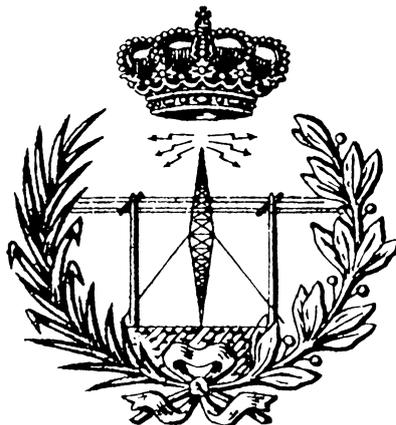


ESCUELA DE INGENIERÍA DE TELECOMUNICACIÓN Y ELECTRÓNICA



TRABAJO DE FIN DE GRADO

DISEÑO DE UN SISTEMA DE OBSERVACIÓN METEOROLÓGICA COMBINANDO ESTACIONES METEOROLÓGICAS Y RADIOENLACES SATELITALES

Titulación: Grado en Ingeniería en Tecnologías de la Telecomunicación

Mención: Sistemas Electrónicos

Autora: Elena Encinas Pérez

Tutores: Dr. Juan Luis Navarro Mesa
José Antonio Godoy Rosario

Fecha: Diciembre de 2020

ESCUELA DE INGENIERÍA DE TELECOMUNICACIÓN Y ELECTRÓNICA



TRABAJO DE FIN DE GRADO

DISEÑO DE UN SISTEMA DE OBSERVACIÓN METEOROLÓGICA COMBINANDO ESTACIONES METEOROLÓGICAS Y RADIOENLACES SATELITALES HOJA DE FIRMAS

Alumna

Fdo.: Elena Encinas Pérez

Tutor

Fdo.: Dr. Juan Luis Navarro Mesa

Tutor

Fdo.: José Antonio Godoy Rosario

Fecha: Diciembre de 2020

ESCUELA DE INGENIERÍA DE TELECOMUNICACIÓN Y ELECTRÓNICA



TRABAJO DE FIN DE GRADO

DISEÑO DE UN SISTEMA DE OBSERVACIÓN METEOROLÓGICA COMBINANDO ESTACIONES METEOROLÓGICAS Y RADIOENLACES SATELITALES

HOJA DE EVALUACIÓN

Calificación: _____

Presidente

Vocal

Secretario/a

Fdo.:

Fdo.:

Fdo.:

Fecha: Diciembre de 2020

AGRADECIMIENTOS

A mis tutores,

Por haberme ofrecido la posibilidad de adentrarme en un proyecto como este, permitiéndome desarrollar aún más los conceptos aprendidos a lo largo de los cuatro años de grado. Por dedicarme todo el tiempo que requería y más, por los fines de semana, e incluso vacaciones, cuando seguíamos en contacto para seguir avanzando, revisando hasta los más pequeños detalles y resolviendo todas las dudas, incluidas las más absurdas.

A mis compañeros y amigos,

Los cuales me han apoyado en todo momento y ayudándome a mantener la ilusión, las ganas de seguir aprendiendo, a ver que las cosas no eran tan difíciles como parecía en algunas ocasiones, a querer siempre investigar un poco más allá y dar lo mejor de mí.

Por ayudarme a mantener presente mi meta en todo momento. Por todas esas noches en las que nadie dormía, donde el protagonista siempre era el café.

A mi familia,

Mi gran apoyo moral. Aunque sin poder estar físicamente, han sido uno de mis pilares fundamentales a lo largo de este proyecto, así como durante toda la trayectoria en estos cuatro años. En especial, a mi padre, mi madre y mi hermano, porque sé que están orgullosos de mi.

El conjunto de todos ellos es lo que me ha llevado a conseguir mis metas y presentar con tanta ilusión este Trabajo de Fin de Grado.

RESUMEN

En este Trabajo de Fin de Grado se ha realizado el diseño de un sistema de observación meteorológica combinando estaciones meteorológicas y radioenlaces satelitales, cuya implementación de algoritmos se lleva a cabo mediante el uso de *Machine Learning* para la detección de precipitación de lluvia en la zona 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.

El método finalmente elegido para abordar este objetivo ha sido redes neuronales de tipo LSTM (*Long Short Term Memory*). Asimismo, se ha llevado a cabo un balanceo de las clases pues nos encontramos ante un problema de clases desbalanceadas y un post-procesado en el que se han calibrado los resultados. Todo esto se enmarca en un proceso de aprendizaje máquina en el que se han utilizado diferentes porcentajes de las muestras (datos) para entrenar y testear. Para estudiar la validez de nuestra metodología con respecto a un mes seleccionado, se realiza el entrenamiento de la red con varios de los meses anteriores, cuyo resultado se emplea para testear el mes elegido inicialmente.

Como resultado se muestran unas métricas de calidad basadas en la sensibilidad, exactitud, especificidad y F1-Score. Estos valores se muestran en múltiples repeticiones y de sus valores medios y desviaciones vemos que el método desarrollado es estable y fiable.

ABSTRACT

In this Final Degree Project, it has been developed the design of an observing system that combines meteorological stations and satellite radiolink, where the implementation of the algorithm is done by using Machine Learning algorithms for the detection and quantification of rain precipitation. 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.

The method that has been used to achieve this objective is neural networks. 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 for a specific month, the network train is developed using the two previous months. The result is used for testing the month previously selected.

After all the work, it can be concluded that the model generates values and deviations of sensitivity (recall), accuracy, specificity and F1-Score valid and very accurate.

ÍNDICE GENERAL

PARTE I

CAPÍTULO 1. INTRODUCCIÓN	1
1.1 OBSERVACIÓN METEOROLÓGICA	1
1.1.1 Estación meteorológica	3
1.2 ALTERNATIVAS BASADAS EN RADIOENLACES	5
1.3 OBJETIVOS	8
1.4 ORGANIZACIÓN DE LA MEMORIA	10
CAPÍTULO 2. FUNDAMENTOS TEÓRICOS	12
2.1 PRECIPITACIÓN	12
2.2 CARACTERÍSTICAS GENERALES DE LOS RADIOENLACES	15
2.2.1 Definición y tipos de radioenlaces	15
2.2.2 Parámetros de un radioenlace	16
2.2.3 Ventajas e inconvenientes de los radioenlaces	17
2.3 ENLACES SATELITALES	17
2.3.1 Parámetros de los radioenlaces satelitales	20
2.3.2 Potencia radiada	21
2.3.3 Satélite HISPASAT	24
2.4 MODELOS DE PROPAGACIÓN	29
2.5 FACTORES ATMOSFÉRICOS QUE INFLUYEN EN LA DEGRADACIÓN DE LA CALIDAD DE UN ENLACE SATELITAL	30
2.5.1 Atenuación por absorción de gases	31
2.5.2 Atenuación por nubes	32
2.5.3 Atenuación por lluvia	33
2.5.4 Atenuación en la capa fundente	39
2.5.5 Desvanecimiento por ángulo bajo	39
2.5.6 Atenuación por centelleo	40
2.5.7 Efecto de la antena mojada	41
CAPÍTULO 3. PREPARACIÓN DEL ENTORNO DE TRABAJO	45

3.1	FUENTES DE DATOS EMPLEADAS	45
3.1.1	<i>Prototipo satelital VIMETRI-MAC</i>	45
3.1.2	<i>Estación meteorológica en el pabellón B de la EITE</i>	52
3.1.2	<i>Estación meteorológica de Tafira – El Zurbarán</i>	53
3.2	ADQUISICIÓN DE DATOS	54
3.2.1	<i>Adquisición de datos de los radioenlaces</i>	54
3.2.2	<i>Adquisición de los datos de estaciones</i>	55
3.2.3	<i>Región de trabajo</i>	56
3.3	CREACIÓN DE LA BASE DE DATOS	57
3.3.1	<i>Tipos de datos</i>	57
CAPÍTULO 4. PRE-PROCESADO Y PROCESADO DE LOS DATOS		60
4.1	TRATAMIENTO DE LOS DATOS METEOROLÓGICOS DE VIMETRI	60
4.1.1	<i>Tipos de GAPS</i>	61
4.1.2	<i>Posibles causas de aparición de GAPS</i>	61
4.2	TRATAMIENTO DE LOS DATOS DE RADIOENLACES DE VIMETRI	62
4.2.1	<i>Lectura de la base de datos</i>	63
4.2.2	<i>Localización de GAPS</i>	63
4.2.3	<i>Transformación del periodo de muestreo de la señal</i>	64
4.2.4	<i>Eliminación de bajas frecuencias de la señal</i>	65
4.3	SINCRONISMO ENTRE SEÑALES DE RADIOENLACES Y DE ESTACIONES METEOROLÓGICAS	67
4.4	PROCESADO DE LOS DATOS METEOROLÓGICOS Y DE RADIOENLACES	68
CAPÍTULO 5.		70
ASPECTOS TEÓRICOS SOBRE ALGORITMOS DE DETECCIÓN DE PRECIPITACIÓN		70
5.1	INTRODUCCIÓN	70
5.1.1	<i>Breve historia de la inteligencia artificial</i>	72
5.2	INTRODUCCIÓN AL APRENDIZAJE MÁQUINA	74
5.2.1	<i>Formas de aprendizaje</i>	76
5.2.2	<i>Bases de datos balanceadas y estrategias de balanceo</i>	81
5.3	REDES NEURONALES	83
5.3.1	<i>Perceptrón multicapa</i>	83
5.3.2	<i>Propagación hacia atrás de errores o backpropagation</i>	86
5.3.3	<i>Redes Neuronales Recurrentes</i>	86
5.3.4	<i>Long-Short Term Memory</i>	88

5.3.5	<i>Cómo entrenar una red neuronal</i>	93
CAPÍTULO 6. ALGORITMO DE DETECCIÓN DE PRECIPITACIÓN COMBINANDO ESTACIONES METEOROLÓGICAS Y RADIOENLACES SATELITALES		95
6.1	DIAGRAMA DE FLUJO	95
6.2	LECTURA DE DATOS Y SINCRONISMO DE LOS MISMOS	97
6.3	DETECCIÓN DE GAPS	98
6.4	DETECTOR ÓPTIMO DE UMBRAL	100
6.5	DETECTOR DE UMBRAL Y MÁRGENES DE CONFIANZA	104
6.6	PROCESO DE SELECCIÓN DE CARACTERÍSTICAS DE ENTRENAMIENTO Y TEST	106
6.7	RED LSTM	107
CAPÍTULO 7. EXPERIMENTOS Y RESULTADOS DE ESTIMACIÓN DE LLUVIA		111
7.1	PARÁMETROS DE CALIDAD	111
7.1.1	<i>Basados en la matriz de confusión</i>	111
7.1.2	<i>Basados en la media y desviación sobre múltiples experimentos</i>	114
7.2	EXPERIMENTOS PARA LA ESTIMACIÓN DE UMBRAL ÓPTIMO	114
7.3	EXPERIMENTOS DE DETECCIÓN MES A MES	119
7.4	PRINCIPALES RESULTADOS	121
CAPÍTULO 8. CONCLUSIONES		128
8.1	CONCLUSIONES	128
8.2	LÍNEAS FUTURAS	129
BIBLIOGRAFÍA		131

PARTE II

PLIEGO DE CONDICIONES	1
CONDICIONES LEGALES Y ADMINISTRATIVAS	1
DERECHOS DE EXPLOTACIÓN Y RESTRICCIONES DE USO	1
GARANTÍAS Y CONSIDERACIONES ADICIONALES	2
HERRAMIENTAS UTILIZADAS	3

PARTE III

PRESUPUESTO	1
DECLARACIÓN JURADA	1
DESGLOSE DEL PRESUPUESTO	2
RESUMEN DEL PRESUPUESTO	8

ÍNDICE DE FIGURAS

FIGURA 1. SISTEMA DE OBSERVACIÓN METEOROLÓGICA. _____	4
FIGURA 2. SOCIEDAD MERCANTIL HISPASAT Y SUS SOCIEDADES DEPENDIENTES COMPONEN HISPASAT. _____	25
FIGURA 3. MAPA DE COBERTURA DE HISPASAT. _____	26
FIGURA 4. SATÉLITES DE LA FLOTA DE HISPASAT. _____	27
FIGURA 5. SATÉLITE HISPASAT 30W-6. _____	28
FIGURA 6. REGIONES CLIMÁTICAS – NASA. _____	34
FIGURA 7. PARÁMETROS DE ENTRADA PARA EL CÁLCULO DE ATENUACIÓN POR LLUVIA. _____	36
FIGURA 8. ATENUACIÓN DE LA ANTENA MOJADA $A\alpha$ EN FUNCIÓN DE LA FRECUENCIA DE LA SEÑAL PARA DIFERENTES INTENSIDADES DE LLUVIA, PARA UNA SOLA ANTENA. _____	42
FIGURA 9. ATENUACIÓN DE ANTENA MOJADA $A\alpha$ EN FUNCIÓN DE LA INTENSIDAD DE LLUVIA EN UNA SOLA ANTENA. SE MUESTRA UNA COMPARACIÓN ENTRE EL MÉTODO ACTUAL (PELÍCULA PLANA) Y UN MÉTODO ANTERIOR (EMPÍRICO) A 27 GHz, Y LAS RELACIONES ENTRE R Y $A\alpha$ PARA 10, 20, 50 Y 70. _____	43
FIGURA 10. RELACIÓN ENTRE LA ATENUACIÓN INDUCIDA POR LA LLUVIA Y LA ATENUACIÓN DE LA ANTENA MOJADA EN FUNCIÓN DE LA FRECUENCIA DE LA SEÑAL Y LA INTENSIDAD DE LA LLUVIA PARA ENLACES DE 1 KM. _____	44
FIGURA 11. BLOQUEO DE CONTINUA PARA LA MEDICIÓN DE POTENCIA RECIBIDA. _____	46
FIGURA 12. DIAGRAMA DE BLOQUES GENERAL DEL PROTOTIPO. _____	47
FIGURA 13. ANALIZADOR DE ESPECTROS PORTÁTIL (LABORATORIO 220, PABELLÓN B -EITE). _____	48
FIGURA 14. CONEXIONES DE RED PARA EL PROTOTIPO. _____	50
FIGURA 15. DIAGRAMA DE FLUJO DEL ALGORITMO DE OBTENCIÓN DE PARÁMETROS METEOROLÓGICOS. _____	51
FIGURA 16. VAISALA RAIN GAUGE RG13. _____	53
FIGURA 17. NIVELES DE POTENCIA DE LA SEÑAL SATELITAL. VIMETRIMAC 2019. _____	55
FIGURA 18. DATOS DE PRECIPITACIÓN MEDIDOS CON PLUVIÓMETRO. VIMETRIMAC 2020. _____	56
FIGURA 19. ARCHIVO DE LA BASE DE DATOS DE LA ESTACIÓN METEOROLÓGICA CORRESPONDIENTE AL MES DE NOVIEMBRE DE 2020. _____	58
FIGURA 20. ARCHIVO DE LA BASE DE DATOS RSL CORRESPONDIENTE AL MES DE NOVIEMBRE DE 2020. _____	59
FIGURA 21. ETAPAS DEL PRE-PROCESADO DE LOS DATOS DE RADIOENLACES. _____	62
FIGURA 22. COMPARATIVA DE SEÑAL LA SEÑAL ORIGINAL CON LA SEÑAL DE REFERENCIA Y LA SEÑAL UMBRAL. _____	66

