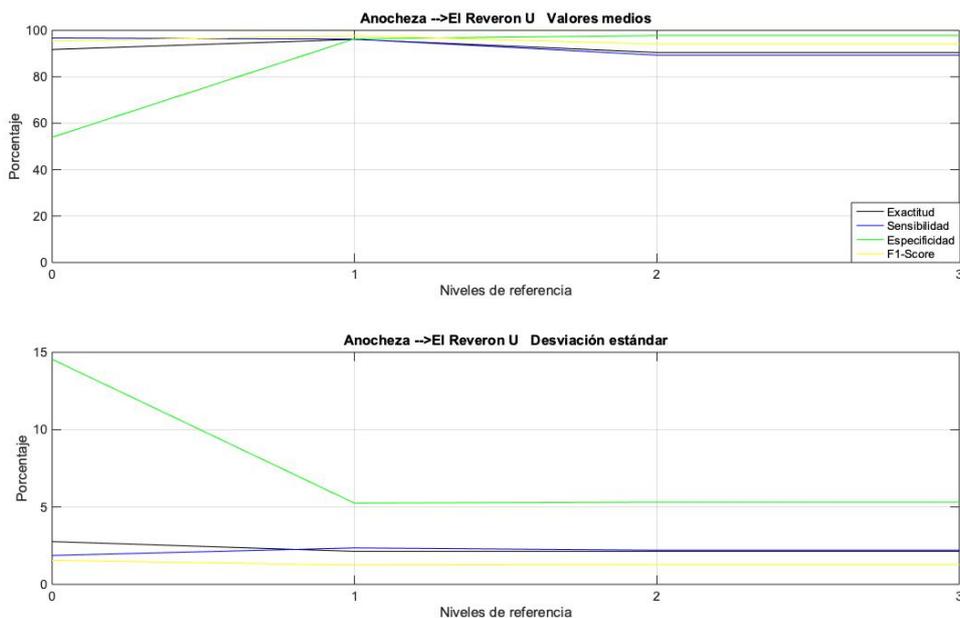


Figura 32 Resultados para el radioenlace Anocheza – El Reverón

La Figura 33 muestra los resultados obtenidos para el radioenlace Suculum - Bailadero. Para este radioenlace tomamos como nivel de referencia el 4 pues es a partir de este nivel donde se presentan los mejores porcentajes para los diferentes parámetros de calidad. Estos porcentajes son muy buenos rozando el 90% o superiores para los valores medios y el 2% o inferiores para la desviación estándar excepto para la sensibilidad donde este valor asciende al 6%.

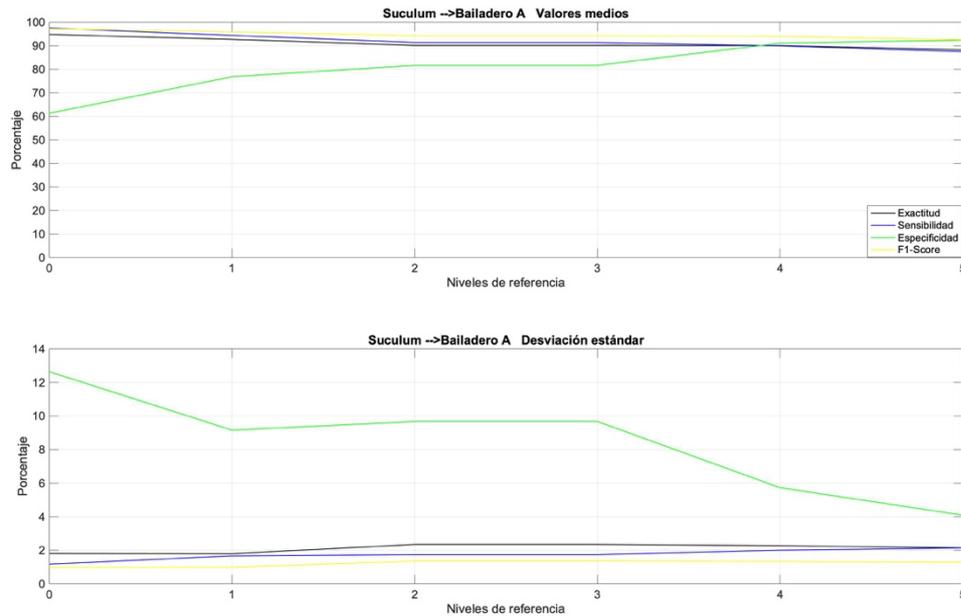


Figura 33 Resultados para el radioenlace Suculum - Bailadero

La Figura 34 muestra los resultados obtenidos para el radioenlace Suculum - Casas Cumbre. Destacamos que se presentan valores muy buenos de valores medios ya que son siempre del 90% o superiores. En cuanto a la desviación estándar, si bien no se obtiene un buen resultado para el nivel 0 (no lluvia en su totalidad) ya que es de un 7,4%, en el siguiente nivel (nivel 1) desciende hasta obtener todos los parámetros de calidad por debajo de un 2%. Por ello en este radioenlace tomaríamos este nivel como referencia.

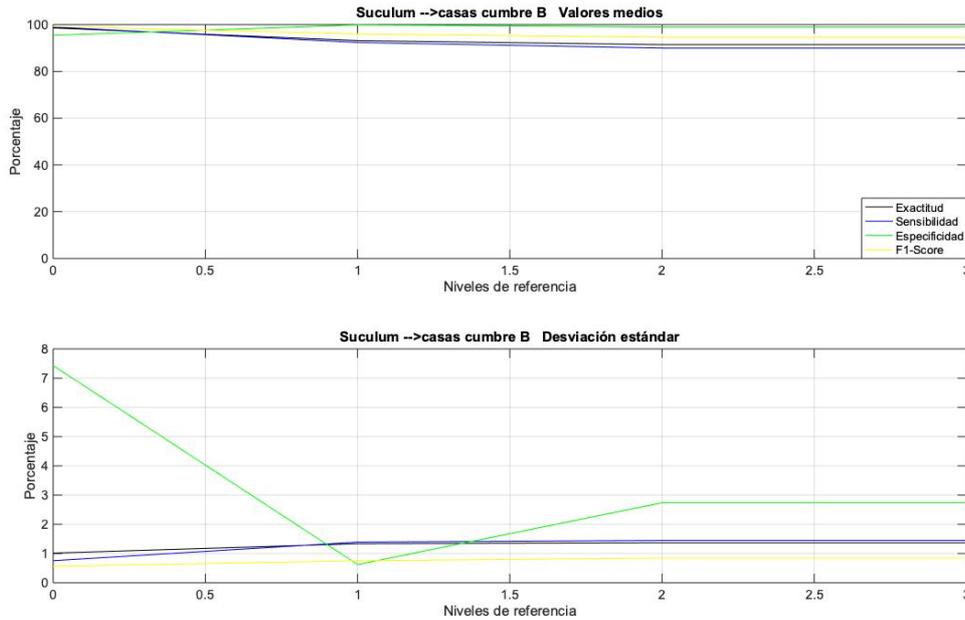


Figura 34 Resultados para el radioenlace Suculum – Casas Cumbre

### 6.3 Experimentos con regresión logística

Después de varios experimentos se llegó a la conclusión que la mejor combinación de experimento para regresión logística es utilizar las características originales y una máscara ya que el 62,5% de los radioenlaces de estudio presentaron mejores valores medios y desviación estándar de sensibilidad, especificidad, precisión y F1-Score para este caso. El 37,5% restante corresponde a 3 radioenlaces, concretamente a los radioenlaces Moriscos – Teror, Anocheza – El Reverón y Suculum – Bailadero.

El radioenlace Suculum – Bailadero A mostró mejores resultados utilizando las características originales y transformadas y 8 máscaras. El radioenlace Moriscos – Teror y Anocheza – El Reverón presentaron mejores resultados para la combinación características transformadas y 8 máscaras. Aunque estos radioenlaces presentan mejores resultados con estas combinaciones de experimentos estos no difieren sustancialmente del experimento calificado como óptimo (características originales y una máscara).

A continuación, se muestran los resultados obtenidos con regresión logarítmica para cada radioenlace por medio de una serie de gráficas en las cuales el eje horizontal corresponde con los distintos niveles de lluvia obtenidos y el eje vertical con los valores medios y desviación estándar de sensibilidad, especificidad,

precisión y F1 dados en porcentaje. Estos parámetros corresponden a los colores azul, verde, negro y amarillo respectivamente.

La Figura 35 muestra los resultados obtenidos para el radioenlace Agaete – Montaña de Gáldar. Obtenemos muy buenos resultados para valores medios pues superamos el 85% en todos los parámetros de calidad. Asimismo, aunque hay variaciones en los porcentajes de desviación estándar estas no superan el 4%. Para este radioenlace podemos tomar como nivel de referencia tanto el nivel 0, como el 1 y el 2 ya que todos presentan valores muy buenos, aunque diferentes para cada parámetro de calidad.

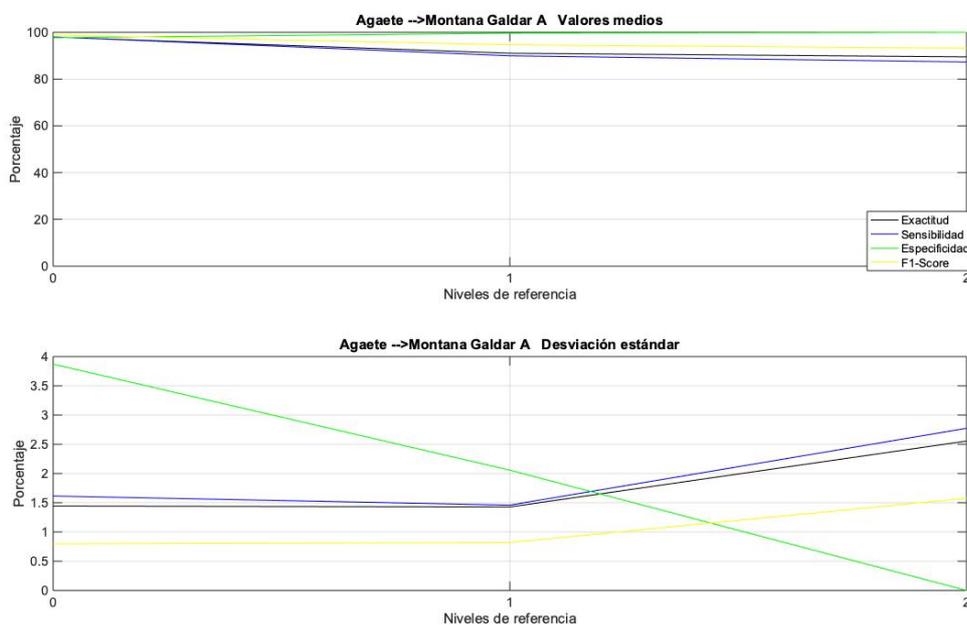


Figura 35 Resultados para el radioenlace Agaete - Montaña de Gáldar

La Figura 36 muestra los resultados obtenidos para el radioenlace Agüimes - Tabaiba. Para este radioenlace tomamos como nivel de referencia el 2. En este nivel es donde se presentan los mejores porcentajes de valores medios y desviación estándar para los diferentes parámetros de calidad llegando a niveles de valores medios que superan el 80% siempre y en muchas ocasiones el 90%. Asimismo, los porcentajes de la desviación estándar a partir del nivel 2 son muy bajos, por debajo del 4%.

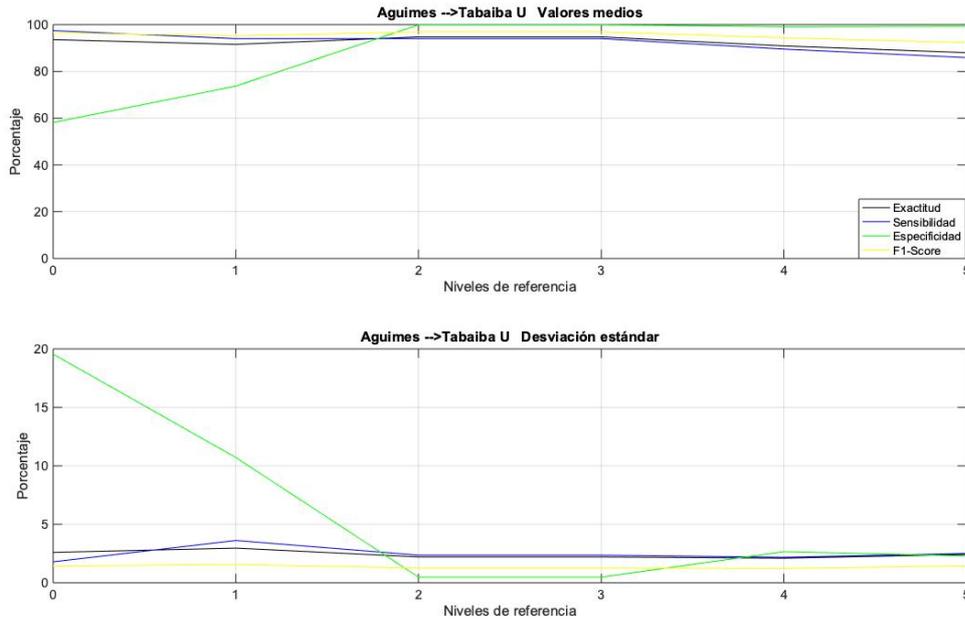


Figura 36 Resultados para el radioenlace Agüimes – Tabaiba

La Figura 37 muestra los resultados obtenidos para el radioenlace Monteluz – Hospital. Se observan porcentajes de valores medios que en los peores casos rondan el 80% para los distintos parámetros de calidad. La desviación estándar de la especificidad es bastante alta pues empieza en un 13,8% en el nivel 0 pero esta disminuye hasta un 6% en el nivel 4 que es el que cogeríamos como referencia. El resto de los parámetros quedan siempre por debajo del 4%.

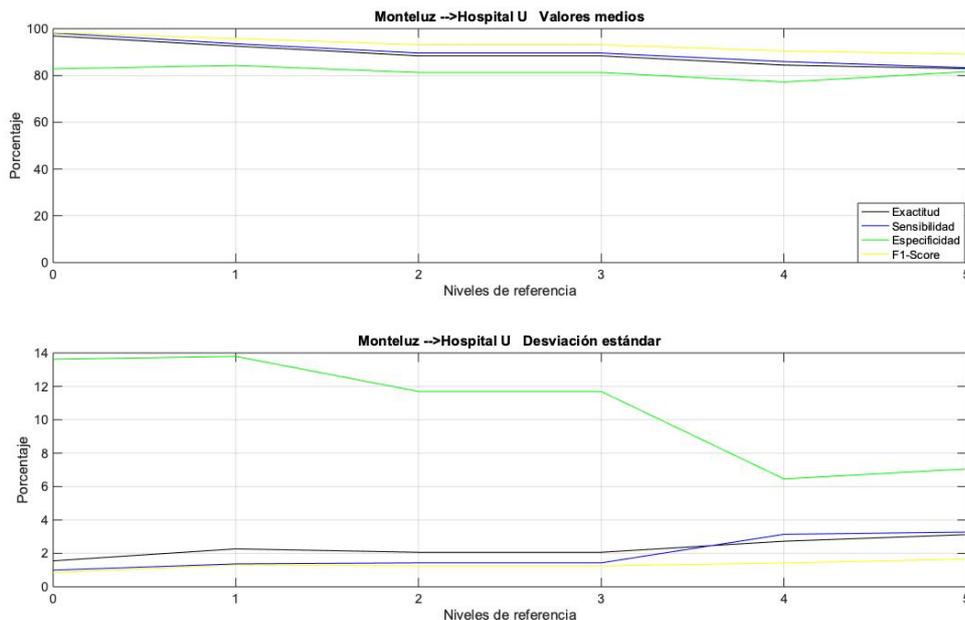


Figura 37 Resultados para el radioenlace Monteluz - Hospital

La Figura 38 muestra los resultados obtenidos para el radioenlace Moriscos – Teror. Destacamos el hecho que con nivel de referencia 0 se dan los mejores resultados, esto es muy bueno ya que es a lo que aspiramos en todos los radioenlaces. Además, los porcentajes son altos para todos los niveles ya que los valores medios de los parámetros de calidad rebasan siempre el 85%. En cuanto a la desviación estándar los porcentajes son buenos pues, aunque hay variaciones, nunca se rebasa el 4%. Además, concretamente para los parámetros de exactitud, sensibilidad y F1-Score estos valores siempre se mantienen por debajo del 1,25%.

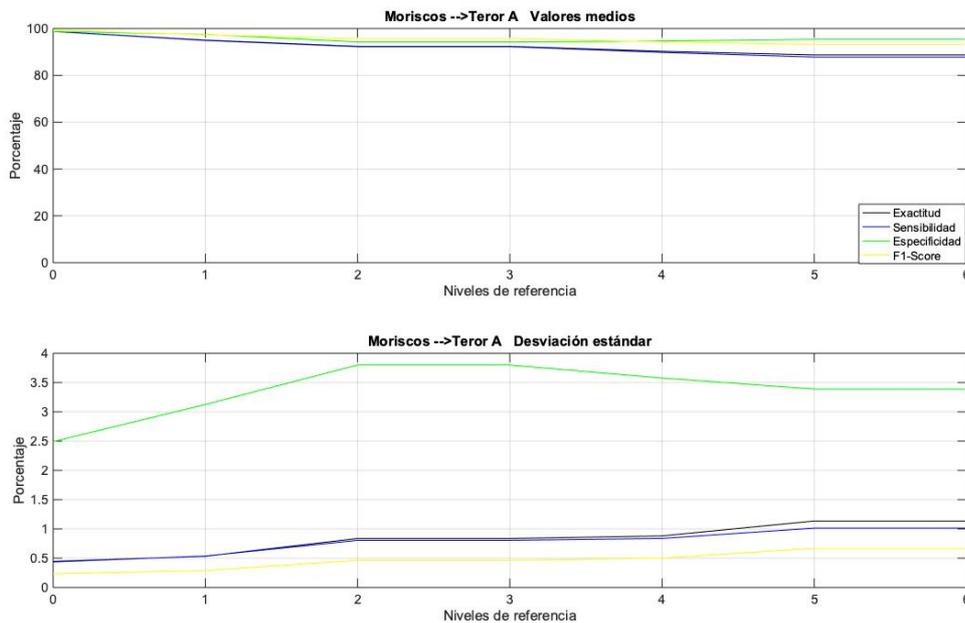


Figura 38 Resultados para el radioenlace Moriscos - Teror

La Figura 39 muestra los resultados obtenidos para el radioenlace Anocheza - Birmagen. Se observa que para todos los parámetros de calidad los porcentajes de valores medios disminuyen conforme cambia el nivel de referencia. Sin embargo, para la sensibilidad, exactitud y F1-Score se mantiene en porcentajes altos superiores al 79%. Asimismo, estos mismos parámetros presentan una desviación estándar muy buena, menor al 4%. En cuanto a la especificidad no se obtienen resultados tan buenos pues llega a valores medios del 60% y porcentajes de desviación estándar superiores al 10%. Por ello en este radioenlace tomamos como nivel de referencia el 2 para nunca superar este 10%.

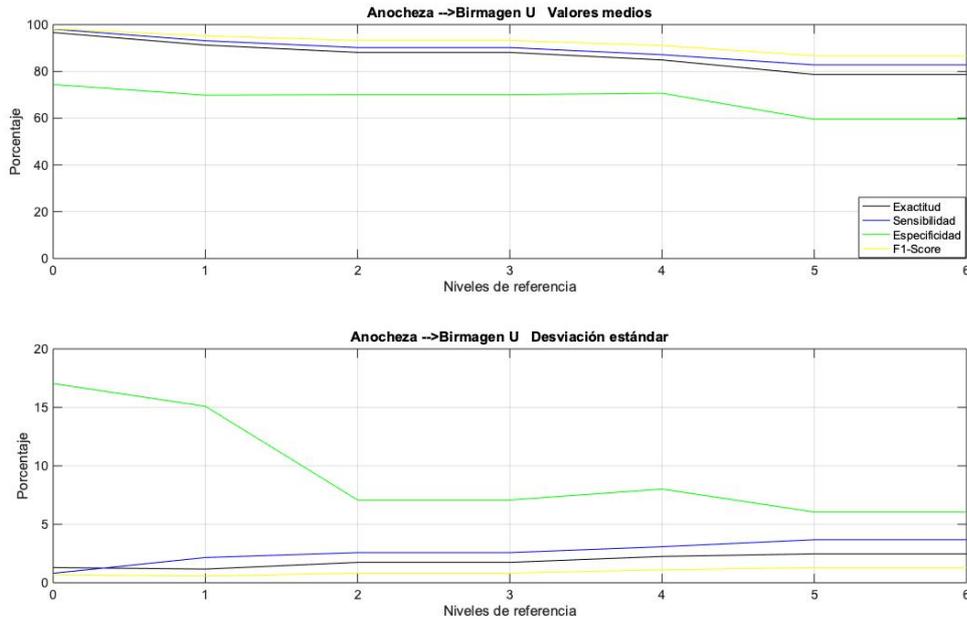


Figura 39 Resultados para el radioenlace Anocheza - Birmagen

La Figura 40 muestra los resultados obtenidos para el radioenlace Anocheza – El Reverón. Para este radioenlace tomamos como nivel de referencia el 2 pues es a partir de este nivel donde se presentan los mejores porcentajes para los diferentes parámetros de calidad. Estos porcentajes son muy buenos superando el 90% para los valores medios y el 4% o inferiores para la desviación estándar.

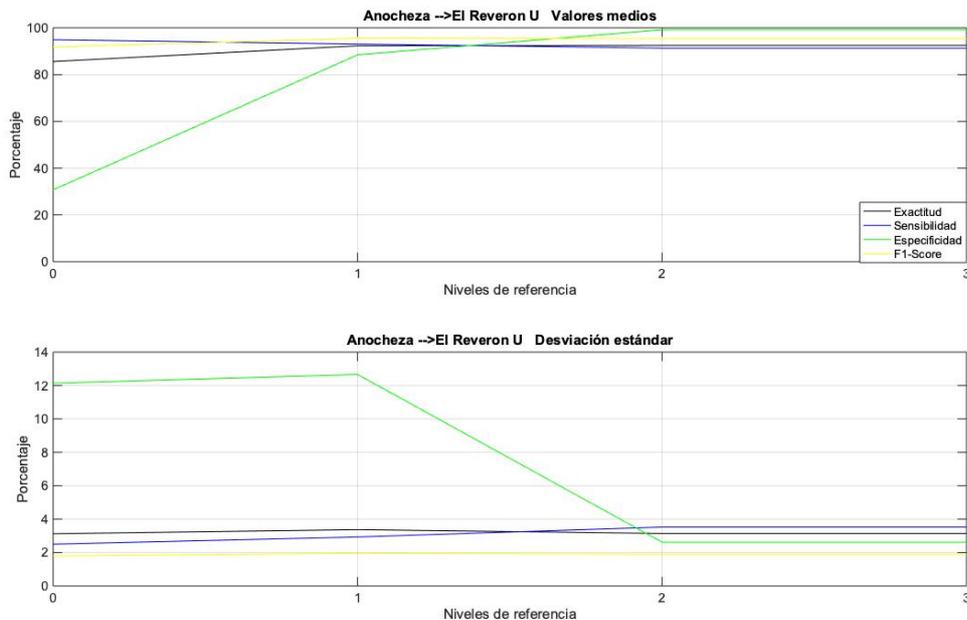


Figura 40 Resultados para el radioenlace Anocheza – El Reverón

La Figura 41 muestra los resultados obtenidos para el radioenlace Suculum - Bailadero. Para este radioenlace tomamos como nivel de referencia el 5 pues es a partir de este nivel donde se presentan los mejores porcentajes para los diferentes parámetros de calidad. Estos porcentajes son muy buenos superando el 85% para los valores medios y menores al 2,5% para la desviación estándar excepto para la sensibilidad donde este valor asciende al 6%.