FIGURA 23. SEÑAL RSL A LA QUE SE LE HA APLICADO UNA APERTURA Y UN CIERRE MORFOLÓGICO. _____	67
FIGURA 24. RSL SATÉLITE: 25 A 30 DE NOVIEMBRE DE 2020. _____	68
FIGURA 25. DATOS DE PRECIPITACIÓN MEDIDOS POR EL PLUVIÓMETRO: 25 A 30 DE NOVIEMBRE DE 2020. _____	69
FIGURA 26. ESQUEMA DE LOS DISTINTOS ENFOQUES DE APRENDIZAJE AUTOMÁTICO. _____	78
FIGURA 27. ESQUEMA DEL FUNCIONAMIENTO DEL ALGORITMO SELF-LEARNING. _____	79
FIGURA 28. AJUSTE ITERATIVO DE LA FRONTERA DE DECISIÓN EN EL APRENDIZAJE SEMI-SUPERVISADO. _____	80
FIGURA 29. BALANCEO DE DATOS. _____	82
FIGURA 30. ESQUEMA BÁSICO DE UNA NEURONA ARTIFICIAL CON CUATRO CARACTERÍSTICAS DE ENTRADA. _____	84
FIGURA 31. REPRESENTACIÓN GRÁFICA DE LA ESTRUCTURA DE UN PERCEPTRÓN MULTICAPA DE DOS NIVELES. _____	85
FIGURA 32. RNN DE UNA NEURONA. _____	87
FIGURA 33. RNN DESPLEGADA EN EL EJE DEL TIEMPO. _____	87
FIGURA 34. DIAGRAMA DE UNA LSTM. _____	89
FIGURA 35. DIAGRAMA DE UNA LSTM – CELDA. _____	90
FIGURA 36. DIAGRAMA DE UNA LSTM - CAPA SIGMOIDEA. _____	91
FIGURA 37. FUNCIONAMIENTO DE UNA LSTM - PASO 1. _____	92
FIGURA 38. FUNCIONAMIENTO DE UNA LSTM - PASO 2. _____	92
FIGURA 39. FUNCIONAMIENTO DE UNA LSTM - PASO 3. _____	93
FIGURA 40. DIAGRAMA DE FLUJO DEL ALGORITMO DE DETECCIÓN. _____	96
FIGURA 41. GAP RSL - ABRIL DE 2020. _____	99
FIGURA 42. SEÑAL RSL A LA QUE SE LE HA QUITADO LA COMPONENTE DE BAJA APLICADO UNA APERTURA Y UN CIERRE MORFOLÓGICO. _____	100
FIGURA 43. APLICACIÓN DE UMBRALES A LA SEÑAL RSL. _____	101
FIGURA 44. APLICACIÓN DE LOS UMBRALES A SEÑAL RSL – ZOOM. _____	101
FIGURA 45. BUCLE PARA EL CÁLCULO DEL UMBRAL ÓPTIMO. _____	102
FIGURA 46. CURVA ROC. _____	103
FIGURA 47. SENSIBILIDAD VS ESPECIFICIDAD. _____	103
FIGURA 48. APLICACIÓN DEL UMBRAL ÓPTIMO A LA SEÑAL RSL. _____	104
FIGURA 49. EVENTO DE LLUVIA REPRESENTADO CON LAS MARCAS DE PRINCIPIO-FIN. _____	105
FIGURA 50. CARACTERÍSTICAS DE LOS EVENTOS DE LLUVIA Y NO LLUVIA. _____	106
FIGURA 51. CAPAS Y OPCIONES PARA EL ENTRENAMIENTO DE LA RED. _____	107
FIGURA 52. CAPAS DE ENTRENAMIENTO DE LA RNN. _____	109
FIGURA 53. OPCIONES DE ENTRENAMIENTO DE LA RNN. _____	110

FIGURA 54. DIAGRAMA DE FLUJO SEGUIDO PARA EL CÁLCULO DEL COEFICIENTE K_{OPT} .	115
FIGURA 55. CURVA ROC - 02/2020.	116
FIGURA 56. ENTRENAMIENTO PARA EL CÁLCULO DEL K ÓPTIMO.	117
FIGURA 57. ARCHIVOS DE K_{OPT} POR MESES.	118
FIGURA 58. LLUVIA REAL VS. LLUVIA DETECTADA - OCTUBRE DE 2020.	119
FIGURA 59. PARÁMETROS DE CALIDAD OBTENIDOS PARA CADA UNO DE LOS MESES DE ESTUDIO UTILIZANDO UN 50% DE LOS DATOS.	123
FIGURA 60. PARÁMETROS DE CALIDAD OBTENIDOS PARA CADA UNO DE LOS MESES DE ESTUDIO UTILIZANDO UN 60% DE LOS DATOS.	123
FIGURA 61. PARÁMETROS DE CALIDAD OBTENIDOS PARA CADA UNO DE LOS MESES DE ESTUDIO UTILIZANDO UN 70% DE LOS DATOS.	124
FIGURA 62. PARÁMETROS DE CALIDAD OBTENIDOS PARA CADA UNO DE LOS MESES DE ESTUDIO UTILIZANDO UN 80% DE LOS DATOS.	125
FIGURA 63. PARÁMETROS DE CALIDAD OBTENIDOS PARA EL MES DE SEPTIEMBRE DE 2020 UTILIZANDO UN 90% DE LOS DATOS.	126
FIGURA 64. PARÁMETROS DE CALIDAD MEDIOS ATENDIENDO AL PORCENTAJE DE TRAIN-TEST UTILIZADO.	126

ÍNDICE DE TABLAS

TABLA 1. VARIABLES METEOROLÓGICAS CAPTURADAS EN LAS EMA. _____	5
TABLA 2. CRITERIOS PARA DETERMINAR LA INTENSIDAD (I) DE PRECIPITACIÓN DÉBIL, MODERADA O FUERTE. _____	13
TABLA 3. PARÁMETROS DE UN RADIOENLACE. _____	17
TABLA 4. PARÁMETROS DE UN RADIOENLACE SATELITAL. _____	20
TABLA 5. CARACTERÍSTICAS TÉCNICAS DEL SATÉLITE HISPASAT 30W-6. _____	29
TABLA 6. TIPOS DE NUBES. _____	33
TABLA 7. TASAS DE LLUVIA. _____	35
TABLA 8. EJEMPLOS DISPONIBLES. _____	75
TABLA 9. DEFINICIÓN GRÁFICA DE TRUE POSITIVE, TRUE NEGATIVE, FALSE POSITIVE Y FALSE NEGATIVE. _____	112
TABLA 10. MATRIZ DE CONFUSIÓN DEL MODELO. _____	112
TABLA 11. DATOS DE K _{OPT} DE LOS MESES DE ESTUDIO. _____	118
TABLA 12. VALIDATION ACCURACY. _____	121

ANEXO

TABLA 13. PARÁMETROS DE CALIDAD CON 50 % DE DATOS PARA EL ENTRENAMIENTO. _____	I
TABLA 14. PARÁMETROS DE CALIDAD CON 60 % DE DATOS PARA EL ENTRENAMIENTO. _____	IV
TABLA 15. PARÁMETROS DE CALIDAD CON 70 % DE DATOS PARA EL ENTRENAMIENTO. _____	VII
TABLA 16. PARÁMETROS DE CALIDAD CON 80 % DE DATOS PARA EL ENTRENAMIENTO. _____	X

ACRÓNIMOS

AEMET	<i>Agencia Estatal de Meteorología</i>
BBS	<i>Broad-Band Service</i>
CDTI	<i>Centro para el Desarrollo Tecnológico e Industrial</i>
CPU	<i>Central Processing Unit</i>
CSV	<i>Comma-Separated Values</i>
EITE	<i>Escuela de Ingeniería de Telecomunicación y Electrónica</i>
EMA	<i>Estación meteorológica Automática</i>
FDM	<i>Multiplexación por División de Frecuencia (Frequency Division Multiplexing)</i>
FI	<i>Frecuencia Intermedia</i>
FM	<i>Modulación en Frecuencia (Frequency Modulation)</i>
FN	<i>False Negative (Falso Negativo)</i>
FP	<i>False Positive (Falso Positivo)</i>
FSK	<i>Modulación por desplazamiento de frecuencia (Frequency Shift Keying)</i>
FSPL	<i>Free-Space Path Loss</i>
FTP	<i>File Transfer Protocol</i>
GISM	<i>Global Ionospheric Scintillation Model</i>
GOS	<i>Global Observing System</i>
GPRS	<i>General Packet Radio Service</i>
IA	<i>Inteligencia Artificial</i>
IBM	<i>International Business Machines Corporation</i>
IDeTIC	<i>Instituto para el Desarrollo Tecnológico y la Innovación en Comunicaciones</i>
ITU	<i>International Telecommunication Union</i>
JOHNIAC	<i>John v. Neumann Numerical Integrator and Automatic Computer</i>
LISP	<i>LISt Processor (Procesamiento de listas)</i>
LNB	<i>Low Noise Block</i>

LSTM	<i>Long-Short Term Memory</i>
MAC	<i>Madeira-Azores-Canarias: Programa Operativo de Cooperación Territorial</i>
MIT	<i>Massachusetts Institute of Technology</i>
ML	<i>Machine Learning</i>
NLP	<i>Natural Language Processing</i>
PIRE	<i>Potencia Isotrópica Radiada Equivalente</i>
PSK	<i>Modulación por desplazamiento de fase (Phase Shift Keying)</i>
QAM	<i>Modulación de amplitud en cuadratura (Quadrature Amplitude Modulation)</i>
QPSK	<i>Modulación por desplazamiento de fase en cuadratura de polarización dual (Quadrature Phase Shift Keying)</i>
RDSI	<i>Red Digital de Servicios Integrados</i>
REC	<i>Red Eléctrica Corporación</i>
RF	<i>Radio Frequency</i>
RNA	<i>Red Neuronal Artificial</i>
RNN	<i>Redes Neuronales Recurrentes</i>
ROC	<i>Receiver Operating Characteristic</i>
RSL	<i>Radio Signal Level</i>
SEPI	<i>Sociedad Estatal de Participaciones Industriales</i>
SSL	<i>Space Service/Loral</i>
SVM	<i>Support Vector Machines</i>
TFG	<i>Trabajo de Fin de Grado</i>
TN	<i>True Negative (Verdadero Negativo)</i>
TP	<i>True Positive (Verdadero Positivo)</i>
UIT	<i>Unión Internacional de Telecomunicaciones</i>
UPM	<i>Universidad Politécnica de Madrid</i>
USB	<i>Universal Serial Bus</i>
VIMETRIMAC	<i>Sistema de vigilancia meteorológica para el Seguimiento de Riesgos Medioambientales</i>
WMO	<i>World Meteorological Organization</i>
WWW	<i>World Weather Watch</i>
XPD	<i>Cross Polarization Discrimination</i>

PARTE I – MEMORIA

CAPÍTULO 1.

INTRODUCCIÓN

Este capítulo contiene una breve introducción en la que se tratará de abordar algunos de los conceptos básicos e importantes relacionados con este Trabajo de Fin de Grado. En primer lugar, se explicará en qué consiste la observación meteorológica. Seguidamente, se expondrán algunas alternativas a los sistemas basados en observación meteorológica, donde los protagonistas serán los radioenlaces.

1.1 Observación meteorológica

La observación meteorológica consiste en medir y determinar variables atmosféricas que en su conjunto conforman las condiciones de la atmósfera en un lugar y un momento determinados, empleando para ello el instrumental adecuado. Desde de la aparición de *World Weather Watch* (WWW) en 1963, el principal mecanismo encargado de proporcionar estos datos en todo el mundo ha sido el Sistema Mundial de Observación (GOS – *Global Observing System*). La continua evolución de las tecnologías ha permitido la aparición de nuevos sistemas de medida y predicción de mayor resolución y exactitud, pudiéndose realizar, en cualquier momento, evaluaciones sobre observaciones provenientes de cualquier fuente. Este desarrollo también se ha visto impulsado debido a la creciente demanda y dependencia de la información meteorológica.

El Sistema Mundial de Observación está formado por una gran red de sistemas de observación distribuidos entre superficie y espacio que son operados por una gran cantidad de agencias nacionales e internacionales.

La Organización Mundial de Meteorología (WMO -*World Meteorological Organization*) es la encargada de coordinar y guiar el Sistema Mundial de Observación y el Sistema Mundial de Telecomunicaciones con el fin de intercambiar los datos obtenidos de la red de sistemas de observación entre los miembros de la WMO, permitiendo complementar unos datos con otros.