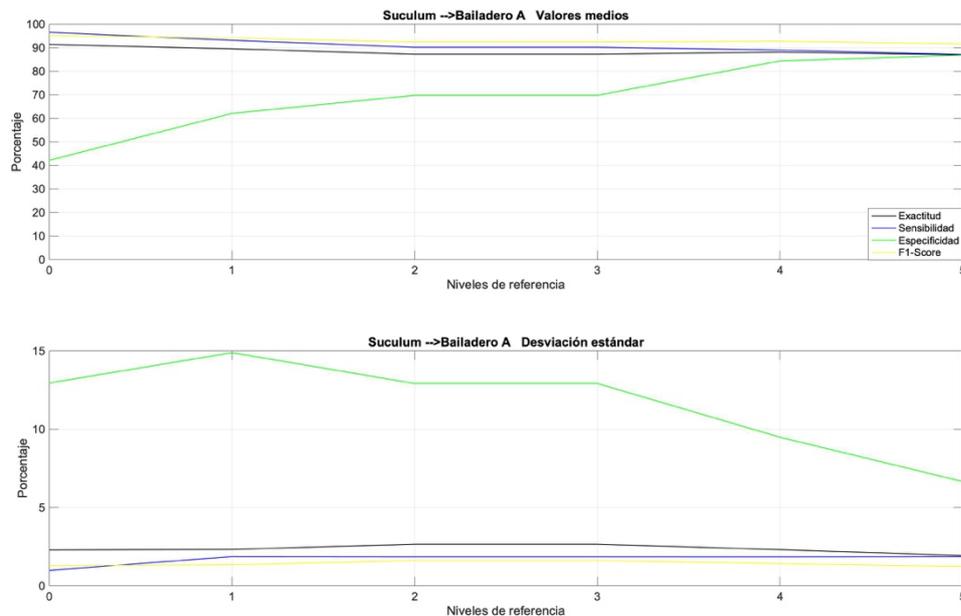


Figura 41 Resultados para el radioenlace Suculum - Bailadero

La Figura 42 muestra los resultados obtenidos para el radioenlace Suculum - Casas cumbre. Destacamos que se presentan valores muy buenos de valores medios ya que los porcentajes son siempre del 90% o superiores. En cuanto a la desviación estándar, si bien no se obtiene un buen resultado para el nivel 0 (no lluvia en su totalidad) ya que es de un 5%, en el siguiente nivel (nivel 1) desciende hasta obtener todos los parámetros de calidad por debajo de un 2%. Por ello en este radioenlace tomaríamos este nivel como referencia.

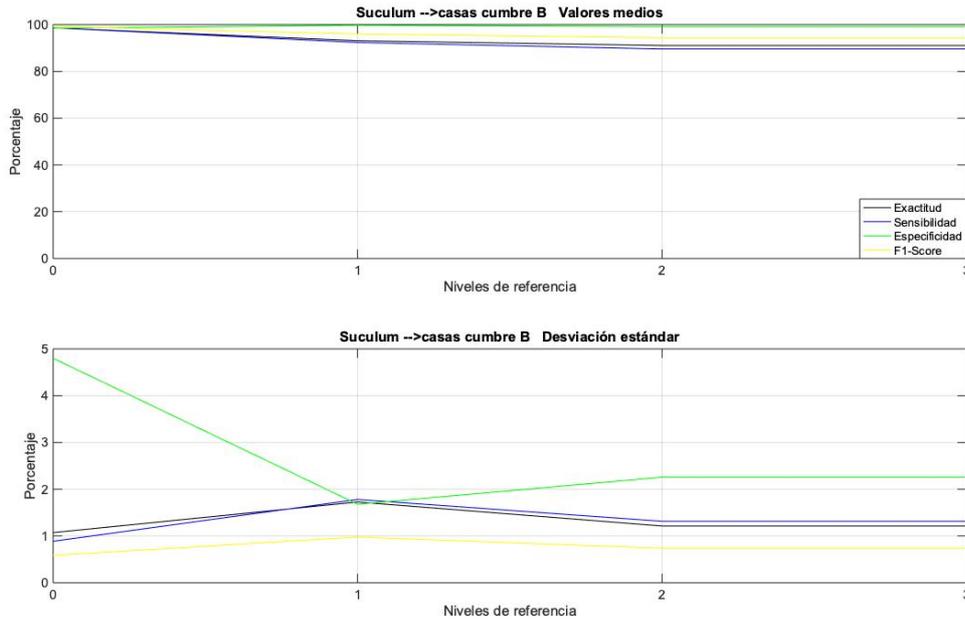


Figura 42 Resultados para el radioenlace Suculum - Casas Cumbre

### 6.4 Experimentos con red neuronal

Observando los resultados se llegó a la conclusión que la mejor combinación es la primera (características originales y 8 máscaras) para todos los radioenlaces pues es la que presenta mejores valores medios y desviación estándar de sensibilidad, especificidad, exactitud y F1-Score. Esto lo podremos comprobar con las Figuras que se presentan en las siguientes páginas.

El eje horizontal de las gráficas corresponde con los distintos niveles de lluvia obtenidos y el eje vertical con los valores medios y desviación estándar de sensibilidad, especificidad, precisión y F1 dados en porcentaje. Estos parámetros corresponden a los colores azul, verde, negro y amarillo respectivamente.

La Figura 43 muestra los resultados obtenidos para el radioenlace Agaete – Montaña de Gáldar. Para este radioenlace tomamos como nivel de referencia el 1. En este nivel es donde se presentan los mejores porcentajes de valores medios y desviación estándar para los diferentes parámetros de calidad siendo superiores al 90%. Asimismo, los porcentajes de la desviación estándar a partir del nivel 1 no superan el 6%

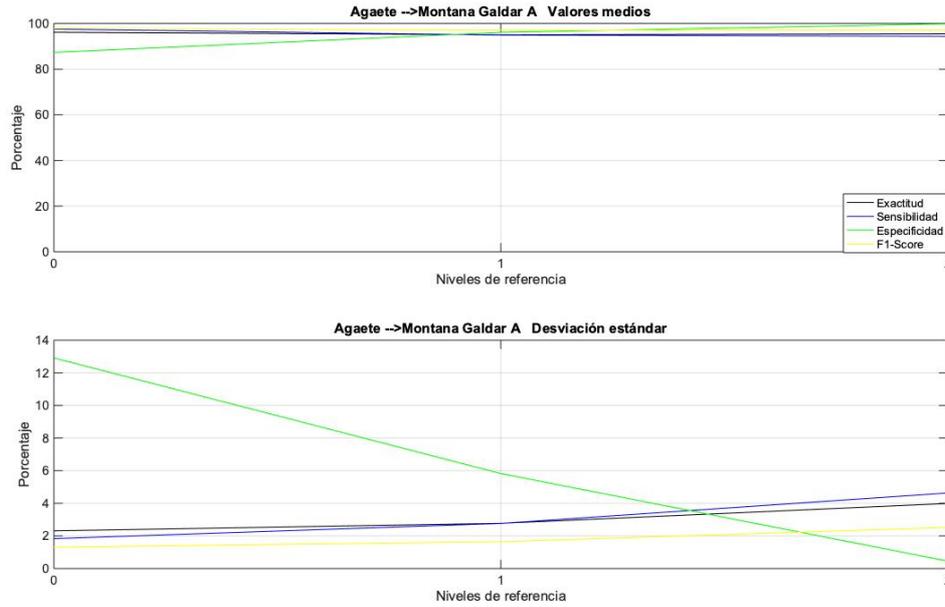


Figura 43 Resultados para el radioenlace Agaete - Montaña de Gáldar

La Figura 44 muestra los resultados obtenidos para el radioenlace Agüimes – Tabaiba. Para este radioenlace tomamos como nivel de referencia el 1. En este nivel es donde se presentan los mejores porcentajes de valores medios y desviación estándar para los diferentes parámetros de calidad alcanzando niveles de valores medios que superan el 95%. Asimismo, los porcentajes de la desviación estándar a partir del nivel 1 son muy bajos, por debajo del 5% y si avanzamos hasta el nivel dos nos encontramos con una desviación estándar del 2%.

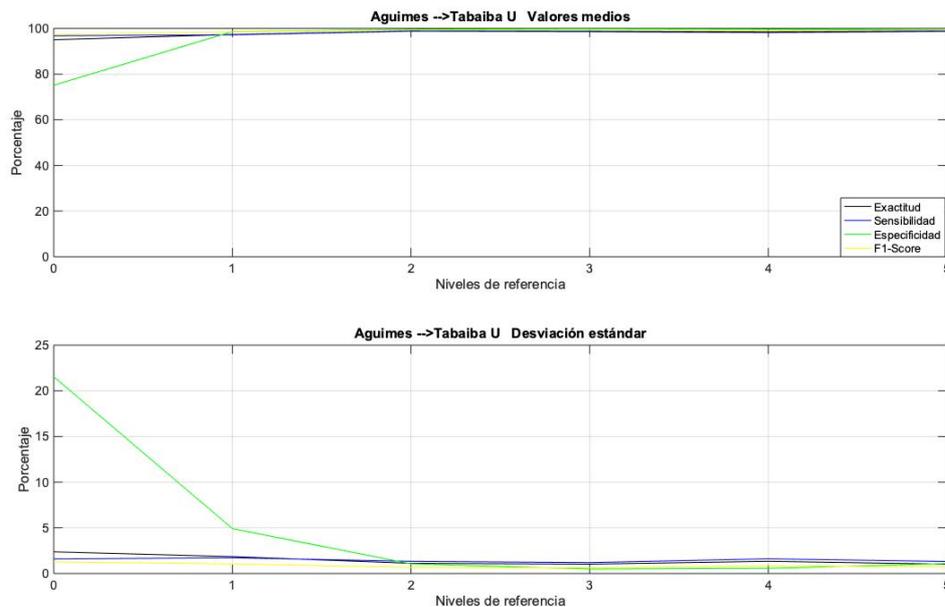


Figura 44 Resultados para el radioenlace Agüimes – Tabaiba

La Figura 45 muestra los resultados obtenidos para el radioenlace Monteluz – Hospital. Para este radioenlace tomamos como nivel de referencia el 1. En este nivel es donde se presentan los mejores porcentajes de valores medios y desviación estándar para los diferentes parámetros de calidad alcanzando niveles de valores medios del 95% y superiores. Asimismo, los porcentajes de la desviación estándar a partir del nivel 1 son muy bajos, por debajo del 1,5%.

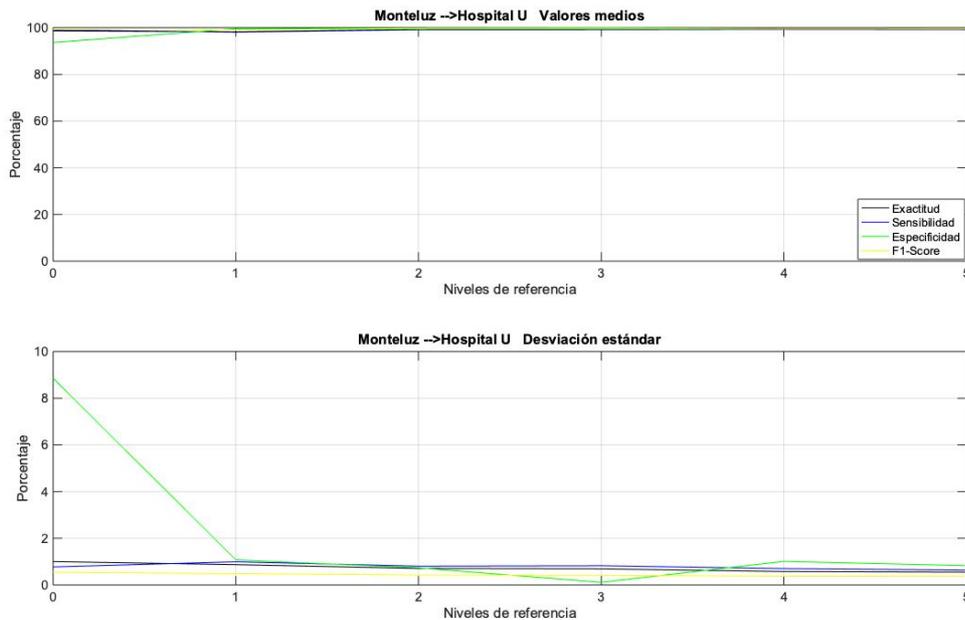


Figura 45 Resultados para el radioenlace Monteluz – Hospital

La Figura 46 muestra los resultados obtenidos para el radioenlace Moriscos - Teror. Destacamos el hecho que con nivel de referencia 0 se dan los mejores resultados de valores medios, que rondan el 100% y un 1% de desviación estándar para la especificidad y menor para todos los demás parámetros de calidad.

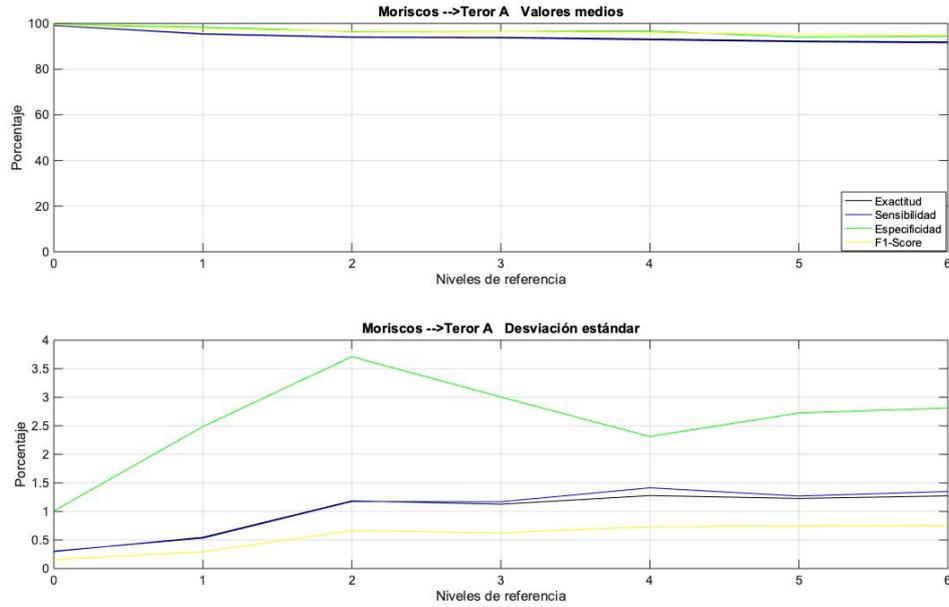


Figura 46 Resultados para el radioenlace Moriscos – Teror

La Figura 47 muestra los resultados obtenidos para el radioenlace Anocheza - Birmagen. En este radioenlace es en el nivel 2 donde se presentan los mejores valores para los diferentes parámetros de calidad alcanzando niveles de valores medios que superan el 95%. La desviación estándar a partir del nivel 1 es muy baja, por debajo del 5%, sin embargo, si avanzamos hasta el nivel dos nos encontramos con una desviación estándar del 2% o menor para todos los parámetros de calidad.

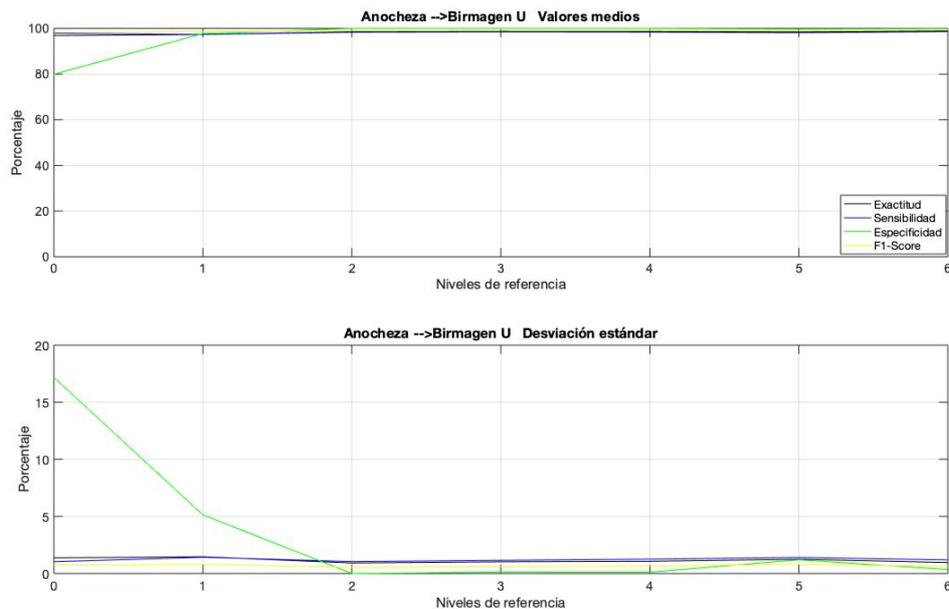


Figura 47 Resultados para el radioenlace Anocheza – Birmagen

La Figura 48 muestra los resultados obtenidos para el radioenlace Anocheza – El Reverón. En este radioenlace tomaríamos como nivel de referencia el 1 pues es en este nivel donde se presentan unos valores medios muy buenos que superan el 95%. Por otro lado, la desviación estándar es baja, por debajo del 4%.

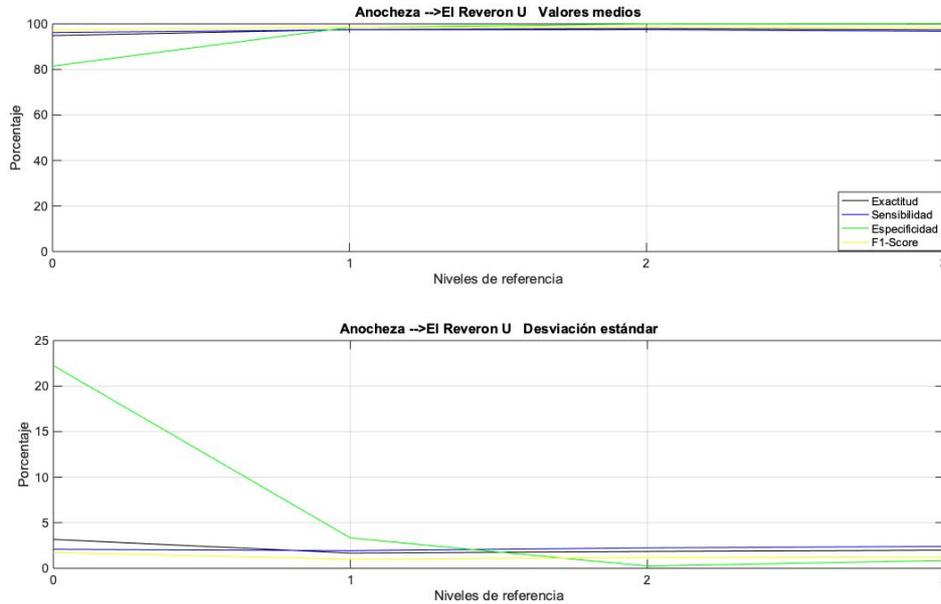


Figura 48 Resultados para el radioenlace Anocheza - El Reverón

La Figura 49 muestra los resultados obtenidos para el radioenlace Suculum - Bailadero. Para este radioenlace tomamos como nivel de referencia el 2 ya que se llega a niveles de valores medios que superan el 95% siempre. Además, la desviación estándar a partir del nivel 2 es baja, por debajo del 2%.

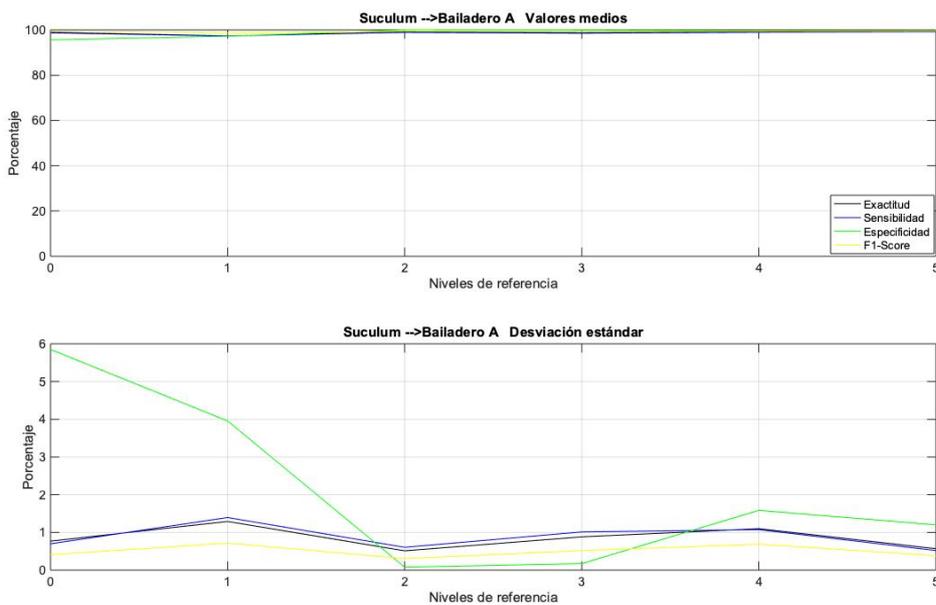


Figura 49 Resultados para el radioenlace Suculum – Bailadero

La Figura 50 muestra los resultados obtenidos para el radioenlace Suculum – Casas Cumbre. Para este radioenlace tomamos como nivel de referencia el 1 ya que se llega a niveles de valores medios que superan el 95% siempre. Además, la desviación estándar a partir del nivel 1 es baja, por debajo del 4% y si avanzamos hasta el nivel 2 ésta es inferior al 2% para todos los parámetros de calidad.

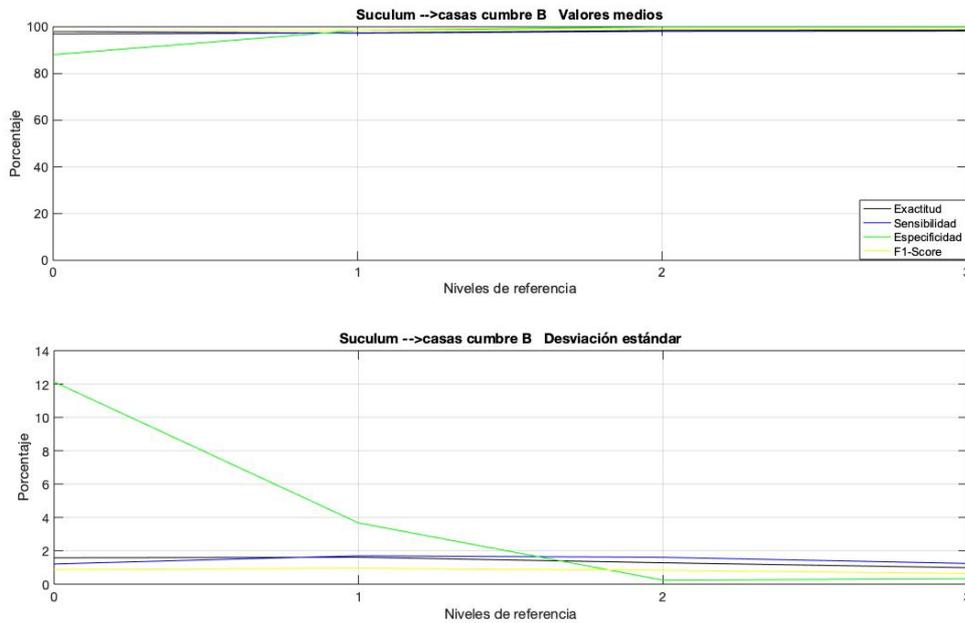


Figura 50 Resultados para el radioenlace Suculum - Casas cumbre

### 6.5 Experimentos con máquina de vectores soportados (SVM)

Se presentan los resultados obtenidos con el método de máquinas de vectores soportados.

Contrastando los resultados se concluyó que se obtienen mejores valores cuando se realiza la combinación característica originales y una máscara en las máquinas de vectores soportados. Esto es así porque todos los radioenlaces presentaron mejores valores medios y desviación estándar de sensibilidad (azul), especificidad (verde), exactitud (negro) y F1-Score (amarillo) para esta opción.

Seguidamente se muestran los resultados obtenidos por medio de una serie de gráficas en las cuales el eje horizontal corresponde con los distintos niveles de lluvia obtenidos y el eje vertical con los valores medios y desviación estándar de sensibilidad, especificidad, precisión y F1 dados en porcentaje. Estos parámetros corresponden a los colores azul, verde, negro y amarillo respectivamente.

La Figura 51 muestra los resultados obtenidos para el radioenlace Agaete – Montaña de Gáldar. En este radioenlace es en el nivel 1 donde se presentan los mejores valores para los diferentes parámetros de calidad alcanzando niveles de valores medios del 90% y superiores. También la desviación estándar en este nivel es baja, del 2% e inferior.