El GOS tiene dos objetivos principales que cumplir:

1. La mejora y optimización de los sistemas meteorológicos globales con el fin de obtener predicciones más efectivas y eficientes, análisis meteorológicos, pronósticos y advertencias más precisos, además, para actividades de monitoreo del clima y el medioambiente y servicios de alerta entre otros.
2. Establecer un estándar para las técnicas y prácticas de observación, así como la planificación de redes a nivel regional.

Actualmente se dispone de 11.000 estaciones terrenales destinadas a las observaciones en la Tierra o en torno a su superficie. Estas estaciones transfieren datos cada 1 a 3 horas, entre los que podemos encontrar presión atmosférica, temperatura, humedad relativa y velocidad y dirección del viento. Estos parámetros meteorológicos son intercambiados globalmente y en tiempo real.

A parte de las terrenales, existe una gran variedad de plataformas, entre las que encontramos las satelitales, marinas, aéreas y radares.

La red de satélites de observación ambiental está compuesta por satélites de órbita polar y satélites geoestacionarios, además de satélites de investigación y desarrollo. Proporcionan datos precisos y actualizados constantemente a lo largo de amplias áreas geográficas haciendo uso de imágenes tomadas por estos satélites.

Aproximadamente 1.300 estaciones componen el sistema de observación aéreo, compuesto por radiosondas y globos, que realizan mediciones de presión, velocidad del viento, humedad y temperatura en altitudes no superiores a 30km

El sistema de observación marino cuenta con alrededor de 4.000 barcos, cuya información recogida se combina con la obtenida por satélites, boyas y plataformas flotantes. Estas estaciones recogen datos similares a los de las estaciones terrestres, pero incluyendo temperatura de la superficie del mar, altura de las olas y período de las mismas.

Los radares meteorológicos permiten la detección de precipitaciones, así como calcular sus trayectorias y determinar en qué estado se encuentra, pudiendo tratarse de llovizna, lluvia, nieve, granizo, etc. Los radares *Doppler* son ampliamente utilizados en redes nacionales y regionales por su capacidad de pronosticar fenómenos climáticos severos en un corto plazo de tiempo.

Pudiéndose añadir otros sistemas de observación meteorológica, como son las observaciones de radiación solar, detección de rayos y mediciones de mareógrafos, así como los perfiladores de viento, los cuales parecen ser útiles para realizar observaciones entre sondeos transmitidos por globos.

1.1.1 Estación meteorológica

Entre los sistemas de captura de datos meteorológicos, uno de los más relevantes son las estaciones meteorológicas automáticas (EMA). Una EMA se define como una “estación meteorológica en la que se realizan y transmiten observaciones automáticamente”. Estos sistemas disponen de una serie de sensores integrados que se encargan de medir diferentes variables meteorológicas, las cuales son enviadas a una unidad central de adquisición de datos [1].

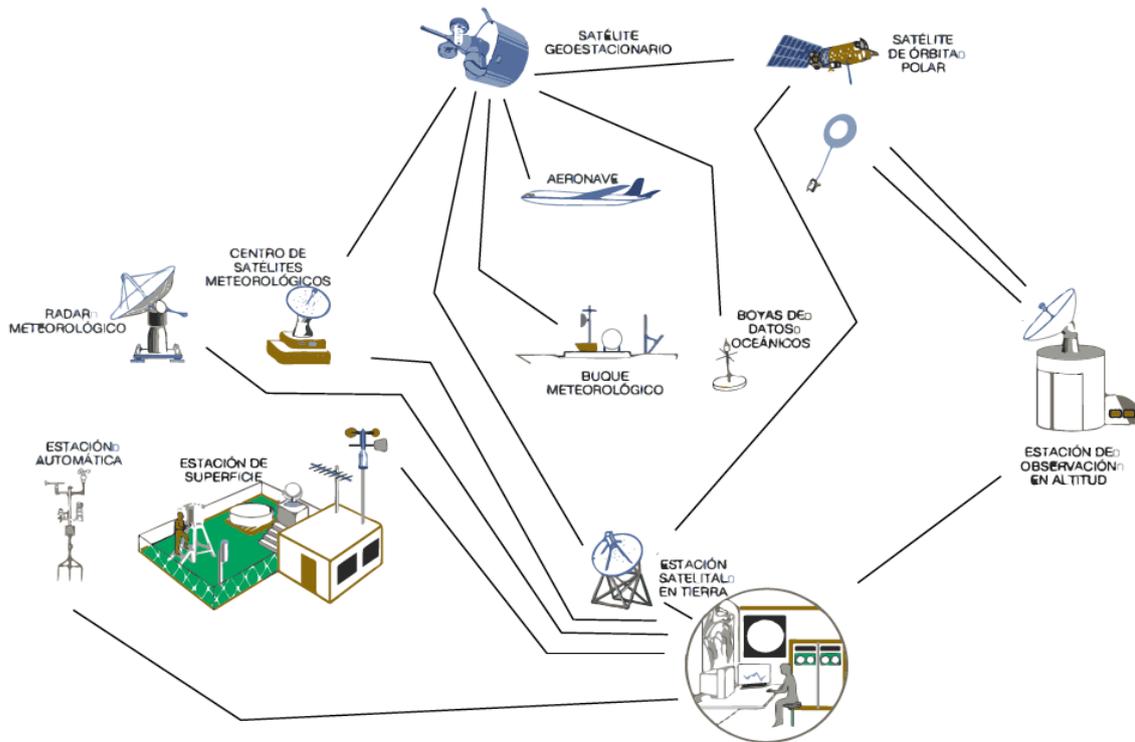


Figura 1. Sistema de Observación Meteorológica.

Nota: Adaptado de "Algoritmos de Aprendizaje Aplicados a la Monitorización Meteorológica mediante Redes Inalámbricas de Comunicaciones Distribuidas", por I. D. Guerra, 2017.

Estos datos capturados por las estaciones meteorológicas automáticas pueden ser procesados directamente en la propia estación o llevarse a cabo en otra parte.

A continuación, se muestran en la Tabla 1 una serie de parámetros que componen las principales variables capturadas por las EMA.

Tabla 1. Variables meteorológicas capturadas en las EMA.

VARIABLE	DESCRIPCIÓN
Dirección del viento	El valor obtenido es el promedio de 10 minutos de la dirección del viento. La dirección indica de donde proviene el viento, su unidad de medición es en grados Dextrorsum (giro en sentido de las manecillas del reloj) donde 0° es norte verdadero
Dirección del viento de ráfaga	La dirección del viento de ráfaga es la dirección de donde proviene la ráfaga más intensa en un lapso de 10 minutos, su unidad de medición es en grados Dextrorsum (giro en sentido de las manecillas del reloj).
Velocidad del viento	La velocidad del viento es el promedio aritmético de las velocidades medidas en un lapso de 10 minutos, su unidad de medición para EMA es en km/h.
Velocidad del viento de ráfaga	La velocidad del viento de ráfaga es la máxima velocidad medida en un intervalo de 10 minutos (se toman muestras cada 5 segundos), su unidad de medición para EMA es en km/h.
Temperatura ambiente promedio	Es la temperatura ambiente promedio de las mediciones realizadas en un lapso de 10 minutos (se toman muestras cada minuto), su unidad de medición es en °C.
Humedad relativa	La humedad relativa es el promedio de las mediciones realizadas en un intervalo de 10 minutos (se toman muestras cada minuto), su unidad de medición es en %.
Presión atmosférica	La presión atmosférica es el promedio de las mediciones realizadas en un lapso de 10 minutos (se toman muestras cada minuto), su unidad de medición es el hpa (Hectopascal).
Precipitación	Es la lámina de precipitación acumulada en un lapso de 10 minutos, su unidad de medición es el mm (milímetro).
Radiación	La radiación solar (global) son los valores promedio medidos en un lapso de 10 minutos (se toman mediciones cada minuto), su unidad de medición es en W/m ² (watt/m ²).

1.2 Alternativas basadas en radioenlaces

Antes de que tuviera lugar la Segunda Guerra Mundial ya se habían realizado las primeras pruebas que incluían radioenlaces. En 1931 el Laboratorio Central de Telecomunicación desarrolló una serie de equipos radioeléctricos que trabajaban a una frecuencia del orden de

1.7GHz. Fue en 1933, dos años más tarde, cuando encontraron su aplicación en un enlace con un solo canal telefónico o telegráfico entre los aeródromos *Lympne* (Inglaterra) y *Saint-Inglevert* (Francia).

En 1947 se desarrollaron los primeros radioenlaces analógicos. Estos empleaban modulaciones en frecuencia (FM - *Frequency Modulation*), un múltiplex analógico de canales analógicos (FDM - *Frequency Division Multiplexing*) o una señal de televisión analógica.

En España, en el año 1956, Telefónica se convirtió en la empresa pionera en instalar un radioenlace de microondas para establecer comunicación entre las ciudades de Madrid y Sevilla.

Pero no fue hasta 1968 cuando aparecieron los primeros radioenlaces de tipo digital. En su inicio, las modulaciones más utilizadas fueron las modulaciones en frecuencia, como la FSK (*Frequency Shift Keying*), o en fase, como QPSK (*Quadrature Phase Shift Keying*). Con el tiempo las redes se fueron digitalizando, consiguiendo de esta manera que las redes tradicionales de telefonía evolucionaran a redes RDSI (*Red Digital de Servicios Integrados*) y las redes de televisión analógica se digitalizaron.

Diferentes técnicas de modulación de mayor nivel y mayor eficiencia espectral, como por ejemplo la QAM (*Quadrature Amplitude Modulation*), comenzaron a utilizarse en la década de los 80. Con el avance en la tecnología, estas redes de radioenlaces tradicionales han sido sustituidas por redes de fibra óptica debido a su amplia capacidad y sus altas prestaciones.

En la última década ha aumentado el interés por aprovechar el efecto de las condiciones meteorológicas sobre los niveles de potencia en las comunicaciones por radio a frecuencias superiores a 1 GHz para superar las limitaciones de los sistemas convencionales.

Mediante el uso de radioenlaces, los datos obtenidos son mucho más exactos que los proporcionados por sistemas meteorológicos. Cuantificar la lluvia en un lugar concreto, medir la velocidad del viento, detectar si hay niebla, detectar calima, la temperatura e incluso detectar conatos de incendio, es posible ya que las mediciones se toman en el área del mismo utilizando las variaciones de la potencia de la señal transmitida.

Los estudios realizados por C. Ho [1] muestran una simulación para la predicción del comportamiento de una señal afectada por distintas condiciones meteorológicas y topográficas. A partir de una serie de pruebas, pudo determinar la capacidad las limitaciones de simulación y modelos subyacentes.

Una investigación de Messer, Zinevich, y Alpert [2], realizada en 2006, estima la cantidad de lluvia mediante el uso de datos obtenidos de una red celular. Posteriormente estos tres mismos investigadores [3], utilizando una modificación de la investigación anterior [2], junto con un modelo tomográfico no lineal diseñado sobre una rejilla de densidad variable, logran la medida de intensidad de lluvia cerca de la superficie (en una superficie de 3.200 Km²) con una desviación menor del 10%.

En trabajos recientes, Alpert, Rayitsfeld et al. [4], comparan dos metodologías utilizadas por otros investigadores. La primera trata de la estimación de la intensidad de lluvia a partir de datos obtenidos de un radioenlace cercano a una estación meteorológica. La segunda, sin embargo, utiliza la interpolación de los datos de varios radioenlaces para la obtención de la intensidad de lluvia en los puntos donde se encuentran las estaciones meteorológicas. Comparando ambas metodologías con los datos obtenidos en estaciones meteorológicas se obtiene una desviación del 6% y del 3%, respectivamente.

Recientemente, Messer, Zinevich y Alpert [5], hicieron uso de una red de comunicaciones inalámbricas, estableciendo una serie de sensores ambientales para la medición de la intensidad de lluvia y la velocidad del viento. Asimismo, David, Alpert y Messer [6] introducen un método nuevo para la monitorización de eventos de niebla mediante el uso de infraestructuras de comunicaciones móviles.

Asimismo, este proyecto se engloba dentro de la Tesis Doctoral de D. Iván Guerra [1], cuyo objetivo ha sido investigar la utilización de algoritmos de aprendizaje para la monitorización meteorológica. A su vez, ambos trabajos, se enmarcan en el proyecto VIMETRI-MAC, cuyo título es "SISTEMA DE VIGILANCIA METEOROLÓGICA PARA EL SEGUIMIENTO DE RIESGOS MEDIOAMBIENTALES", código MAC/3.5b/065, que tiene la Universidad de las Palmas de Gran Canaria como Jefe de Filas, y la Universidad de Madeira y el Laboratório Regional de Engenharia Civil como Socios

(ambos en Madeira). Además, tiene como participantes asociados la Dirección General de Telecomunicaciones y Nuevas Tecnologías (Gobierno de Canarias), Dirección General de Seguridad y Emergencias (Gobierno de Canarias), Agencia Estatal de Meteorología. Fue 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.

1.3 Objetivos

Objetivo principal:

Implementación de algoritmos de *Machine Learning* (ML) para la detección y cuantificación de precipitación basado en los niveles de potencia involucrados en comunicaciones satelitales y estaciones meteorológicas 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:

Objetivo 1: Estudio de las precipitaciones y su influencia sobre la atenuación de señales en radioenlaces satelitales.

Con este objetivo, se pretende estudiar cómo conociendo el par “parámetros-degradación” en un radioenlace satelital podemos obtener información que nos permita establecer una hipótesis acerca de los mecanismos de propagación. De esta manera, nos centramos en las precipitaciones debidas a lluvia. En este sentido, la atmósfera terrestre y el espacio exterior, hasta-desde el satélite, se establece como un sistema oculto para un observador situado en el receptor (p.e, antena parabólica en tierra). Este sistema tiene capacidad de modificar la señal transmitida, que se caracteriza por parámetros conocidos

desde el diseño, y cuya respuesta constituye una señal afectada por el sistema atmósfera-espacio.