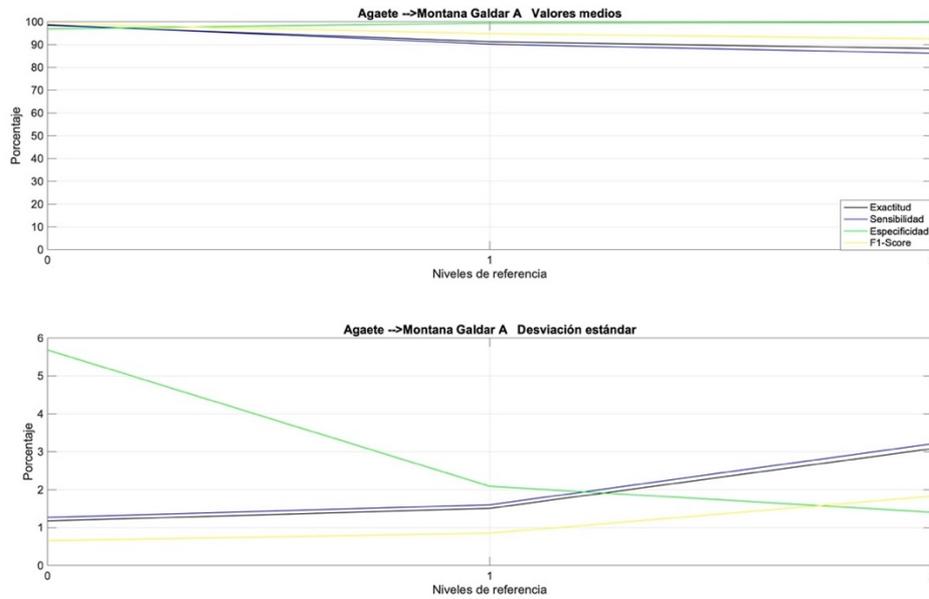


Figura 51 Resultados del radioenlace Agaete - Montaña de Gáldar

La Figura 52 muestra los resultados obtenidos para el radioenlace Agüimes – Tabaiba. Para este radioenlace tomamos como nivel de referencia el 2 ya que se llega a niveles de valores medios que superan el 90%. Además, la desviación estándar a partir del nivel 2 es baja, por debajo del 2%.

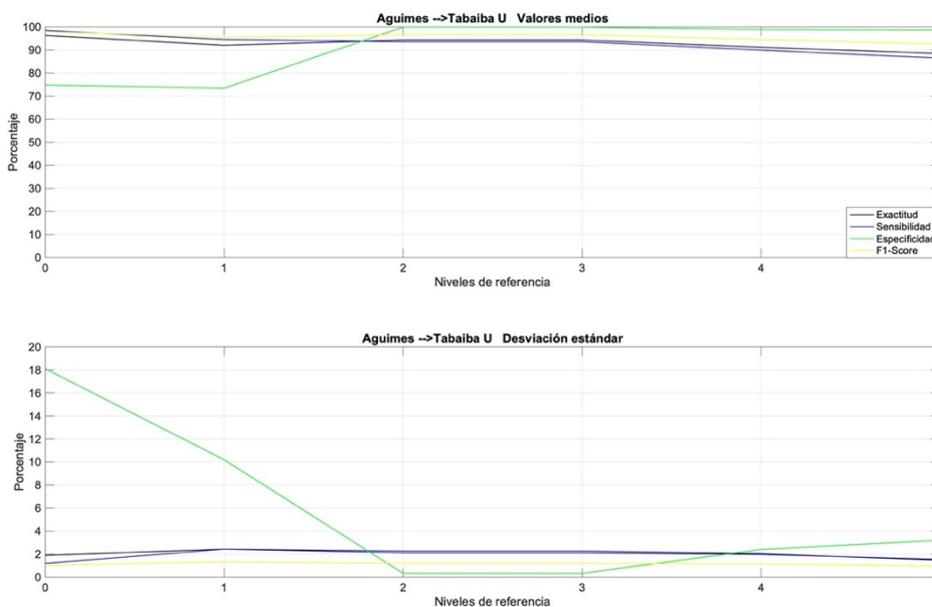


Figura 52 Resultados para el radioenlace Agüimes – Tabaiba

La Figura 53 muestra los resultados obtenidos para el radioenlace Monteluz - Hospital. Destacamos el hecho que con nivel de referencia 0 se dan los mejores resultados de valores medios, estos rondan el 100%. Asimismo, es en este nivel donde la desviación estándar alcanza un nivel más bajo, del 2% (para la especificidad) y por debajo del 0,5% para los demás parámetros de calidad.

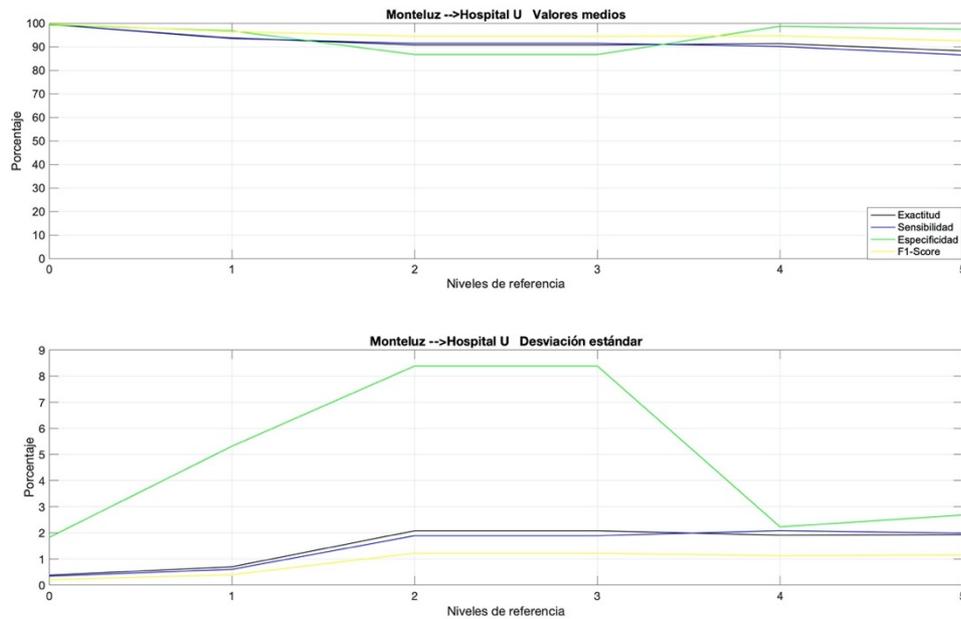


Figura 53 Resultados para el radioenlace Monteluz - Hospital

La Figura 54 muestra los resultados obtenidos para el radioenlace Moriscos – Teror. Al igual que en la Figura 53 destacamos el hecho que con nivel de referencia 0 se dan los mejores resultados de valores medios, estos rondan el 100%. Asimismo, es en este nivel donde la desviación estándar alcanza un nivel más bajo, del 1,2% (para la especificidad) y por debajo del 0,5% para los demás parámetros de calidad.

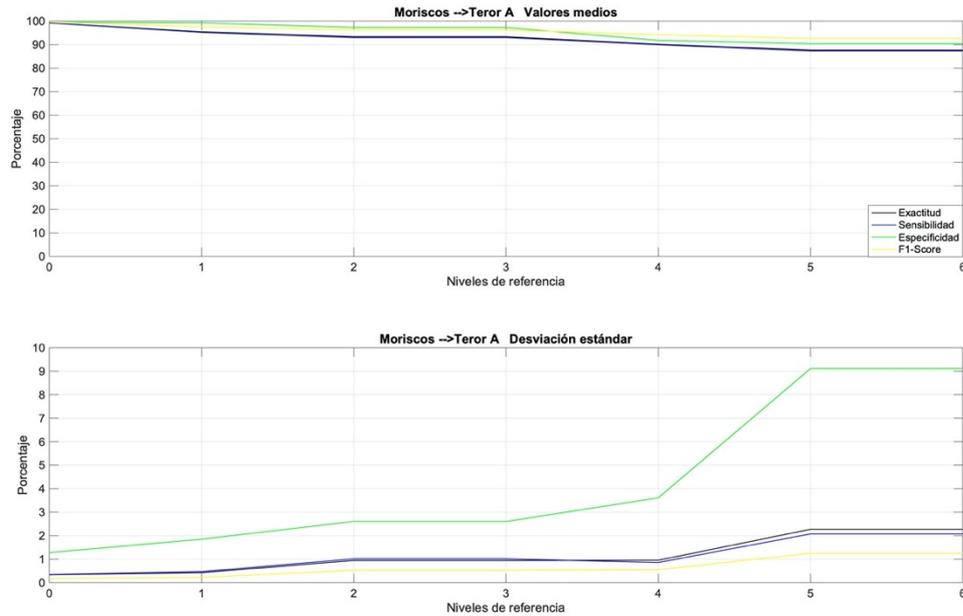


Figura 54 Resultados para el radioenlace Moriscos – Teror

La Figura 55 muestra los resultados obtenidos para el radioenlace Anocheza - Birmagen. Con nivel de referencia 0 se dan los mejores resultados de valores medios, estos rondan el 100%. Asimismo, es en este nivel donde la desviación estándar alcanza un nivel más bajo, del 3% (para la especificidad) y por debajo del 1% para los demás parámetros de calidad.

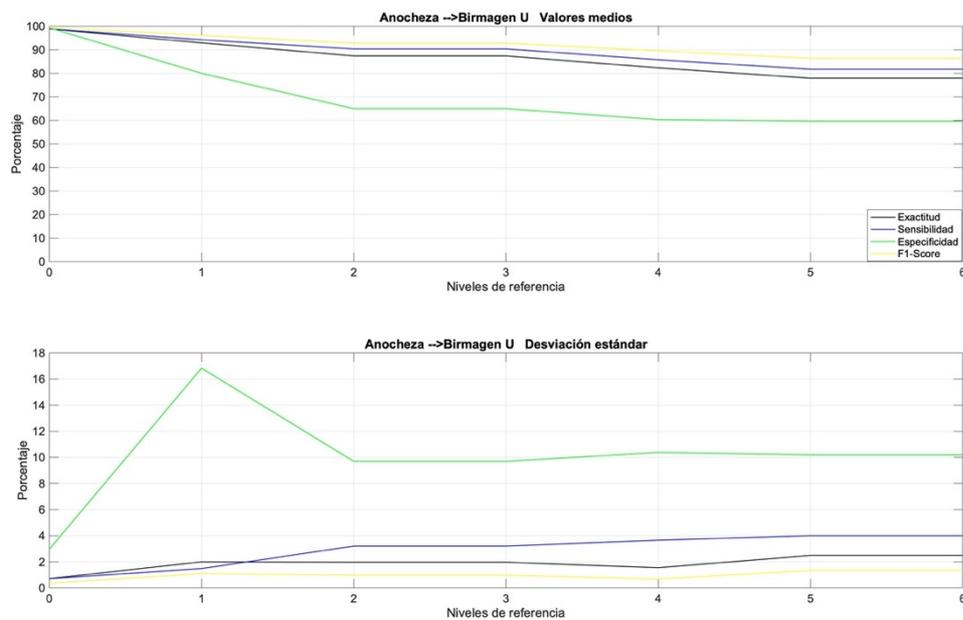


Figura 55 Resultados para el radioenlace Anocheza – Birmagen

La Figura 56 muestra los resultados obtenidos para el radioenlace Anocheza – El Reverón. Para este radioenlace tomamos como nivel de referencia el 1. En este nivel es donde se presentan los mejores porcentajes de valores medios y desviación estándar para los diferentes parámetros de calidad alcanzando niveles de valores medios del 90% y superiores. La desviación estándar de la especificidad es bastante alta ya que ronda el 11%, sin embargo, los demás parámetros de calidad presentan una desviación bastante baja, del 4%.

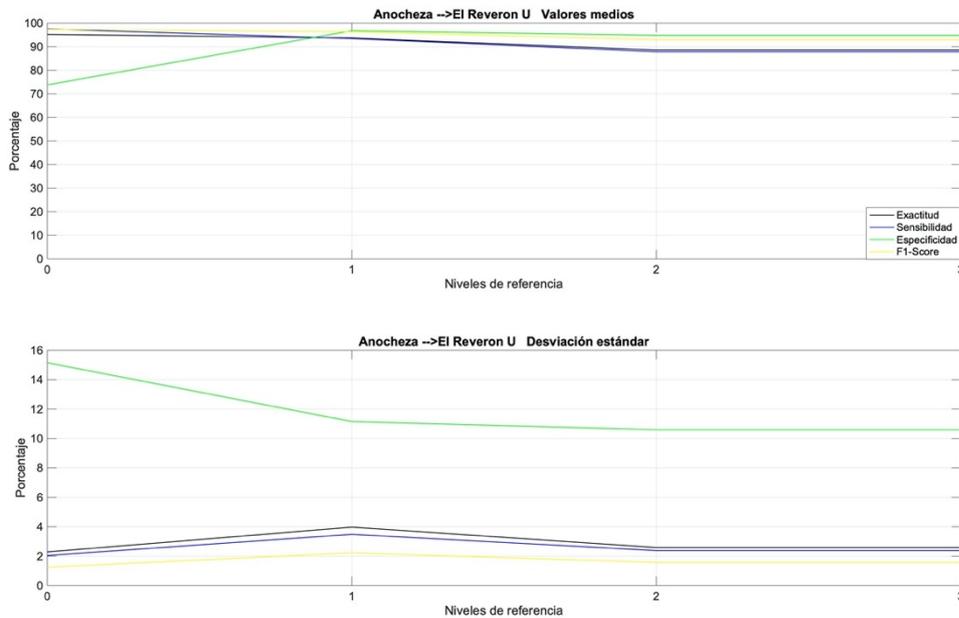


Figura 56 Resultados para el radioenlace Anocheza - El Reverón

La Figura 57 muestra los resultados obtenidos para el radioenlace Suculum - Bailadero. Para este radioenlace tomamos como nivel de referencia el 2 ya que se alcanza niveles de valores medios del 90% y superiores. Por otro lado, la desviación estándar es baja, por debajo del 2% para este nivel.

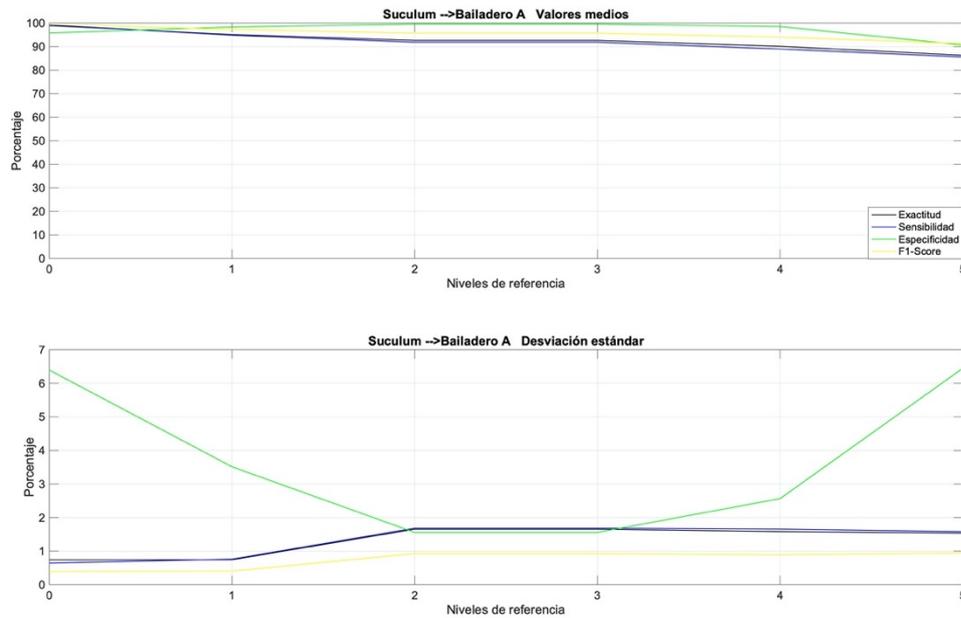


Figura 57 Resultados para el radioenlace Suculum – Bailadero

La Figura 58 muestra los resultados obtenidos para el radioenlace Suculum – Casas Cumbre. Resaltamos que con nivel de referencia 0 se dan los mejores resultados de valores medios, estos rondan el 100%. Asimismo, es en este nivel donde la desviación estándar alcanza un nivel más bajo, del 1% (para la especificidad) e inferiores para los demás parámetros de calidad.

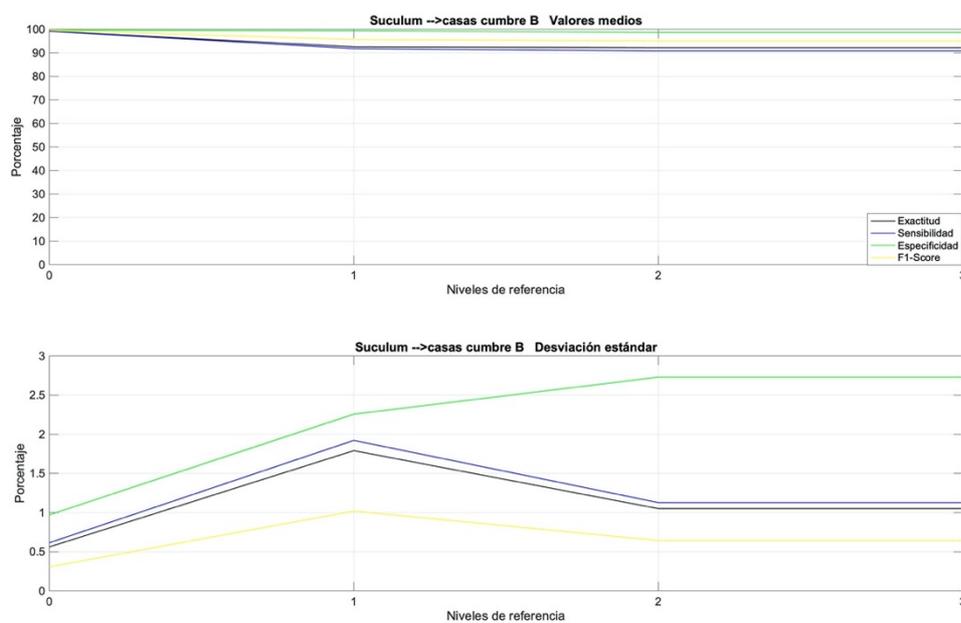


Figura 58 Resultados para el radioenlace Suculum - Casas Cumbre

## 6.6 Resultados finales

En este apartado presentamos el método que mejor ha resultado para cada uno de los ocho radioenlaces que hemos estudiado en este TFG. Asimismo, para cada radioenlace, en las Tablas 10 a 17 presentamos los mejores resultados de nivel de referencia, parámetros de calidad (sensibilidad, exactitud, especificidad, F1-Score) dados con sus media y desviación. En la columna “Error” mostramos los errores de cuantificación. Hemos de aclarar que, dada la forma en que hemos obtenidos los errores, un signo negativo nos indica tendencia a sobredimensionar, es decir, dar valores de lluvia superiores a los reales, y el signo positivo indica la tendencia contraria.

### Agaete – Montaña de Gáldar

En la tabla 10 se muestra los mejores resultados obtenidos para el radioenlace Agaete – Montaña de Gáldar de los distintos parámetros de calidad. Estos se obtienen con regresión lineal (características originales y una máscara).

Método	Nivel ref. / Lluvia acum.	Exactitud $\bar{x} \pm \sigma$	Sensibilidad $\bar{x} \pm \sigma$	Especificidad $\bar{x} \pm \sigma$	F1-Score $\bar{x} \pm \sigma$	Error $\bar{x} \pm \sigma$
Regresión lineal	0	98,77% ±	98,79% ±	98,73% ±	99,31% ±	-0,1727 ±
	0 mm/h	1,123%	1,202%	2,693%	0,6273%	0.4543

Tabla 10 Resultados finales para el radioenlace Agaete – Montaña de Gáldar

### Agüimes – Tabaiba

En la tabla 11 se presenta los mejores valores obtenidos para el radioenlace Agüimes – Tabaiba de los distintos parámetros de calidad. Estos se obtienen con red neuronal (características originales y ocho máscaras).

Método	Nivel ref. / Lluvia acum.	Exactitud $\bar{x} \pm \sigma$	Sensibilidad $\bar{x} \pm \sigma$	Especificidad $\bar{x} \pm \sigma$	F1-Score $\bar{x} \pm \sigma$	Error $\bar{x} \pm \sigma$
Red Neuronal	1	97,32% ±	97,2% ±	98,61% ±	98,48% ±	-0,0224 ±
	0,2 mm/h	1,85%	1,726%	4,916%	1,048%	0,1996

Tabla 11 Resultados finales para el radioenlace Agüimes - Tabaiba

### Monteluz – Hospital

En la tabla 12 se muestra los mejores resultados obtenidos para el radioenlace Monteluz – Hospital obtenidos de los distintos parámetros de calidad. Estos se obtienen con red neuronal (características originales y ocho máscaras).

Método	Nivel ref. / Lluvia acum.	Exactitud $\bar{x} \pm \sigma$	Sensibilidad $\bar{x} \pm \sigma$	Especificidad $\bar{x} \pm \sigma$	F1-Score $\bar{x} \pm \sigma$	Error $\bar{x} \pm \sigma$
Red Neuronal	1	98,25% ±	98,11% ±	99,5% ±	99,01% ±	-0,0078 ±
	0,2 mm/h	0,8704%	0,994%	1,075%	0,4896%	0,0882

Tabla 12 Resultados finales para el radioenlace Monteluz - Hospital

### Moriscos – Teror

En la tabla 13 se presenta los mejores valores obtenidos para el radioenlace Moriscos – Teror de los distintos parámetros de calidad. Estos se obtienen con red neuronal (características originales y ocho máscaras).

Método	Nivel ref. / Lluvia acum.	Exactitud $\bar{x} \pm \sigma$	Sensibilidad $\bar{x} \pm \sigma$	Especificidad $\bar{x} \pm \sigma$	F1-Score $\bar{x} \pm \sigma$	Error $\bar{x} \pm \sigma$
Red Neuronal	0	99,13% ±	99,1% ±	99,62% ±	99,54% ±	-0,0088 ±
	0 mm/h	0,2969%	0,3048%	1,009%	0,1563%	0,1999

Tabla 13 Resultados finales para el radioenlace Moriscos - Teror

### Anocheza – Birmagen

En la tabla 14 se muestra los mejores valores de los distintos parámetros de calidad obtenidos para el radioenlace Anocheza - Birmagen. Estos se obtienen con SVM (características originales y una máscara).

Método	Nivel ref. / Lluvia acum.	Exactitud $\bar{x} \pm \sigma$	Sensibilidad $\bar{x} \pm \sigma$	Especificidad $\bar{x} \pm \sigma$	F1-Score $\bar{x} \pm \sigma$	Error $\bar{x} \pm \sigma$
SVM	0	98,87% ±	98,85% ±	99,31% ±	99,4% ±	-0,3416 ±
	0 mm/h	0,7087%	0,7065%	2,962%	0,3717%	1,3823

Tabla 14 Resultados finales para el radioenlace Anocheza – Birmagen

### Anocheza -El Reverón

En la tabla 15 se expone los mejores valores obtenidos para los distintos parámetros de calidad para el radioenlace Anocheza – El Reverón. Estos se obtienen con red neuronal (características originales y ocho máscaras).

Método	Nivel ref. / Lluvia acum.	Exactitud $\bar{x} \pm \sigma$	Sensibilidad $\bar{x} \pm \sigma$	Especificidad $\bar{x} \pm \sigma$	F1-Score $\bar{x} \pm \sigma$	Error $\bar{x} \pm \sigma$
Red	2	97,95% ±	97,44% ±	99,94% ±	98,68% ±	-0,0530 ±
Neuronal	0,4 mm/h	1,848%	2,231%	0,2372%	1,164%	0,3622

Tabla 15 Resultados finales para el radioenlace Anocheza – El Reverón

### Suculum – Bailadero

En la tabla 16 se presenta los mejores valores obtenidos para los distintos parámetros de calidad para el radioenlace Suculum - Bailadero. Estos se obtienen con red neuronal (características originales y ocho máscaras).

Método	Nivel ref. / Lluvia acum.	Exactitud $\bar{x} \pm \sigma$	Sensibilidad $\bar{x} \pm \sigma$	Especificidad $\bar{x} \pm \sigma$	F1-Score $\bar{x} \pm \sigma$	Error $\bar{x} \pm \sigma$
Red	2	99,09% ±	98,92% ±	99,99% ±	99,45% ±	-0,0172 ±
Neuronal	0,4 mm/h	0,5128%	0,6054%	0,0787%	0,3062%	0,1828

Tabla 16 Resultados finales para el radioenlace Suculum - Bailadero

### Suculum – Casas Cumbre

En la tabla 17 se muestra los mejores valores obtenidos para los distintos parámetros de calidad para el radioenlace Agaete – Montaña de Gáldar. Estos se obtienen con red neuronal (características originales y ocho máscaras).

Método	Nivel ref. / Lluvia acum.	Exactitud $\bar{x} \pm \sigma$	Sensibilidad $\bar{x} \pm \sigma$	Especificidad $\bar{x} \pm \sigma$	F1-Score $\bar{x} \pm \sigma$	Error $\bar{x} \pm \sigma$
Red	2	98,47% ±	98,06% ±	99,91% ±	99,00% ±	-0,0160 ±
Neuronal	0,4 mm/h	1,282%	1,608%	0,2392%	0,8268%	0,1170

Tabla 17 Resultados finales para el radioenlace Suculum – Casas Cumbre



## Capítulo 7 – Conclusiones y Líneas Futuras

En este capítulo presentan las conclusiones obtenidas tras el desarrollo del TFG. Además, se exponen las posibles líneas de investigación para futuros trabajos relacionamos con técnicas de aprendizaje máquina para detección y cuantificación de lluvia mediante radioenlaces terrenos a frecuencia de gigahercios.