Objetivo 2: Bases de datos de los radioenlaces satelitales y las estaciones meteorológicas próximas.

Una vez estudiadas las precipitaciones y su influencia sobre la atenuación, debemos gestionar la base de datos que disponemos, con la finalidad de que sirva como elemento de aprendizaje en etapas posteriores donde se hará uso de técnicas de ML (*Machine Learning*). Así, usaremos dos bases de datos: 1) Niveles de potencia proporcionados por los receptores satelitales VIMETRI 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.

Se crearán las bases para poder integrar de forma conjunta la información de los sistemas satelital y estaciones meteorológicas.

Objetivo 3: Esquema general del sistema integrado satélite-estación.

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.

Objetivo 5: Particularmente, se estudiará técnicas basadas en redes neuronales, máquinas de vectores soportados (*Support Vector Machines, SVM*), etc. Enfatizar también en el estudio de las técnicas más apropiadas para extraer las características más relevantes (corrección del efecto de humedad en las antenas, análisis por ventanas deslizantes, etc.).

1.4 Organización de la memoria

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

Capítulo 1. Introducción.

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.

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. Preparación del entorno de trabajo.

Analizamos las distintas fuentes de datos al igual que la misma adquisición de los datos y el previo tratamiento que se debe hacer para utilizarlos.

Capítulo 4. Pre-procesado y procesado de los datos.

Breve introducción a los gaps, tipos, y cómo afectan a nuestra base de datos. Se explica el proceso seguido para sincronizar los datos de las diferentes fuentes.

Capítulo 5. Aspectos teóricos sobre algoritmos de detección de precipitación.

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 6. Algoritmo de detección de precipitación combinando estaciones meteorológicas y radioenlaces satelitales.

Describimos el flujo de trabajo seguido por el algoritmo desarrollado para la detección de eventos de lluvia.

Capítulo 7. Experimentos y resultados de estimación de lluvia.

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 8. 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, donde se enumeran las referencias bibliográficas empleadas en este TFG.

Además, se incluye un apartado de anexo donde se encuentran tablas de datos recogidos durante el desarrollo del proyecto.

CAPÍTULO 2.

FUNDAMENTOS TEÓRICOS

2.1 Precipitación

Una precipitación se puede definir como el producto líquido o sólido de la condensación del vapor de agua procedente de las nubes o del aire y termina su trayectoria en el suelo. Existen diferentes tipos de precipitación, entre los que se encuentran: lluvia, granizo, nieve, rocío, cencellada blanca, escarcha y precipitación de la niebla.

El término lluvia hace referencia a las gotas de agua que caen desde una nube cuyo diámetro es superior a 0.5mm. La creación de las gotas se produce con el choque de millones de pequeñas gotas de agua en las nubes, dando lugar a la creación de gotas mayores. Cuando estas nuevas gotas que se forman son demasiado pesadas se precipitan desde las nubes en forma de lluvia. A diferencia de la forma con la que se suele representar las gotas, estas no poseen forma de lágrima. Su forma varía con su tamaño y con la influencia de otros factores como las ráfagas de viento [1].

Para el cálculo de la cantidad total de precipitación que llega al suelo en un determinado período, se emplean términos de profundidad vertical de agua, o su equivalente en el caso de formas sólidas, que cubriría una proyección horizontal de la superficie de la Tierra. En el caso de una nevada se expresa como el espesor de nieve reciente que cubre una superficie horizontal plana. La unidad de precipitación es la profundidad lineal, normalmente en milímetros (volumen/área) o en $\text{kg} \cdot \text{m}^{-2}$ (masa/área) para la precipitación líquida.

Con respecto a la resolución, a la precipitación por día se le debería aplicar una resolución de 0.2 mm y, de ser posible, una resolución de 0.1 mm. A las cantidades semanales o

mensuales se les debería aplicar una resolución de, al menos, 1mm. Las mediciones diarias deben realizarse a horas fijas y comunes para toda una red meteorológica. La precipitación inferior a 0.1mm se denomina generalmente traza. La intensidad de precipitación se expresa igualmente en medidas lineales por unidad de tiempo, normalmente en milímetros por hora (mm/h). En la Tabla 2 se exponen los criterios para determinar la intensidad.

Tabla 2. Criterios para determinar la intensidad (i) de precipitación débil, moderada o fuerte.

VARIABLE	RANGO	INTENSIDAD
Llovizna	$i < 0.1 \text{ mm/h}$	Débil
	$0.1 \leq i < 0.5 \text{ mm/h}$	Moderada
	$i \geq 0.5 \text{ mm/h}$	Fuerte
Lluvia (chubasco)	$i < 2.5 \text{ mm/h}$	Débil
	$2.5 \leq i < 10.0 \text{ mm/h}$	Moderada
	$10.0 \leq i < 50.0 \text{ mm/h}$	Fuerte
	$i \geq 50.0 \text{ mm/h}$	Violenta
Nieve	$i < 1.0 \text{ mm/h}$ (equivalente en agua)	Débil
	$1.0 \leq i < 5.0 \text{ mm/h}$ (equivalente en agua)	Moderada
	$i \geq 5.0 \text{ mm/h}$ (equivalente en agua)	Fuerte

Dependiendo de la temperatura existente en los kilómetros inferiores de la atmósfera variará el tipo de precipitación que finalmente llegue a la superficie terrestre.

Otro fenómeno a tener en cuenta es el rocío, fenómeno que tiene lugar por la noche de proporciones relativamente pequeñas. Este fenómeno es de especial interés en las zonas áridas, puesto que en estas regiones puede alcanzar el mismo orden de magnitud que la lluvia. Con el fin de evaluar la contribución hidrológica del rocío, es necesario distinguir entre tres tipos:

- Como resultado del transporte hacia abajo de la humedad atmosférica condensada en superficies frías, denominado rocío de precipitación.

- Por el vapor de agua procedente del suelo y de las plantas y condensado en superficies frías, denominado rocío de destilación
- Como agua exudada por hojas, conocido como gutación.

Las tres fuentes pueden contribuir simultáneamente a la formación del rocío, aunque solo la primera aporta agua adicional a la superficie, y la última tiene generalmente como resultado una pérdida neta. Otra fuente de humedad es la procedente de gotitas de niebla o de nube reunidas por las hojas y ramitas de la vegetación, que llegan al suelo por goteo o resbalando sobre los tallos [1].

El hielo puede acumularse en las superficies como resultado de varios fenómenos: la acumulación de hielo como consecuencia de la precipitación congelada, denominada con frecuencia hielo liso o cencellada transparente. La cencellada blanca, depósito de hielo opaco, de aspecto granular y color blanco o lechoso, formado por la rápida congelación de gotas de agua subfundidas al tomar contacto con un objeto expuesto. La escarcha (denominada generalmente helada) es un depósito de cristales de hielo entremezclados, se forma cuando el aire con una temperatura de punto de rocío por debajo del nivel de congelación se satura por enfriamiento.

La niebla consiste en gotas de agua muy pequeñas suspendidas en la atmósfera, que forman una nube sobre la superficie terrestre. Las gotas de niebla tienen un diámetro del orden de 1 a 40 μm , y velocidades de caída comprendidas entre menos de 1 y 5 cm/s aproximadamente. En realidad, las velocidades de caída de las gotas de niebla son tan bajas que, incluso con vientos débiles, pueden desplazarse casi horizontalmente. En caso de niebla, la visibilidad horizontal es en general inferior a 5 km, y raramente se observa cuando la temperatura y el punto de rocío difieren en más de 2°C. Desde un punto de vista hidrológico, existen zonas forestales elevadas que experimentan frecuentes episodios de niebla como resultado de la advección de nubes sobre la superficie de la montaña, de donde, si se considera solo la precipitación, se puede subestimar notablemente la entrada de agua [7].

2.2 Características generales de los radioenlaces

Este apartado está centrado en los radioenlaces bidireccionales dado que esta es una de las características que definen a los radioenlaces satelitales, los cuales son objetos de este TFG.

Los radioenlaces normalmente transportan señales multiplexadas, aunque existen casos en los que emiten una única señal, a estos se les llama ‘radioenlaces monocanales’. Por ello, la portadora suele ser modulada y no se realiza la transmisión en banda base.

Debido a esto, lo normal es que cada radioenlace transmita dos portadoras moduladas (comunicación dúplex a cuatro hilos equivalentes), una por cada uno de los sentidos, subida y bajada. La palabra ‘radiocanal’ hace referencia a la pareja de portadoras asociadas a un radioenlace. Los sistemas dúplex permiten la comunicación bidireccional (en los dos sentidos). Los sistemas a cuatro hilos equivalentes emplean un único soporte físico para transmitir la información y frecuencias distintas para cada uno de los sentidos de transmisión.

El rango de frecuencias más utilizado es el comprendido entre 2 y 50 GHz. No obstante, se están comenzando a explotar radioenlaces con frecuencias superiores.

El uso del espectro de frecuencias es limitado, de ahí que deban establecerse planes para regular su uso. Dado que se requieren dos frecuencias por radiocanal, es necesario que éstas se encuentren suficientemente separadas para evitar acoples e interferencias entre las señales de ambos sentidos. En ciertas ocasiones es necesario aplicar técnicas de tratamiento de la señal como las codificaciones o las ecualizaciones debido a las limitaciones de frecuencia que originan distorsiones y otra serie de efectos indeseados [8].

2.2.1 Definición y tipos de radioenlaces

La palabra radioenlace hace referencia a la interconexión que se establece mediante ondas radioeléctricas entre dos terminales. Esta interconexión se lleva a cabo con el fin de

transmitir información entre dos puntos situados a una distancia determinada a la que se denomina ‘longitud del radioenlace’.

De acuerdo con sus características, los radioenlaces se podrían clasificar de la siguiente forma:

- Según los terminales:
 - Fijo o móvil
 - Terrenal o espacial
- Según la modulación:
 - Analógico
 - Digital

Atendiendo los terminales, los radioenlaces se pueden dividir en dispositivos móviles o fijos. Además, podrían dividirse en terrenales o espaciales. Se denomina radioenlace fijo a aquel cuyas terminales se encuentran siempre situadas en el mismo espacio físico, mientras que, si alguno de los terminales se traslada, recibe el nombre de radioenlace móvil. Los de clase terrenal son aquellos que tienen todos sus terminales situados en la superficie terrestre. Recibe el nombre de radioenlace satelital en el caso de que uno de los terminales se encuentre en el espacio.

Dependiendo de su modulación, se pueden clasificar en radioenlaces analógicos o digitales. Se dice que un radioenlace es analógico cuando para la modulación de la portadora se emplea una modulación analógica, por el contrario, se considera digital si su portadora cuenta con modulación digital, ya sea binaria o multinivel. Normalmente se emplean modulaciones de fase coherente (diversos tipos de PSK - *Phase Shift Keying*) y modulaciones multinivel mixtas de fase y amplitud (como las M-QAM, con M igual a 16, 64 o 128) [8].

2.2.2 Parámetros de un radioenlace

A continuación, se va a explicar los parámetros principales a tener en cuenta a la hora del diseño del radioenlace, tales como pérdidas, interferencias, desvanecimientos, etc.

Tabla 3. Parámetros de un radioenlace.

PARÁMETRO	UNIDAD
Potencia transmitida y potencia recibida	<i>dBm</i>
Ganancia de la antena transmisora y de la antena receptora	<i>dB</i>
Pérdidas en los terminales, originadas por el acoplamiento de los circuitos a las antenas	<i>dB</i>
Pérdida básica de propagación	<i>dB</i>
Factor de ruido del receptor	<i>dB</i>
Relación portadora a ruido	<i>dB</i>
Probabilidad de error (enlaces digitales)	<i>%</i>

2.2.3 Ventajas e inconvenientes de los radioenlaces

Entre las ventajas que pueden ofrecer los radioenlaces destacan su rápida y fácil instalación con coste reducido. Además, es posible superar las irregularidades presentes en el terreno, solo hay que atender a la regulación de los equipos, ya que las características del medio pueden considerarse constantes en la banda de trabajo, y permite una interconexión sencilla entre dos puntos alejados.

Por el contrario, se requiere una visibilidad directa, se necesita linealidad para repetidores y antenas directivas, necesitan un gran ancho de banda de la señal en radioenlaces digitales, además de que la segregación de canales no es tan flexible como en sistemas por cable.

2.3 Enlaces satelitales

Los enlaces entre estaciones terrenas y los satélites o entre satélites están constituidos por radiación electromagnética dirigida en forma de haces, similares en algunas de sus características a los enlaces entre estaciones ubicadas sobre la superficie terrestre.

Existen tres tipos de enlaces:

- Enlace de subida de las estaciones terrenas a los satélites.
- Enlace de bajada de los satélites a las bases terrenas.
- Enlace inter-satelital.

Los enlaces de subida y de bajada consisten en portadoras de RF (*Radio Frecuency*) moduladas, mientras que los enlaces satelitales pueden ser tanto RF, como ópticas. Las portadoras son moduladas por señales de banda base, por lo general, cuando se trata de información para propósitos de comunicación. Las conexiones entre usuarios finales requieren de enlaces de bajada, enlaces de subida y posiblemente uno o varios enlaces satelitales.

Para lograr que los enlaces por satélite cumplan con los requisitos de una determinada red de comunicación deben considerarse las características del equipo de las estaciones terrenas y los transpondedores de los satélites que forman parte de la misma, las del medio de propagación y los efectos de radiaciones no deseadas de origen externo. La banda de frecuencia en que opere una red determinada hace que algunos de los factores mencionados tengan una importancia menor o mayor en el diseño de los enlaces.

El diseño correcto de un enlace satelital asegura la recepción de una señal de buena calidad, evitando el desperdicio de recursos técnicos y económicos, y optimizando la capacidad del satélite y estaciones terrenas.

La señal emitida por la estación transmisora debe llegar a la receptora con la potencia suficiente para garantizar la calidad esperada de la comunicación, a pesar de las pérdidas y el ruido introducidos en su propagación y recepción, de tal forma que en el punto de destino la relación de la potencia de la portadora al ruido acumulado que se simboliza por C/N , incluyendo todas las fuentes de interferencia, que tenga el valor requerido para la red considerada. La relación C/N mínima útil depende del tipo de información, su acondicionamiento, su modulación y si está codificada o no.

De hecho, la finalidad última del diseño de un enlace completo, incluyendo los tramos de subida y de bajada, es cumplir con el valor específico de C/N requerido, o con otra relación

equivalente que en el caso de señales digitales es frecuentemente E_b/N_0 o energía por Bit de información transmitido entre la densidad de ruido.

Este proceso desglosado se explica de la siguiente manera:

En la estación terrena transmisora los elementos más importantes son el amplificador de potencia y la antena que radia la señal portadora amplificada. Cuando la señal llega al satélite, hay que considerar las características de la antena receptora y del preamplificador de bajo ruido conectado después de ella. En el trayecto de subida, la señal portadora sufrirá atenuación por propagación en el espacio libre, por absorción en la atmósfera, por lluvia, por despolarización, y por apuntamiento o desalineación de los ejes de las antenas de transmisión y recepción. Asimismo, dentro de su propio ancho de banda, tal vez recibirá interferencias de RF y, sin lugar a dudas, la introducción de ruido. Si se le descuenta las atenuaciones a la señal de RF original que sale de la antena transmisora de la base terrena y se le añade la ganancia de la antena receptora del satélite se obtiene C . El ruido acumulado se puede englobar como un ruido total equivalente N , que se toma después de la antena receptora. Conocida la potencia C y la potencia de ruido total N , se obtiene el cociente $(C/N)_s$, donde el subíndice s indica que corresponde a un enlace de subida.

La señal portadora, junto con todo el ruido acumulado, es procesada en el interior del satélite, y poco antes de partir de regreso hacia la Tierra, es amplificada en el transpondedor que le corresponde, por su propio amplificador de potencia. Si hay más de una sola portadora dentro del amplificador, irremediamente se producirá ruido de intermodulación. A mayor de salida del amplificador, habrá mayor ruido de intermodulación, por lo que, a bordo del satélite, se define un cociente de potencia de la portadora sobre el ruido de intermodulación, designado como $(C/N)_i$, donde el índice i indica la etapa de intermodulación, y el valor del cociente de potencias lo especifica, por medio de tablas y gráficas, el propio fabricante del satélite.

Como tercera y última etapa se tiene al enlace de bajada. Aquí sucede lo mismo que con el enlace de subida. La señal portadora parte revitalizada de la antena transmisora del satélite, después de haber sido amplificada por el amplificador de potencia del transpondedor y la ganancia directiva de la antena del satélite. Conforme desciende, se atenúa por las mismas

razones que en el enlace de subida, y al llegar a la estación receptora, después de ser amplificada por la antena, queda “algo” de potencia inmediatamente antes de la entrada del amplificador de bajo ruido. Se introducirá ruido de diversas fuentes dentro del ancho de banda de RF de la señal portadora, y se puede calcular otro cociente de potencias $(C/N)_b$, donde b indica que se trata del enlace de bajada. Se debe recordar que la frecuencia de la portadora en el enlace de bajada es diferente a la del enlace de subida, de modo que los niveles de atenuación y de ruido también son diferentes.

Como se puede ver, para el cálculo del cociente $(C/N)_{Total}$, es necesario tomar en cuenta muchas variables y características del medio, por lo que, antes de explicar cómo se calcula un enlace, es importante conocer ciertos conceptos que serán utilizados antes de explicar el procedimiento correspondiente [8].

2.3.1 Parámetros de los radioenlaces satelitales

En este caso se va a exponer los parámetros principales a tener en cuenta a la hora del diseño del radioenlace satelital.

Tabla 4. Parámetros de un radioenlace satelital.

PARÁMETRO	UNIDAD
Inclinación	°
Distancia del vano	<i>km</i>
Frecuencia RF	<i>Hz</i>
Diámetro de la antena parabólica receptora	<i>m</i>
Eficiencia estimada	-
Longitud de onda	<i>m</i>
Pérdidas de espacio libre	%
Atenuación debida a la absorción y dispersión por hidrometeoros	-
Potencia transmitida y potencia recibida	<i>dBm</i>

2.3.2. Potencia radiada

Ya se explicó las características de las antenas, en las estaciones terrenas y en los satélites, por lo que se sabe que éstas tienen una ganancia máxima que depende de su área efectiva y la longitud de onda de operación.

Si la antena transmisora no fuera directiva, sino isotrópica, entonces la potencia de transmisión P_T se distribuiría de manera uniforme en un frente de onda esférico, a una distancia o radio r desde la antena, la superficie esférica sobre la que dicha potencia estaría distribuida tendría un área igual a $4\pi r^2$, y por cada unidad de superficie habría una cierta densidad de flujo de potencia (nivel de iluminación), F_d , igual a:

$$F_d = \frac{P_T}{4\pi r^2} [W/m^2] \quad (2.1)$$

Sin embargo, la antena transmisora no es isotrópica, sino directiva, con una ganancia G_T , que es máxima en la dirección hacia el satélite, siempre y cuando esté bien orientada. Esta ganancia aumenta la densidad de flujo de potencia, por lo que, en la dirección hacia el satélite realmente es igual a:

$$F_d = \left(\frac{P_T}{4\pi r^2} \right) G_T [W/m^2] \quad (2.2)$$

Como ya se sabe, la PIRE (Potencia Isotrópica Radiada Equivalente), es la potencia efectiva que es radiada inicialmente, y es calculada con el producto:

$$PIRE = P_T G_T [W] \quad (2.3)$$

Suele ser representada en decibelios siendo:

$$[PIRE]_{dBW} = 10 \log P_t + 10 \log G_T [dBW] \quad (2.4)$$

Conociendo la ganancia de la antena en decibeles suele ser $10 \log G_T = [G_T]_{dB}$.

Ahora, al colocar una antena con área efectiva de recepción A_{ef} frente a la onda viajera con una densidad de flujo de potencia F_d , se obtiene la potencia capturada por la antena receptora:

$$Potencia\ interceptada\ o\ recibida = (F_d)(A_{ef}) [W] \quad (2.5)$$

Utilizando los conceptos de antenas vistos anteriormente, considerando a la antena receptora con una ganancia G_R y cambiando a la potencia interceptada o recibida por C , siendo la señal portadora transmitida desde la base terrena, tenemos que:

$$C = (F_d)(A_{ef}) = \left(\frac{P_T G_T}{4\pi r^2}\right) \left(\frac{G_R \lambda^2}{4\pi}\right) = (PIRE)(G_R) \left[\frac{\lambda^2}{(4\pi)^2 r^2}\right] [W] \quad (2.6)$$

Al tercer término de la ecuación se le conoce como ganancia debido a la propagación en el espacio libre y como se puede ver es muy pequeña debido a la longitud de onda y el valor r , ya que se expresa en miles de kilómetros.

Las pérdidas de potencia más considerables en un enlace satelital ascendente o descendente se deben a la enorme distancia que hay entre el satélite y las estaciones terrenas. Aunque su valor exacto depende de la frecuencia, tales pérdidas son del orden de 200 dB en la subida y otros 200 dB en la bajada para satélites geoestacionarios. A dicha disminución de potencia debida a la distancia viajada por una señal portadora se le refiere como atenuación por propagación en el espacio libre (FSPL, *Free Space Path Loss*). Siendo la atenuación el inverso de la ganancia, tenemos que:

$$FSPL = \left(\frac{4\pi r}{\lambda}\right)^2 \quad (2.7)$$

Por lo tanto:

$$FSPL = \left(\frac{4\pi r}{\lambda}\right)^2 \quad (2.8)$$

$$C = \frac{(PIRE)(G_R)}{FSPL} [W] \quad (2.9)$$

En decibelios quedaría:

$$[C_{dbW}]_{subida} = [PIRE]_{dbW} + [G_R]_{dBi} - [FSPL]_{dB} \quad (2.10)$$

Para el caso del enlace de bajada, se tienen los mismos parámetros. Ahora el satélite tiene un PIRE, igual al producto de la potencia de salida del amplificador de cierto transpondedor por la ganancia de su antena transmisora:

$$PIRE_{satélite} = (P_{Ttranspondedor})(G_T)_{satélite} \quad (2.11)$$

Las pérdidas por propagación en el descenso se calculan usando la frecuencia de bajada, y la potencia de la portadora recibida por la estación terrena:

$$[C_{dbW}]_{bajada} = [PIRE_{satélite}]_{dbW} + [G_{Restación\ terrena}]_{dBi} - [FSPL]_{dB} \quad (2.12)$$

Para efectos prácticos, el FSPL se expresa como:

$$FSPL = \left(\frac{4\pi r f}{c}\right)^2 \quad (2.13)$$

Debido a que se acostumbra a calcular un enlace a una frecuencia de operación determinada y no a cierta longitud de onda.

El rango satelital, es la distancia r que hay entre una estación terrena y un satélite. Esta distancia se calcula geoméricamente utilizando el ángulo de elevación θ , las coordenadas geográficas de la base terrena y la posición del satélite geoestacionario.

Primero se obtiene el ángulo de cobertura θ' :

$$\theta' = \text{angcos}[\cos l \cos \Delta L] \quad (2.14)$$

Con este ángulo se calcula el rango satelital S :

$$S = 35,786\sqrt{1.4199 - 0.4199 \cos \theta'} \quad [km] \quad (2.15)$$

Para mayor exactitud, habría que tomar en cuenta la altitud de la estación sobre el nivel del mar. Sin embargo, la diferencia que se obtiene en el rango S es muy pequeña y generalmente se desprecia en los cálculos de niveles de potencia. En cambio, para los cálculos de atenuación por lluvia sí es muy importante considerar dicha altitud.

2.3.3. Satélite HISPASAT

HISPASAT es el operador español de satélites de comunicaciones, líder en la distribución de contenidos en español y portugués.

Con más de 25 años de experiencia, el Grupo HISPASAT mantiene una fuerte presencia en la Península Ibérica y América Latina, donde es ya el cuarto operador satelital. HISPASAT se ha posicionado de forma sólida en mercados de alto crecimiento y posee una base estable de clientes estratégicos.

HISPASAT distribuye a través de su potente flota de satélites más de 1.250 canales de televisión y radio y es un motor clave de la industria aeroespacial española [9].

La Sociedad Mercantil HISPASAT y sus sociedades dependientes componen HISPASAT con la siguiente estructura:

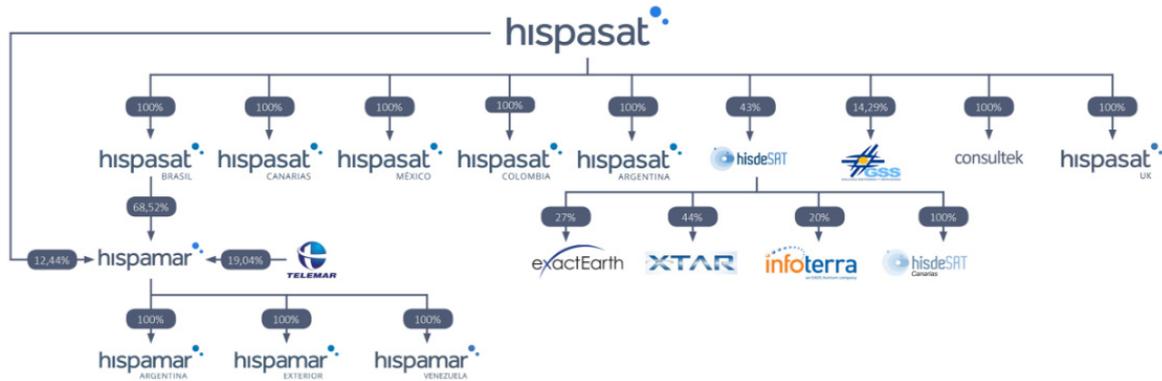


Figura 2. Sociedad Mercantil HISPASAT y sus sociedades dependientes componen HISPASAT.
 Nota: Adaptado de <https://www.hispasat.com/es/> [9].

Desde su constitución en 1989, HISPASAT ha tenido para el Gobierno español y para los usuarios del sistema satelital un alto valor estratégico. Gracias a HISPASAT, España dispone de un sistema nacional de comunicaciones por satélite en bandas de frecuencia especiales y de alta seguridad capaz de satisfacer sus necesidades de comunicación en áreas de interés geoestratégico.

HISPASAT es una pieza clave en el desarrollo de la industria aeroespacial y de telecomunicaciones española, generando importantes retornos, directos e indirectos, que elevan su valor de empresa. La empresa facilita además los medios tecnológicos para intensificar las comunicaciones con América, reforzando los vínculos culturales y de información.

Dado este carácter estratégico, la compañía cuenta en su capital con la presencia del sector público español (Sociedad Estatal de Participaciones Industriales -SEPI- y Centro para el Desarrollo Tecnológico e Industrial -CDTI-) y Red Eléctrica Corporación (REC), que el 3 de octubre de 2019 pasó a ser el accionista mayoritario [9].

2.3.3.1 Flota de satélites

A continuación, se muestra el mapa de cobertura de la flota de satélites de los que dispone HISPASAT. Para este proyecto se ha empleado el satélite llamado “Hispasat 30W-6”, cuyas características se expondrán más adelante.



Figura 3. Mapa de cobertura de HISPASAT.
Nota: Adaptado de <https://www.hispasat.com/es/> [9].