### 7.1 Conclusiones

A lo largo de este Trabajo de Fin de Grado se ha logrado un objetivo principal, la implementación de algoritmos de Machine Learning para la detección y cuantificación de precipitación de lluvia basándonos en los niveles de potencia involucrados en comunicaciones terrenas a frecuencias de microondas distribuidas por las zonas de observación. Asimismo, este objetivo principal se ha visto complementado por unos objetivos secundarios como son:

- La realización del estudio de las precipitaciones y su influencia sobre la atenuación de señales en radioenlaces terrenos
- El estudio de los datos para llevar a cabo un preprocesado y un procesado que nos permite transformar los datos de tal manera que podemos hacer un uso sencillo de ellos.

De los resultados obtenidos y presentados se extraen las siguientes conclusiones. El método que mejores resultados de detección y cuantificación presenta es la red neuronal. Como se puede ver en el capítulo anterior en el apartado (6.6), de los test realizados para los ocho radioenlaces el método que mejores valores presenta para cada uno es este teniendo casos que en general pasan del 98% en los diferentes parámetros de calidad (sensibilidad, exactitud, Especificidad y F1-Score).

## 7.2 Líneas futuras

Las posibles líneas futuras de trabajos que se podrían llevar a cabo para complementar este Trabajo de Fin de Grado son:

- Selección de un modelo “generalizable” que sirva para cualquier radioenlace, esto es, que no sea específico de cada uno.
- Integración en la base de datos de los radioenlaces datos con un tiempo de muestreo menor a quince minutos. Así, obtendríamos resultados más precisos.
- Integración de nuevas estaciones meteorológicas más cercanas a los radioenlaces pues mejoraría la sincronía de los datos recogidos. Una opción sería contactar con más instituciones públicas, empresas o particulares que compartan sus datos.
- Estudio de nuevas formas de extracción de características que puedan mejorar la diferenciación de lluvia y no lluvia. Esto podría ser de interés en radioenlaces donde nos pudiéramos encontrar con casuísticas nuevas.
- Implementación con técnicas de Machine Learning de una metodología que nos permita una predicción a un tiempo futuro (por ejemplo, 15, 30 minutos u horas) puesto que en este TFG se ha realizado fundamentalmente una observación.



## Bibliografía

- AEMET. (2020). Obtenido de <http://www.aemet.es/es/portada>
- AEMET. (7 de julio de 2020b). Obtenido de <http://www.aemet.es/es/eltiempo/observacion/ultimosdatos?k=coo&l=C658X&w=0&datos=det&x=h24&f=temperatura>
- Amat Rodrigo, J. (Abril de 2017). *Máquinas de Vector Soporte (Support Vector Machines, SVMs)*. Obtenido de [https://www.cienciadedatos.net/documentos/34\\_maquinas\\_de\\_vector\\_soporte\\_support\\_vector\\_machines](https://www.cienciadedatos.net/documentos/34_maquinas_de_vector_soporte_support_vector_machines)
- Berk, R. A. (2017). *Statistical Learning from a Regression Perspective*. University of Pennsylvania; Philadelphia, PA: Springer.
- Cherkassky, V., & Murier, F. (2007). *Learning from data*.
- Cibernéticas, S. S. (2020). Obtenido de SEAC: <http://www.seac.es/es/productos/estaciones/>
- Cook, D. R., & Weisberg, S. (1999). *Applied regression including computing and graphics*. New York: Wiley.
- Daicy Cuenca, D. L. (julio de 2020). *SUPPORT VECTOR MACHINE*. Obtenido de [https://rpubs.com/Dario\\_BSC/SVM](https://rpubs.com/Dario_BSC/SVM)
- González Gómez, P. d. (2017). *Diseño de técnicas de estimación de lluvia a partir de niveles de potencia de radioenlaces*.
- Guerra Moreno, I. D. (2017). *"Algoritmos de aprendizaje aplicados a la monitorización meteorológica mediante redes inalámbricas de comunicaciones distribuidas"*. Tesis doctoral, Universidad de las Palmas de Gran Canaria.
- Gunn, G. D. (1949). *The terminal velocity of fall for water droplets in stagnant air, J. Meteorol., vol. 6, (4), pp. 243-248*.
- Hagit Messer, A. Z. (2006). *Environmental Monitoring by Wireless Communication Networks*. BREVIA.
- He, H., Bai, Y., Garcia, E., & Li, S. (2008). ADASYN: Adaptive Synthetic Sampling Approach for Imbalanced Learning. *International Joint Conference on Neural Networks*. Granada.

- Ho, C. (2005). Modeling and Simulation for Realistic Propagation Environments of Communications Signals at SHF Band. *Annual ITEA Technology Review Conference Sponsored by the International Test and evaluation Association (ITEA)*. Atlanta, Georgia.
- ITU. (2005). *UIT-R P.838-3. Modelo de la atenuación específica debida a la lluvia para los métodos de predicción, 2005*.
- ITU. (1995). *UIT-R, "Proyecto de manual sobre radiometeorología," marzo 1995*.
- ITU\_R. (2013). *UIT-R P.839-4 Modelo de estimación de la altura de la lluvia para utilizar en los métodos de predicción, 09/2013*.
- ITU-R'. (1997). *UIT-R P.676-3. "Atenuación debida a los gases atmosféricos"*.
- ITU-R. (2015). *ITU-R V.431-8. Nomenclatura de las bandas de frecuencias y de las longitudes desde onda empleadas en telecomunicaciones, 2015*.
- Kaur, G. (2012). *Meteorological data mining techniques: A survey. International Journal of Emerging Technology and Advanced Engineering, 2(8), 325-327, 2012*.
- Kubat, M. (2017). *An introduction to machine learning*. Springer.
- Liebe, H. e. (1993). *Propagation modeling of moist air and suspended water/ice particles at frequencies below 1000 GHz. In In AGARD, Atmospheric Propagation Effects Through Natural and Man- Made Obscurants for Visible to MM-Wave Radiation 11 p (SEE N94-30495 08-32), volume 1*.
- Menacho, C. (2014). Modelo de regresión lineal con redes neuronales. *Anales científicos, 75(2), 253-260*.
- Núñez Vivero, L. M. (2018). *Segmentación de imágenes de resonancia magnética cerebral mediante redes neuronales artificiales convolucionales*. Madrid.
- Núñez, S., & Gentile, E. (2004). *La observación Meteorológica, 2004*.
- Navarro Mesa, J. L., & Ravelo García, A. (2017-2021). Proyecto "*Sistema de vigilancia meteorológica para el seguimiento de riesgos medioambientales*". Acrónimo: VIMETRIC-MAC. Entidades beneficiarias: Universidad de las Palmas de Gran Canaria (Jefe de Filas), Universidade da Madeira, Laboratório Regional de Engenharia Civil (Madeira). Participantes asociados: Dirección General de Telecomunicaciones y Nuevas Tecnologías (Gobierno de Canarias), Dirección General de Seguridad y Emergencias (Gobierno de Canarias), Agencia Estatal de Meteorología. Duración: Desde 01/01/2017 hasta 30/06/2021

- RAE. (29 de 04 de 2020). RAE. Obtenido de <https://dle.rae.es/meteoro>
- Rayitsfeld, A. (2012). *Rana Samuels, Artem Zinevich, Uri Hadar, Pinhas Alpert. Atmospheric Research. Comparison of two methodologies for long term rainfall monitoring using a commercial microwave communication system, 2012.*
- Renom, M. (2011). *Principios básicos de las mediciones atmosféricas, 2011.*
- RG13/RG13H, P. (2020). Obtenido de vaisala: <https://www.vaisala.com/es/products/instruments-sensors-and-other-measurement-devices/weather-stations-and-sensors/rg13-rg13h>
- Rodrigo, J. A. (Agosto de 2016). *Regresión logística simple y múltiple.* Obtenido de [https://www.cienciadedatos.net/documentos/27\\_regresion\\_logistica\\_simple\\_y\\_multiple#regresión\\_log%C3%ADstica\\_múltiple](https://www.cienciadedatos.net/documentos/27_regresion_logistica_simple_y_multiple#regresión_log%C3%ADstica_múltiple)
- Rodríguez Jiménez, R. M. (2004). *Águeda Benito Capa, Adelaida Portela Lozano. Meteorología y Climatología, 2004.*
- UIT-R. (2019). P.676-12. *Atenuación debida a los gases atmosféricos y efectos conexos. Propagación de las ondas radioeléctricas.*
- Valencia, U. I. (07 de 05 de 2020). Obtenido de <https://www.universidadviu.es/que-es-el-sistema-tetra/>
- WMO. (29 de 04 de 2020). *Observation components of the Global Observing System.* Obtenido de <https://www.wmo.int/pages/prog/www/OSY/Goscomponents.html>
- Zinevich, A. (2009). *Hagit Messer, Pinhas Alpert. Frontal rainfall observation by a comercial microwave communication network. Tel Aviv University, Israel, 2009.*



**Parte II:**

**Pliego de Condiciones**



## Pliego de Condiciones

El presente Trabajo Fin de Grado comprende toda una serie de métodos y aplicaciones con técnicas de aprendizaje máquina para detección y cuantificación de lluvia mediante radioenlaces terrenos a frecuencia de gigahercios. En el presente pliego de condiciones se trata tanto las condiciones legales, administrativas, derechos de explotación, restricciones de uso y garantías adicionales como las herramientas software como hardware necesarios para la elaboración del presente trabajo.

### Condiciones legales y administrativas

A continuación, se exponen los aspectos de tipo legal y administrativo que habrán de ser tenidos en cuenta para los futuros usuarios. Estas condiciones establecen las obligaciones, derechos y deberes que los autores adquieren con los usuarios del sistema implementado en este trabajo de fin de grado, y a su vez de los usuarios con los autores.

### Derechos de explotación y restricciones de uso

La Universidad de Las Palmas de Gran Canaria es quien ostenta los derechos de explotación y cualquier usuario debe de estar de acuerdo en obligarse por los términos y condiciones establecidas en este Pliego de Condiciones, aceptando todos sus términos y estipulaciones; de igual manera se deberá de cumplir lo establecido en el “*Reglamento general de los trabajos de fin de grado y de fin de máster de las titulaciones oficiales de la Universidad de Las Palmas de Gran Canaria*”, aprobado por acuerdo de Consejo de Gobierno de la ULPGC de 28 de noviembre de 2019 (BOULPGC 5 diciembre 2019), que en su capítulo VI reglamenta los derechos de propiedad intelectual, de explotación industrial y protección de datos personales.

El uso de programas y base de datos o de una copia en ordenador personal de los mismos, será bajo la autorización expresa del autor, el o los tutores del trabajo fin de grado y el director de la Escuela de Ingeniería de Telecomunicación y Electrónica de la Universidad de Las Palmas de Gran Canaria.

El usuario no podrá realizar ingeniería inversa, decompilación o desamblado de los programas. El usuario podrá transferir los programas y bases de datos a un tercero, siempre que no tenga copias de estos, incluyendo posibles actualizaciones o retener material escrito adicional que acompañe a los programas y la base de datos.

## **Garantías y consideraciones adicionales**

La autora y los directores del TFG garantizan que los programas y la base de datos funcionarán correctamente en el momento de la correcta instalación. También se garantiza que el soporte en el cual está grabado el programa no contendrá defectos en el momento de la adquisición de este.

Con la única excepción de lo expresamente expuesto en el párrafo anterior, los programas y bases de datos han sido creados sin garantías de ningún tipo. No se asegura, garantiza, o realiza ninguna declaración respecto al uso o los resultados derivados de la utilización de los programas o la base de datos. Tampoco se garantiza que la operación de los programas sea ininterrumpida o sin errores.

La confiabilidad de operación del programa puede ser afectada por factores adversos a los que se denominan fallas del sistema. En estos se incluyen fluctuaciones en el sistema de alimentación eléctrico, errores en el funcionamiento del hardware del ordenador personal, sistema operativo o entorno del ordenador personal o entorno de desarrollo y ejecución usado para realizar la aplicación, errores de instalación, problemas de compatibilidad de software y hardware, fallas o funcionamientos incorrectos de equipos de control, fallas por uso, o errores por parte del usuario o autor de la aplicación.

Cualquier aplicación donde una falla de sistema pueda crear un riesgo de daño a la propiedad o personas, no puede ser imputada al programa. Para evitar daños, el usuario deberá tomar razonablemente, todas las precauciones necesarias.