Posición orbital	Satélite	Transpondedores	Año de Lanzamiento
30° Oeste	Hispasat 30W-4	28 Ku	2002
30° Oeste	Hispasat 30W-5	53 Ku, Ka	2010
30° Oeste	Hispasat 30W-6	40 Ku, 7 Ka, 10 C	2018
36° Oeste	Hispasat 36W-1	20 Ku, Ka	2017
55,5° Oeste	Hispasat 55W-2		
61° Oeste	Amazonas 2	54 Ku, 10 C	2009
61° Oeste	Amazonas 3	33 Ku, 19 C, 9 Ka	2013
74° Oeste	Hispasat 74W-1	-	2014
61° Oeste	Amazonas 5	24 Ku, 34 Ka	2017
70° Oeste	Hispasat 70W-1		

*Figura 4. Satélites de la flota de HISPASAT.
Nota: Adaptado de <https://www.hispasat.com/es/> [9].*

2.3.3.2 Hispasat 30W-6

El Hispasat 30W-6, se ubica en 30° Oeste y servirá de reemplazo al Hispasat 30W-4, aportando a HISPASAT capacidad adicional en banda Ku para las regiones andina y brasileña. Asimismo, con el Hispasat 30W-6 (Hispasat 1F) se amplían las capacidades transatlánticas de HISPASAT en sus conectividades Europa-América y América-Europa y se añadirá capacidad en banda C en América y en banda Ka con cobertura europea para que HISPASAT siga incrementando sus servicios de banda ancha en la región.



Figura 5. Satélite Hispasat 30W-6.
Nota: Adaptado de <https://www.hispasat.com/es/> [9].

El Hispasat 30W-6 tiene una vida útil prevista de 15 años y fue construido sobre la plataforma satelital 1300 de SSL (*Space Service Loral*). Cuenta con una potencia de 11,5 kilovatios y una carga útil multimisión distribuida en:

- 40 transpondedores en banda Ku, que darán servicio en dos áreas de cobertura:
 - Europa y norte de África: con cobertura para la Península Ibérica, Islas Baleares, Canarias, Azores, Cabo Verde y Madeira; Mauritania, Marruecos, países mediterráneos del noroeste de África y la mayor parte del continente europeo visible desde 30° Oeste.
 - Panamérica: con cobertura desde Canadá hasta la Patagonia, excluyendo Brasil.
- 6 haces en banda Ka, con cobertura en la Península Ibérica, Islas Baleares y Canarias, noroeste de África y sudeste y centro de Europa.
- 1 haz en banda Ka BSS (*Broad-Band Service*), con cobertura centrada en la Península Ibérica.
- 10 transpondedores en banda C cubriendo el Sur de EE.UU., área del Mar Caribe y todo el continente Sudamericano.

Las características técnicas del satélite se muestran en la siguiente tabla:

Tabla 5. Características técnicas del satélite Hispasat 30W-6.
 Nota: Adaptado de <https://www.hispasat.com/es/> [9].

DESCRIPCIÓN	HISPASAT 30W-6
Posición orbital y coberturas	30°O, 3 coberturas en Ku y C y 7 spots Ka
Transponedores	40 transponedores en banda Ku, 10 en banda C y hasta 6 spots en Ka
Ancho de banda	36 MHz – 72 MHz en Ku y C
Nº de antenas	5
Procesador a bordo	No
Masa seca	-
Masa de lanzamiento	-
Potencia carga útil	10.5 kW
Potencia disponible	11.5 kW
Vida útil	Más de 15 años
Fabricante	SSL (<i>Space Service/Loral</i>)
Lanzador	<i>SpaceX</i>

2.4 Modelos de propagación

La atmósfera es una capa gaseosa de aproximadamente 10.000 km de espesor que rodea al planeta Tierra, en la que se producen todos los fenómenos climáticos y meteorológicos que afectan al planeta. La atmósfera se divide en 5 capas importantes: troposfera, estratosfera, mesosfera, termosfera y exosfera. La troposfera es la capa inferior (la más próxima a la superficie terrestre) de la atmósfera. En esta capa suceden los fenómenos que componen lo que llamamos clima. Estos fenómenos están siempre presentes y, dependiendo de su intensidad, afecta a las telecomunicaciones causando deterioro en las transmisiones vía satélite.

La importancia de la comunicación satelital radica en que casi todo el globo terráqueo puede ser cubierto y se adapta a cualquier aplicación de voz, datos, video, etc. La calidad de un enlace satelital es un aspecto muy importante a ser considerado. Para buscar formas de mejorar la calidad de un enlace es necesario estudiar los factores atmosféricos que afectan a la señal satelital [10].

A la hora de estudiar la propagación de las ondas electromagnéticas a través de un medio, se deberá tener en cuenta una serie de conceptos básicos. Una onda de radio transmitida desde un punto en cualquier dirección se propagará partiendo de ese punto a la velocidad de la luz, la energía de dicha onda viajará en línea recta a menos que exista un obstáculo que se lo impida, y que, de no existir obstáculo, esa energía continuará su trayecto indefinidamente.

En este capítulo se tratará uno de los principales conceptos a tener en cuenta, como es el concepto de pérdida en espacio libre.

Las pérdidas en espacio libre estrictamente hablando no existen como tales, la onda transmitida cumple la ley de la conservación de la energía.

Este concepto hace referencia a un diferencial entre la potencia transmitida en un punto frene a la potencia recibida en otro punto debido a esa transmisión.

Desde el punto de vista físico podemos considerar que la propagación de la energía electromagnética en el espacio libre presenta dependencia con la distancia siguiendo en forma de inversa del cuadrado [1] [11].

2.5 Factores atmosféricos que influyen en la degradación de la calidad de un enlace satelital

En este apartado se presentarán las características y fundamentos físicos de los distintos factores que pueden influir en la degradación de la calidad de la señal y los modelos teóricos que los definen.

En la recomendación UIT-R P.530-16 [12] están determinados los efectos que intervienen en la propagación de una señal y que se debe tener en cuenta en el diseño de cualquier sistema radioeléctrico en condiciones de visibilidad directa:

- Atenuación de una señal, debido a:
 - Presencia de gases en la atmósfera.
 - Existencia de partículas sólidas de la atmósfera.
 - Aparición de precipitaciones.
- Desvanecimiento, como consecuencia de:
 - Difracción ocasionada por obstáculos existentes en el trayecto de la señal.
 - Propagación de la señal a lo largo de varios caminos (multitrayecto).
 - Dispersión en el haz (desenfoque).

Se debe tener en cuenta los problemas que puedan ocasionar estos efectos en la propagación de la señal, por lo que se recomienda tenerlos en cuenta y contabilizarlos siempre que se realice este proceso. Además de los ya mencionados, existen otros efectos que también pueden dificultar que una señal llegue a su destino, entre los que se encuentra la distorsión de la señal, a causa de desvanecimientos selectivos en frecuencia y retardo en su recepción debido a la propagación de la señal por múltiples trayectos; la disminución de la discriminación cruzada (XPD - *Cross Polarization Discrimination*), debida a multitrayectos o a efectos meteorológicos; o las variaciones en los ángulos de llegada o salida de la señal, entre otro [8].

El objetivo principal del estudio es determinar el grado de influencia de los factores atmosféricos sobre un enlace satelital, empleando modelos de atenuación existentes para mostrar el comportamiento de las señales [10].

2.5.1 Atenuación por absorción de gases

Los mayores contribuyentes para la atenuación gaseosa son el oxígeno y el vapor de agua, la absorción gaseosa puede ser un efecto significativo en un diseño de sistema de comunicación satelital dependiendo de la frecuencia específica de operación.

El método y procedimiento para determinar la atenuación por absorción de gases (oxígeno y vapor de agua) está dado por la recomendación UIT-R-P.618 [13]. La fórmula general para realizar el cálculo de la atenuación está dada por:

$$A_t = A_w + A_o \quad [dB] \quad (2.16)$$

Donde A_w es la atenuación por vapor de agua y A_o es la atenuación provocada por el oxígeno [10].

2.5.2 Atenuación por nubes

Las nubes se encuentran constantemente en la tropósfera. La atenuación por nubes se produce cuando las señales satelitales viajan a través de ellas. Existen varios modelos de atenuación por nube, que fueron desarrollados por Gun and East, Liebe, Staelin, Slobin, y Altshuler.

Para el cálculo de atenuación por nubes los modelos elegidos e implementados en la aplicación son el modelo de Liebe y el modelo de Staelin, ya que ambos muestran casi el mismo comportamiento.

2.5.2.1 Modelo de Liebe (Liebe model)

HJ. Liebe creó el modelo en base a la aproximación de Rayleigh sobre la teoría de desvanecimiento de Mie en conjunto con un modelo de relajación que considera la permitividad del agua. El método para el cálculo de la atenuación por nubes según el modelo de Liebe se encuentra detallado en un artículo escrito por GC. Gerace.

2.5.2.2 Modelo de Staelin

El presente modelo deriva de la aproximación de Rayleigh tal como lo hace el modelo de HJ. Liebe, el método para el cálculo de la atenuación según el modelo de Staelin también se encuentra detallada en el artículo de GC. Gerace. La fórmula general de atenuación por nubes es dada por:

$$A = \frac{1}{3} \cdot a \cdot H_{ci} \quad [dB] \quad (2.17)$$

Donde:

a : Atenuación específica, en dB/km

H_{ci} : variable de extensión vertical, en km. Depende del tipo de nube, como se muestra en la tabla 1

Tabla 6. Tipos de nubes.

TIPO DE NUBE	EXTENSIÓN VERTICAL (km) H_c	EXTENSIÓN HORIZONTAL (km) L_c	CONTENIDO DE AGUA (g/m ³)
<i>Cumulonimbus</i>	3,0	4,0	1,0
<i>Cumulus</i>	2,0	3,0	0,6
<i>Nimbostratus</i>	0,8	10,0	1,0
<i>Stratus</i>	0,6	10,0	0,4

2.5.3 Atenuación por lluvia

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. 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 de lluvia a frecuencias de gigahercios.

Debido a que la lluvia es uno de los factores más perjudiciales en las comunicaciones satelitales, estudiamos sus efectos. Las gotas de agua dispersas en la atmósfera interactúan con la radiación, causando atenuación. Según Gagliard, los efectos de la lluvia se vuelven más severos cuando las longitudes de onda se aproximan al tamaño de la gota de agua. Se puede decir que mientras más alta sea la frecuencia de operación, mayor será la atenuación.

El dato de la tasa de caída de la lluvia R , puede ser obtenido de la institución meteorológica local y también se pueden tomar en cuenta los datos de la Tabla 7, basados en la zonificación de regiones climáticas del globo terráqueo que se muestra en la Figura 6 [14].

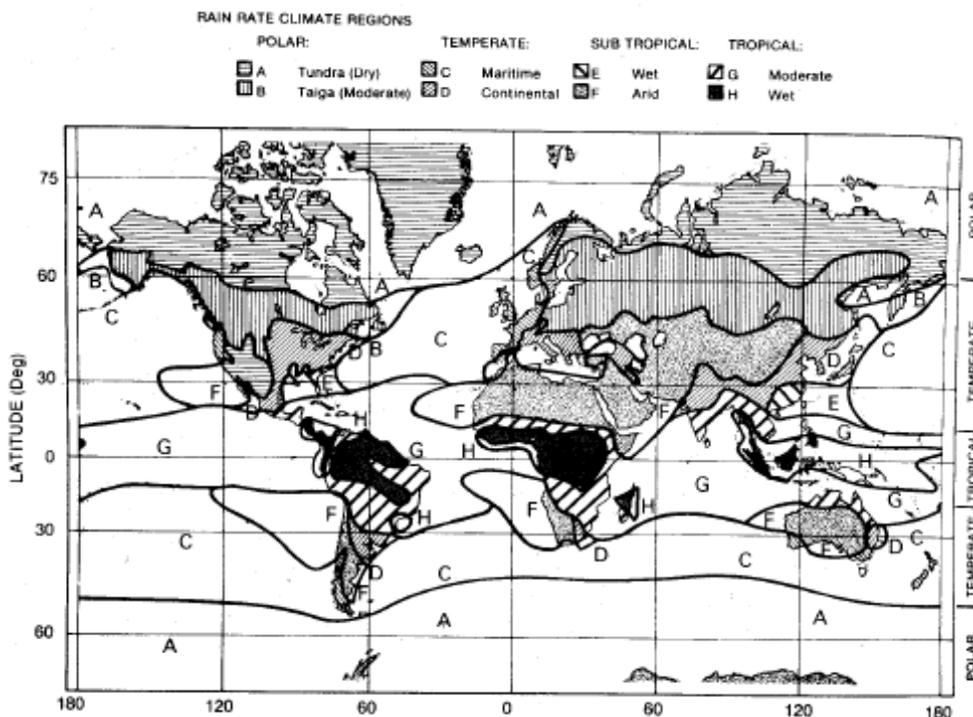


Figura 6. Regiones climáticas – NASA.

Nota: Adaptado de "Estudio de los Factores Atmosféricos que Influyen En La Degradación De Calidad De Un Enlace Satelital", por R.N.Oros Molina, 2010, ACTA NOVA

Tabla 7. Tasas de lluvia.

Porcentaje del año	Regiones climáticas de la lluvia										Minutos por año	Horas por año
	A	B	C	D ₁	D ₂	D ₃	E	F	G	H		
0,001	28	54	80	90	102	127	164	66	129	251	5,3	0,09
0,002	24	40	62	72	86	107	144	51	109	220	10,5	0,18
0,005	19	26	41	50	64	81	117	34	85	178	26	0,44
0,01	15	19	28	37	49	63	98	23	67	147	53	0,88
0,02	12	14	18	27	35	48	77	14	51	115	105	1,75
0,05	8	9,5	11	16	22	31	52	8,0	33	77	263	4,38
0,1	6,5	6,8	7,2	11	16	22	35	5,5	22	51	526	8,77
0,2	4,0	4,8	4,8	7,5	9,5	14	21	3,8	14	31	1052	17,5
0,5	2,5	2,7	2,8	4,0	5,2	7,0	8,5	2,4	7,0	13	2630	43,8
1,0	1,7	1,8	1,9	2,2	3,0	4,0	4,0	1,7	3,7	6,4	5260	87,66
2,0	1,1	1,2	1,2	1,3	1,8	2,5	2,0	1,1	1,6	2,8	10520	175,3

Los parámetros más importantes para realizar el cálculo de la atenuación por lluvia se muestran en la Figura 7.

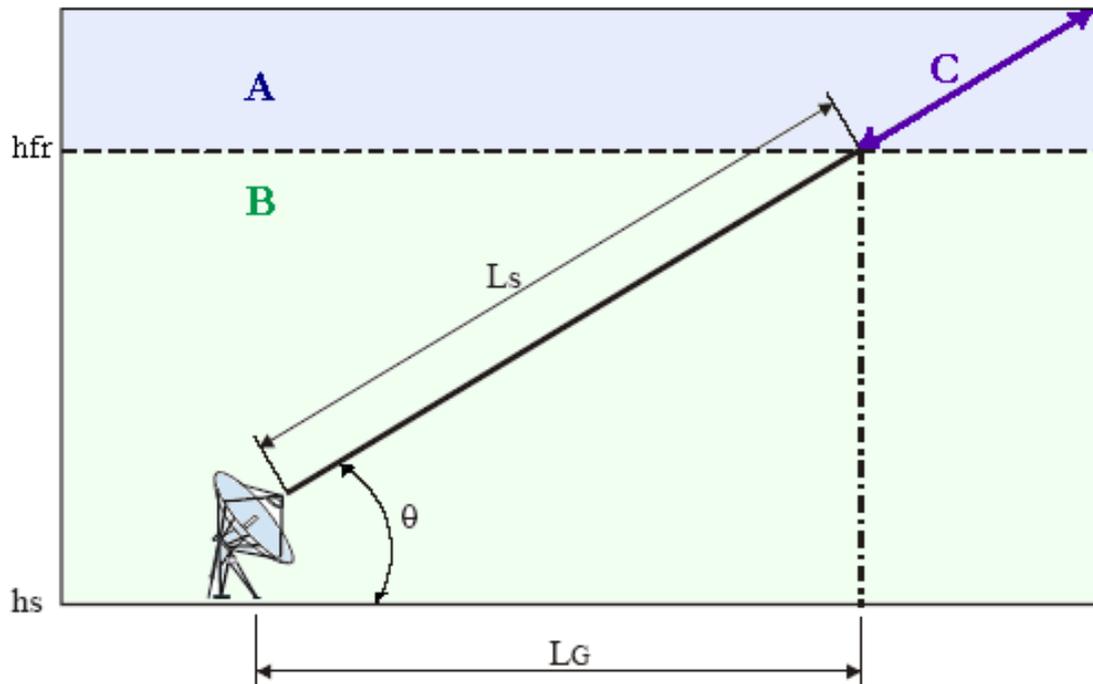


Figura 7. Parámetros de entrada para el cálculo de atenuación por lluvia.

Nota: Adaptado de "Estudio de los Factores Atmosféricos que Influyen En La Degradación De Calidad De Un Enlace Satelital", por R.N.Oros Molina, 2010, ACTA NOVA.

Donde:

A: Precipitación congelada

B: Precipitación líquida

C: Ruta Tierra – Espacio

h_{fr} : Altura efectiva de la lluvia

h_s : Altitud de la estación terrena

L_s : Longitud de ruta oblicua (slant – path length)

L_g : Proyección horizontal de la longitud de ruta oblicua

θ : Ángulo de elevación.

El modelo de atenuación por lluvia empleado en este estudio proviene de la recomendación UIT-R P.618 [13]. La fórmula general de atenuación por lluvia está dada por [10]:

$$A_{0.01} = \gamma \cdot L_e \quad [dB] \quad (2.18)$$

Donde:

γ : Atenuación específica de la lluvia [dB/km]

L_e : Longitud de ruta efectiva a través de la lluvia

2.5.3.1 Cálculo de la atenuación específica debida a la lluvia para modelos de predicción

La obtención de la atenuación específica (γ_R) como consecuencia de la lluvia se estudia en la recomendación UIT-R P.838-3 [15], en la que se establece un proceso de estimación. Se puede definir como:

$$\gamma_R = k \cdot R^\alpha \quad (2.19)$$

La atenuación específica de la lluvia se expresa en dB/km y es, por lo tanto, función de la intensidad de lluvia (R en mm/h) y dos factores que dependen de la frecuencia de trabajo (k y α).

Estas tres variables deben ser estimadas. A continuación, se describe cómo puede hacerse y qué ecuaciones deben emplearse.

1. Determinación de la intensidad de la lluvia (R)

R es la intensidad de lluvia excedida durante un p% del tiempo, siendo p el porcentaje de tiempo, con un tiempo de integración de un minuto, y depende de la situación geográfica en la que se encuentra el radioenlace. La recomendación UIT-R P.387-6 [16] indica un método para establecer la intensidad de lluvia rebasada en un porcentaje p del año medio con un

tiempo de integración de un minuto y para cualquier lugar del mundo. Este método puede utilizarse cuando carecen de medidas locales, aunque es preferible utilizar éstas si se dispone de ellas.

2. Determinación de k y α

Tal y como se ha mencionado tanto k como α son parámetros que dependen de la frecuencia y, además, del tipo de polarización utilizada. Es habitual trabajar con una polarización lineal o circular. Las polarizaciones lineales, a su vez, pueden ser convencionales (horizontales o verticales) o inclinadas

En la Recomendación UIT-R P.838 [15] se muestra el procedimiento para estimar los valores de k y α . Las ecuaciones que lo permiten son:

$$k = \frac{[k_H + k_V + (k_H - k_V) \cdot \cos^2 \theta \cdot \cos 2\tau]}{2} \quad (2.20)$$

$$\alpha = \frac{[k_H \alpha_H + k_V \alpha_V + (k_H \alpha_H - k_V \alpha_V) \cdot \cos^2 \theta \cdot \cos 2\tau]}{2k} \quad (2.21)$$

- k_H y k_V las componentes de k para polarización horizontal y vertical respectivamente
- α_H y α_V las componentes de α para polarización horizontal y vertical respectivamente
- τ es el ángulo de inclinación de la polarización con respecto a la horizontal, en grados. Para la polarización circular, este ángulo valdrá 45°
- θ es el ángulo de elevación del trayecto, en grados

La Recomendación UIT-R P-383-3 [15] no sólo muestra las ecuaciones para estimar los valores de k_H , k_V , α_H y α_V , sino que también presenta sendas gráficas en las que se muestran estos parámetros en función de la frecuencia y tablas de valores. Con estas herramientas pueden obtenerse los parámetros de manera más rápida [8].

2.5.4 Atenuación en la capa fundente

La capa fundente es la región donde las partículas de nieve y hielo se convierten en gotas de lluvia, tiene una temperatura de 0 °C. El grosor de la capa fundente, D_m , se encuentra en el orden de 500 m. La atenuación producida por el derretimiento de partículas de hielo, puede alcanzar valores significativos, particularmente si se trata de enlaces que tienen antenas con un bajo ángulo de elevación [17]. La fórmula general de atenuación en la capa fundente está dada por [10]:

$$A_m = \alpha_m \cdot L_m \quad [dB] \quad (2.22)$$

Donde:

α_m : Atenuación específica [dB/km]

L_m : Longitud de trayecto a través de la capa de fusión

2.5.5 Desvanecimiento por ángulo bajo

Para ángulos de elevación por debajo de los 5°, se producen cambios de niveles de energía debido a irregularidades en gran escala del índice de refracción de la troposfera y a los cambios horizontales producidos por los vientos ligeros.

Se realiza un ajuste a la desviación estándar del centelleo troposférico de la recomendación ITU-R P. 618 [13] para proceder con el cálculo de desvanecimiento por ángulo bajo [10].

$$\sigma_t = \sigma + \sigma_0 \cdot (e^{a \cdot (5-\theta)} - 1) \quad (2.23)$$

Donde:

σ : Desviación estándar del centelleo troposférico calculada con la ecuación 49

σ_0 : Desviación estándar de centelleo troposférico para una frecuencia de 4 GHz,
un diámetro de antena de 4 m y a es una constante empírica igual a 0,11

2.5.6 Atenuación por centelleo

El centelleo troposférico, es producido por las variaciones en pequeña escala del índice de refracción en la troposfera, ocasionando que las señales satelitales lleguen al receptor por medio de varios trayectos. En general el centelleo se encuentra en función de la estación, clima local, frecuencia y ángulo de elevación.

Este efecto depende de diversos parámetros y se da tanto en la tropósfera como en la ionósfera. Es mayor en las frecuencias altas y en ángulos de elevación bajos. También es mayor en las regiones tropicales, en verano y en tiempo nublado.

En frecuencias alrededor de 11 GHz, puede haber variaciones de la amplitud de la señal de 0.3dB el 1% del tiempo y en 20 GHz las variaciones cresta a cresta son generalmente menores de 1 dB para ángulos superiores a 10°.

En el caso del centelleo en la ionosfera, capa inestable de partículas ionizadas a más de 150 km de altura, varía considerablemente con la fase del ciclo de las manchas solares, con la hora del día y con otros factores. En este caso, el mayor efecto se produce en bandas muy bajas, con menor intensidad entre 1 y 6 GHz y excepcionalmente a mayores.

Tanto el centelleo troposférico como el ionosférico tienen en común su bajo valor durante la mayor parte del tiempo en los ángulos de elevación más comunes y duraciones muy cortas en las ocasiones en que alcanzan mayor intensidad. Por tanto, es usual que en condiciones de cielo despejado los enlaces queden protegidos para este propósito a las frecuencias altas con los márgenes de potencia reservados para las condiciones de lluvia.

2.5.6.1 Modelo de centelleo ionosférico

Para predecir la intensidad del centelleo ionosférico en los trayectos Tierra-espacio, se recomienda la utilización del modelo de centelleo ionosférico mundial (GISM - *Global Ionospheric Scintillation Model*). El GISM permite predecir el valor del índice S4, la profundidad del desvanecimiento de amplitud, así como los valores eficaces de la desviación de fase y de la desviación angular debidas al centelleo en función de la ubicación del satélite y de la estación en tierra, de la fecha, hora y de la frecuencia de trabajo. Este modelo se basa

en el método de filtrado de fases múltiples. Los valores por defecto de los principales parámetros internos del modelo son los siguientes:

- Pendiente del espectro de intensidad $p = 3$.
- Tamaño medio de las irregularidades $L0 = 500$ km.
- Desviación típica de las fluctuaciones de la densidad de electrones $\sigma_{Ne} = 0,2$.

Teniendo en cuenta la curvatura de los rayos, se calculan las características de la ionosfera lejana mediante una subrutina que utiliza el modelo ionosférico NeQuick. El código fuente del GISM y la documentación pertinente se pueden consultar en la parte del sitio web del UIT-R correspondiente a la Comisión de Estudio 3 de Radiocomunicaciones [18].

2.5.7 Efecto de la antena mojada

La mayoría de las antenas de enlace de microondas son antenas de bocina con una cubierta protectora de algún tipo a través de la cual se propaga la señal. La humidificación de estas antenas (en particular, la cubierta protectora) con la lluvia causa una atenuación adicional que puede conducir a una sobreestimación severa de la intensidad de la lluvia. La forma más fácil y efectiva de evitar esto es, en primer lugar, evitar que las antenas se mojen. Sin embargo, debido a que la mayoría de las antenas de enlaces de microondas de compañías comerciales de comunicación celular no están protegidas de la lluvia, ha sido necesario realizar una investigación sobre el efecto de este fenómeno y tratar de encontrar algoritmos de corrección.

2.5.7.1 Calibración de la función de atenuación de antena mojada

La relación entre la intensidad de la lluvia y el grosor de la capa de agua formada en las antenas de enlace depende del tipo de antena (por ejemplo, forma, material de cubierta), la dirección del viento, los obstáculos circundantes, etc. Asumiremos aquí que todos los enlaces usan antenas similares, y que el grosor de la película de agua depende solo de R .

La Figura 8 muestra la relación entre la atenuación de la antena mojada y la frecuencia para diferentes intensidades de lluvia para una sola antena. De este gráfico se deduce que

existe una atenuación de antena mojada significativa para todas las intensidades. El aumento de la atenuación con la intensidad de la lluvia es mayor con pequeñas intensidades de lluvia. Esto significa que es importante saber cuándo llueve y cuándo se secan las antenas, ya que la diferencia en la atenuación de la antena mojada es bastante grande (utilizando esta aproximación para la atenuación de la antena mojada). Estas observaciones también se pueden hacer a partir de la Figura 9, donde la atenuación de la antena mojada a 27 GHz se muestra en función de R . Se puede ver que las dos relaciones son bastante similares para las intensidades de baja precipitación, pero se producen desviaciones significativas a intensidades más altas. Sin embargo, como la contribución relativa de la atenuación de la antena mojada es mucho mayor a bajas intensidades de lluvia (es decir, la atenuación a lo largo del camino se vuelve mayor en relación con la atenuación de la antena mojada con el aumento de las intensidades de lluvia), las desviaciones a intensidades más altas son relativamente menos importantes.