Las garantías y consideraciones aquí expresadas se regirán por las leyes del estado español. El usuario acepta la jurisdicción exclusiva de los tribunales de este país con relación a cualquier disputa que pudiera derivarse de los derechos y garantías estipuladas en este documento.

En ningún caso serán la autora, ni los tutores del TFG, ni la Escuela de Ingeniería de Telecomunicación y Electrónica de la Universidad de Las Palmas de Gran Canaria responsables de los perjuicios directos, indirectos, incidentales, probatorios o de los consiguientes, gastos, lucro cesante, pérdida de ahorros, interrupción de negocios, pérdida de información comercial o de negocio, o cualquier otra pérdida que resulte del uso o de la incapacidad de usar los programas o la base de datos. El usuario conoce y acepta esta asignación de riesgo, así como el resto de las garantías, derechos, deberes y restricciones. Se rechaza cualquier otra garantía, derecho, deber o restricción que no haya sido citada directamente en el Pliego de Condiciones.

## Herramientas utilizadas

Los recursos empleados para el desarrollo del presente Trabajo Fin de Grado son:

**Ordenadores:** En los que se ha llevado a cabo todas las modificaciones de los programas, toda la batería de experimentos, fase de preprocesado, extracción de características, entrenamiento, test y extracción de resultados finales.

Las características técnicas de los equipos son las siguientes:

- Procesador Intel® Core™ i7 a 3 GHz, 16 GB de memoria RAM y disco duro de 1 TB SSD y Sistemas Operativos macOS Mojave (ver. 10.14.6) a 64 bits.
- Procesador Quad-Core Intel Core i5 a 1,4GHz (4 núcleos), 8 GB de memoria RAM, disco duro de 121 GB y Sistemas Operativos macOS Catalina (ver. 10.15.4) a 64 bits.

**Microsoft Office Word para Mac versión 16.39:** Aplicación de la suite ofimática Microsoft Office utilizada para la redacción del documento.

**Microsoft Office Excel para Mac versión 16.39:** Aplicación de la suite ofimática Microsoft Office utilizada para la elaboración de las tablas presentes en el TFG.

**Microsoft Office Power Point para Mac versión 16.39:** Aplicación de la suite ofimática Microsoft Office utilizada para la elaboración de la presentación del presente TFG.

**MatLab® versión 2019b,** 64 bits, para el desarrollo de la totalidad de los algoritmos implementados en este TFG.

**Base de datos:** Específica para el presente TFG en el marco del Proyecto VIMETRI-MAC (Navarro Mesa & Ravelo García, 2017-2019). Los datos meteorológicos han sido proporcionados por la AEMET, y los datos de radioenlaces recogidos por los equipos de las antenas emisoras han sido proporcionados por el Gobierno de Canarias, Dirección General de Telecomunicaciones, obtenidos dentro de su plan de Red de Emergencias y Seguridad de Canarias (RESCAN). Los datos solo se podrán usar previa autorización expresa de los responsables del Proyecto VIMETRI-MAC y, en su caso, se hará con la debida referencia a la propiedad y autoría.



# Parte III:

# Presupuesto



## Presupuesto

### Declaración Jurada

Doña Sonia Fuente González, autora del presente Trabajo Fin de Grado, declara que:

El Trabajo Fin de Grado con título “*Técnicas de Aprendizaje Máquina para Detección y Cuantificación de Lluvia Mediante Radioenlaces Terrenos a Frecuencia de Gigahercios*”, desarrollado en la Escuela de Ingeniería de Telecomunicación y Electrónica de la Universidad de Las Palmas de Gran Canaria, con un peso de 12 ECTS, correspondiente a 300 horas, de trabajo presencial y no presencial, tiene un coste total de *diecisiete mil nueve euros con noventa y seis céntimos (17.009,96 €)*, correspondiente a la suma de las cantidades consignadas y justificadas en los apartados que se señalan a continuación.

Firmando la presente para que así conste a los efectos oportunos.

Las Palmas de Gran Canaria, 23 de julio de 2020

Autora del TFG:

Sonia Fuente González

## Desglose del presupuesto

El presupuesto del TFG realizado se ha obtenido según los precios de mercado actual, y de las orientaciones del Colegio Oficial de Graduados e Ingenieros Técnicos de Telecomunicación (COITT).

Dicho informe alberga la estimación de los gastos generados durante su realización y se fundamenta sobre los siguientes conceptos, detallados a continuación:

- C-1. Trabajo tarifado por tiempo empleado.
- C-2. Amortización del inmovilizado material.
- C-3. Costes de redacción del Trabajo Fin de Grado.
- C-4. Derechos de visado del COITT.
- C-5. Gastos de tramitación y envío.

### C-1. Trabajo tarifado por tiempo empleado

A partir del 2008 la Comisión Nacional de los Mercados y la Competencia (CNMC) estableció la liberalización del mercado y por tanto está prohibido el establecimiento de honorarios profesionales por parte de los Colegios Profesionales, no pudiéndose establecer precios de los trabajos profesionales.

La información obtenida, a través del COITT de Canarias, nos indica que el rango en el que se mueve lo que debe de cobrar un ingeniero por su trabajo está entre los 50 y 75 €/hora en función del tipo de proyecto; por lo cual para mi TFG he optado por un precio de 50 €/hora.

Para la realización del presente TFG, se ha empleado 300 horas de trabajo efectivo (3 horas x 5 días x 4 semanas x 5 meses), por lo tanto, los honorarios asociados al total de horas trabajadas en este TFG han sido:

$$H = N_h * C_h = 300 * 50 = 15.000,00 \text{ €}$$

Los honorarios totales por tiempo dedicado libres de impuestos ascienden a *quinze mil euros* (15.000,00 €).

## C-2. Amortización del inmovilizado material

Entre los recursos materiales utilizados para la realización de este TFG se incluyen las herramientas software de desarrollo de los algoritmos del sistema, los paquetes software usados para la redacción de la memoria, y el sistema operativo bajo el que se ejecutó el trabajo. Asimismo, se incluyen los equipos hardware usados para dar soporte a estas herramientas.

Se estipula el coste de amortización para un período de 3 años. Para ello, se utilizará un sistema de amortización lineal o constante, en el que se supone que el inmovilizado material se deprecia de forma constante a lo largo de su vida útil. La cuota de amortización anual se calcula usando la siguiente fórmula:

$$\text{Cuota de amortización anual} = \frac{\text{Valor de adquisición} - \text{Valor residual}}{\text{Número de años de vida útil}}$$

El valor residual es el valor teórico que se supone que tendrá el elemento después de su vida útil.

### Amortización del material hardware

Puesto que la elaboración del TFG ha precisado de 5 meses de trabajo y el cálculo del coste de amortización se estipula en un periodo de tres años, los costes serán calculados como los derivados del tiempo de utilización que se ha requerido por cada uno de los elementos hardware.

En la siguiente tabla se muestra la relación de cada uno de los elementos hardware con su valor de adquisición, su valor residual y el coste de amortización finalmente obtenido.

<b>Elementos Hardware</b>	<b>Coste</b>	<b>Valor residual (3 años)</b>	<b>Amortización (5 meses)</b>
Ordenador Quad-Core Intel Core i5 a 1,4GHz (4 núcleos)	1.377 €	300 €	149,58 €
Ordenador Intel® Core™ i7 a 3 GHz, 16 GB de memoria RAM	1.299 €	350 €	131,81 €
<b>Total, Hardware</b>			<b>281,39 €</b>

Por tanto, el coste total del hardware empleado ha sido de *doscientos ochenta y un euro con treinta y nueve céntimos* (281,39 €).

### Amortización del material software

Para el cálculo del coste de amortización del software se manejan las mismas premisas que en el caso anterior. En la tabla se presentan los costes de amortización del material software:

Software	Número de elementos	Coste	Valor residual (3 años)	Amortización (5 meses)
MatLab® versión 2019b, 64 bits Educativa	1	250 €/año	0 €	104,17 €
Sistema Operativo macOS Mojave ver. 10.14.6	1	0,00 €	0 €	0,00 €
Sistema Operativo macOS Catalina ver. 10.15.4	1	0,00 €	0 €	0,00 €
Paquete Microsoft Office para Mac versión 16.39	2	7,00 €/mes	0 €	35,00 €
Bases de datos de AEMET y RESCAN	---	0,00 €	0 €	0,00 €
<b>Total, Software</b>				<b>139,17 €</b>

Por tanto, el coste total del software empleado ha sido de *ciento treinta y nueve euros con diecisiete céntimos* (139,17 €).

### C-3. Costes de redacción del Trabajo Fin de Grado

El coste asociado a la redacción del TFG se ha calculado según la siguiente expresión recomendada por el COITT:

$$R = 0,07 \cdot P \cdot C_h$$

Donde  $P$  es el presupuesto del TFG obtenido y  $C_h$  un coeficiente de ponderación en función del presupuesto estipulado por el COITT, que varía según

el valor de  $P$ . Dicho valor se ha calculado sumando los costes de las secciones anteriores correspondientes a la amortización del inmovilizado material, tanto hardware como software, al trabajo tarifado por tiempo empleado y los gastos por desplazamiento. Así, el valor de  $P$  asociado a este TFG ha sido:

Concepto	Coste
Trabajo tarifado por tiempo empleado	15.000,00 €
Amortización del material hardware	281,39 €
Amortización del material software	139,17 €
<b>Total</b>	<b>15.420,56 €</b>

Que según el COITT tiene un coeficiente de ponderación de 1.00, pues es el valor fijado para proyectos cuyo presupuesto es inferior a 30.050 €, que es nuestro caso. Teniendo en cuenta el presupuesto calculado en la tabla anterior, se tiene que:

$$R = 0,07 \cdot P \cdot Ch = 0,07 \cdot 15.420,56 \cdot 1 = 1.079,44 \text{ €}$$

Por lo tanto, el importe final de redacción del TFG asciende a la cantidad de *mil setenta y nueve euros con cuarenta y cuatro céntimos (1.079,44 €)*.

#### C-4. Derechos de visado

En cuanto al visado, hoy en día no es obligatorio, y sus gastos depende mucho del tipo de proyecto, por tanto, me he ceñido a un formulario genérico utilizado por el COITT, que establece que para la redacción de proyectos y trabajos en general, los derechos de visado se calculan de acuerdo con la siguiente expresión:

$$V = 0,006 \cdot P \cdot C$$

Donde  $P$  representa el presupuesto total y  $C$  es el coeficiente reductor en función de dicho presupuesto. El presupuesto total se obtiene de la suma de las secciones anteriores correspondientes al trabajo tarifado por tiempo empleado, la amortización del inmovilizado material y la redacción del TFG, como se observa en la tabla:

Concepto	Coste
Trabajo tarifado por tiempo empleado	15.000,00 €
Amortización del material hardware	281,39 €
Amortización del material software	139,17 €
Redacción del TFG	1.079,44 €
<b>Total</b>	<b>16.500,00 €</b>

En función del presupuesto obtenido, se extrae el valor del coeficiente reductor del presupuesto C, según el COITT para presupuestos menores de 30.050 € viene definido con un valor de 1,00. Por consiguiente:

$$V = 0,006 \cdot P \cdot C = 0,006 \cdot 16.500,00 \cdot 1,00 = 99,00 \text{ €}$$

Finalmente, los costes por derecho de visado del TFG ascienden a *noventa y nueve euros (99,00 €)*.

#### **C-5. Gastos de tramitación y envío**

En cuanto a los gastos de tramitación y envío son fijos y se estipulan hoy en día, por parte del COITT en nueve euros (9,00 €).

## Resumen del presupuesto

### Resumen de presupuesto

C-1. Trabajo tarifado por tiempo empleado.....	15.000,00 €
C-2. Amortización del inmovilizado material.....	420,56 €
- Material de hardware: 281,39 €	
- Material de software: 139,17 €	
C-3. Costes de redacción del Trabajo Fin de Grado...	1.079,44 €
C-4. Derechos de visado del COITT.....	99,00 €
C-5. Gastos de tramitación y envío.....	9.00 €
	<hr/>
Presupuesto total antes de impuestos:	<b>16.608,00 €</b>
Aplicación de Impuestos (7% IGIC):	1.162,56 €
	<hr/>
<b>PRESUPUESTO TOTAL:</b>	<b>17.770,56 €</b>
	<hr/>

El presupuesto total, con los impuestos incluidos, asciende a la cantidad de *diecisiete mil setecientos setenta euros con cincuenta y seis céntimos (17.770,56 €)*.

Las Palmas de Gran Canaria, a 23 de julio de 2020

Fdo.: Sonia Fuente González