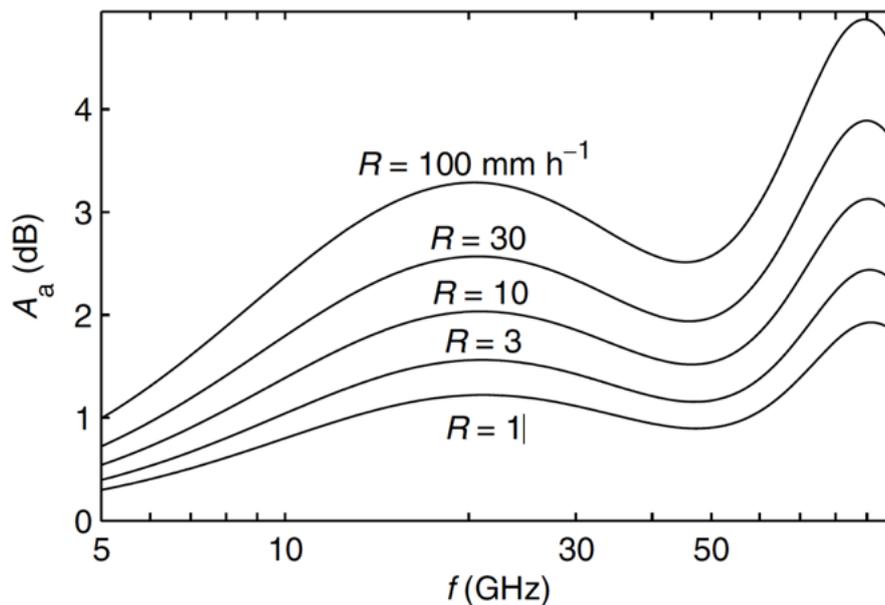


Figura 8. Atenuación de la antena mojada A_a en función de la frecuencia de la señal para diferentes intensidades de lluvia, para una sola antena.

Nota: Adaptado de "Microwave link rainfall estimation: Effects of link length and frequency, temporal sampling, power resolution, and wet antenna attenuation", por H. Leijinse, R. Uijlenhoet, J.N.M Stricker, 2008, Elsevier Ltd.

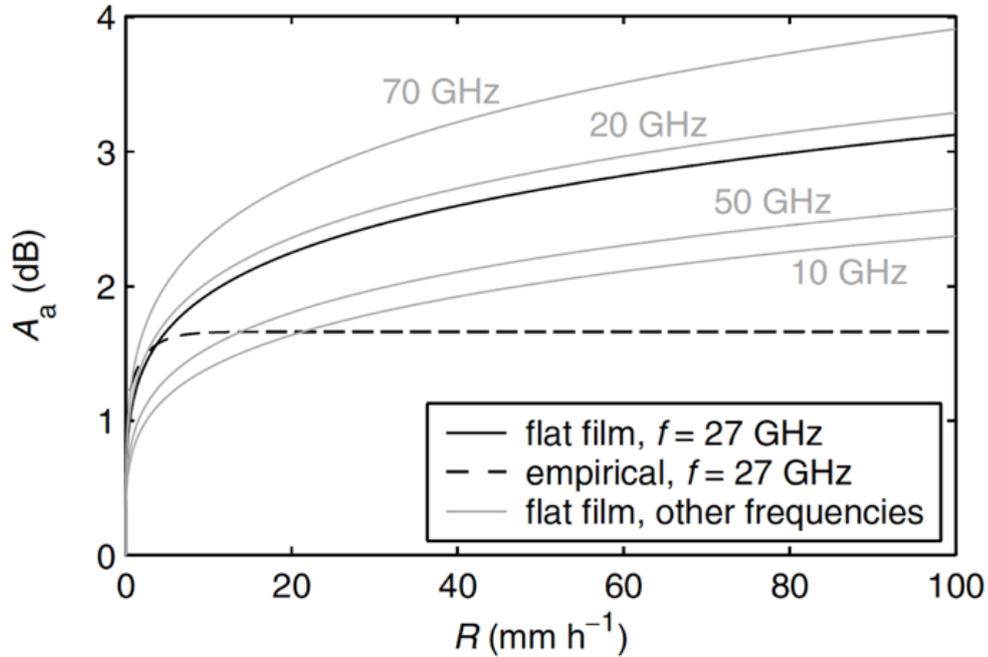


Figura 9. Atenuación de antena mojada A_a en función de la intensidad de lluvia en una sola antena. Se muestra una comparación entre el método actual (película plana) y un método anterior (empírico) a 27 GHz, y las relaciones entre R y A_a para 10, 20, 50 y 70.

Nota: Adaptado de "Microwave link rainfall estimation: Effects of link length and frequency, temporal sampling, power resolution, and wet antenna attenuation", por H. Leijonse, R. Uijlenhoet, J.N.M Stricker, 2008, Elsevier Ltd.

2.5.7.2 Corrección para la atenuación de antena mojada

Las señales de enlace de microondas se pueden corregir para atenuar la antena mojada.

El método empleado supone que la intensidad de la lluvia es constante en todo el enlace, y que la relación entre la intensidad de la lluvia y la atenuación de la antena mojada es conocida. La atenuación total es la suma de la atenuación integrada en el camino por las gotas de lluvia en el camino del enlace y dos veces (dos antenas) la atenuación de la antena mojada.

$$A_m = L \left(\frac{R}{a} \right)^{1/b} + 2A_a(R) \quad (2.24)$$

Como ambas cantidades son funciones de la intensidad de la lluvia, R puede determinarse resolviendo (implícitamente) esta ecuación. Por supuesto, se introducen errores cuando la intensidad de la lluvia no es la misma en las dos antenas y a lo largo del enlace. La razón del primer término al segundo término de la ecuación mostrada anteriormente para un enlace de 1 km se muestra en la Figura 10, en función de la frecuencia del enlace y la intensidad de la lluvia. Tenga en cuenta que los valores para diferentes longitudes de enlace se pueden obtener simplemente multiplicando los valores de la Figura 10 con la longitud del enlace. Se puede ver que la atenuación causada por la humectación de las antenas domina la causada por la lluvia a lo largo del camino para bajas intensidades de lluvia y en bajas frecuencias. La frecuencia para la cual la relación tiene un máximo para una R determinada depende en cierta medida de la propia R , pero se ve que está en el rango $f > 45 \text{ GHz}$.

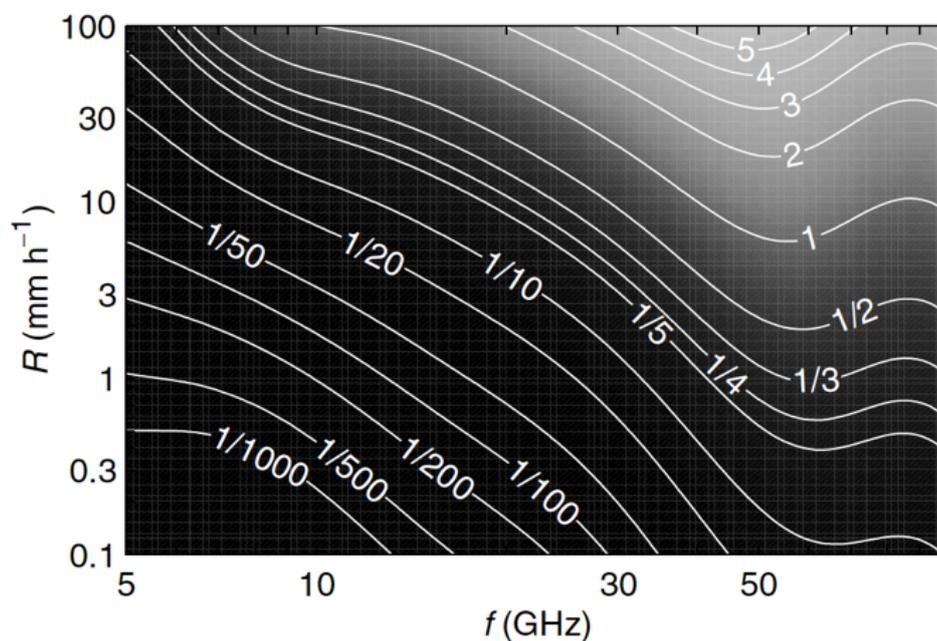


Figura 10. Relación entre la atenuación inducida por la lluvia y la atenuación de la antena mojada en función de la frecuencia de la señal y la intensidad de la lluvia para enlaces de 1 km.

Nota: Adaptado de "Microwave link rainfall estimation: Effects of link length and frequency, temporal sampling, power resolution, and wet antenna attenuation", por H. Leijinse, R. Uijlenhoet, J.N.M Stricker, 2008, Elsevier Ltd.

CAPÍTULO 3.

PREPARACIÓN DEL ENTORNO DE TRABAJO

3.1 Fuentes de datos empleadas

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.

3.1.1 Prototipo satelital VIMETRI-MAC

En la azotea del Pabellón B de la EITE se ha instalado una antena parabólica con su respectivo LNB (*Low Noise Block*) y se ha orientado hacia el satélite HISPASAT 30W-6, que se encuentra a 30° O de elevación. Además, se ha instalado una estación meteorológica Vaisala modelo WXT536 que está conectada a un pluviómetro modelo RG13, también proporcionado por Vaisala. La captación de los datos de estos dispositivos se lleva a cabo a través del prototipo diseñado, que se encuentra ubicado dentro de la caja de registro. Este montaje es el que se presenta en la Figura 11, y constituye un conjunto diseñado para investigación y desarrollo [19].



*Figura 11. Bloqueo de continua para la medición de potencia recibida.
Nota: Adaptado de “Documentación de resultados del proyecto VIMETRI-MAC/MAC 3.5b/065”, por N.Molina, 2019.*

La Figura 12 representa un diagrama de bloques del prototipo diseñado, señalando de forma numérica las etapas que definen su funcionamiento. En función de los pasos indicados con círculos violeta y números blancos en el diagrama, se detalla el funcionamiento de cada una de las etapas que conforman el prototipo.

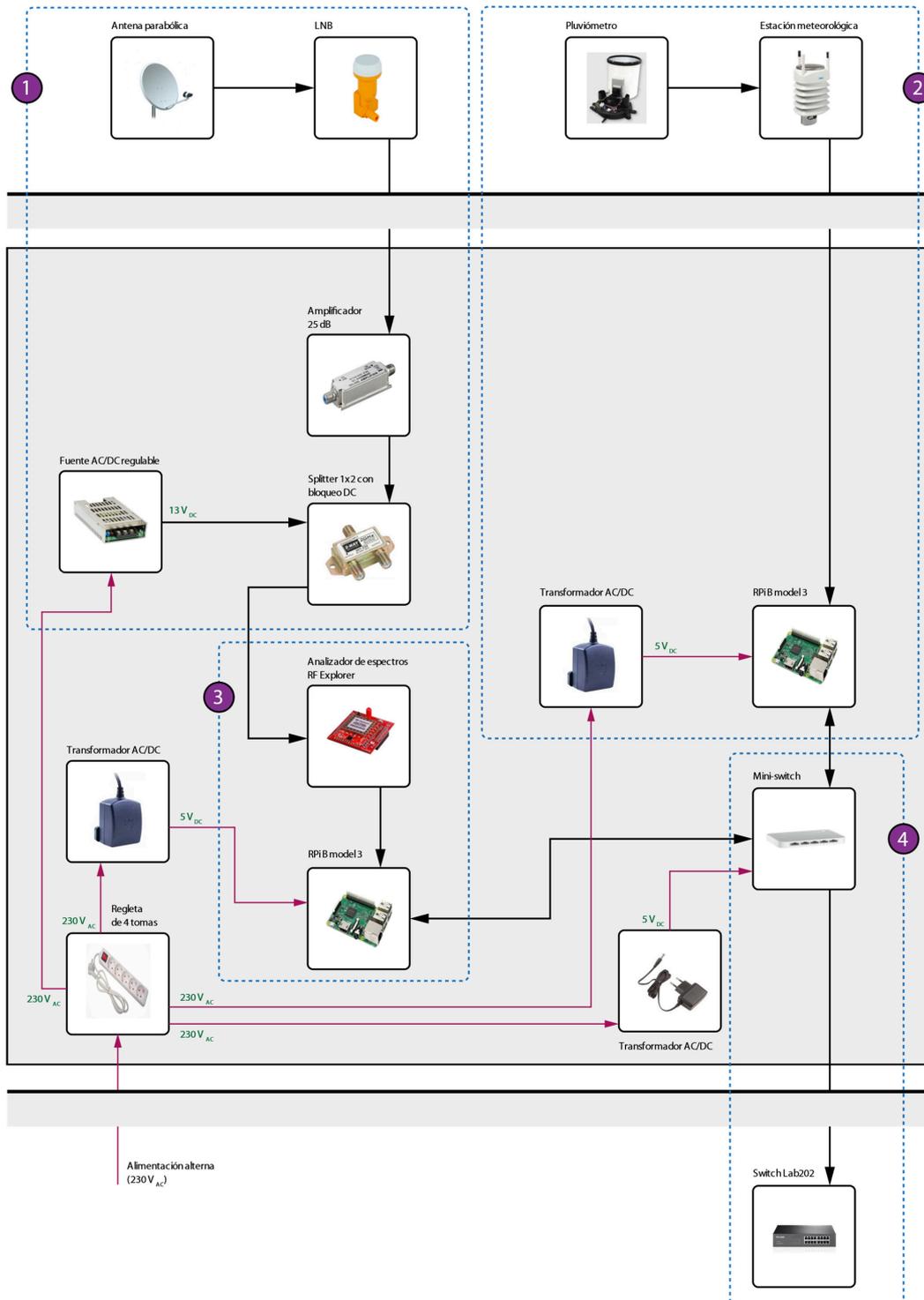


Figura 12. Diagrama de bloques general del prototipo.

Nota: Adaptado de “Documentación de resultados del proyecto VIMETRI-MAC/MAC 3.5b/065”, por N.Molina, 2019.

Etapa 1: instalación del sistema de adquisición de datos radioeléctricos.

Tras realizar la instalación del soporte para la antena parabólica de la Figura 12, esta debe orientarse con una elevación de 30° hacia el oeste, apuntando al satélite HISPASAT 30W-6.

Es necesario usar un analizador de espectro portátil como el que aparece en la Figura 13.



Figura 13. Analizador de espectros portátil (Laboratorio 220, Pabellón B -EITE).

Nota: Adaptado de "Documentación de resultados del proyecto VIMETRI-MAC/MAC 3.5b/065", por N.Molina, 2019.

Al visualizar el espectro que se recibe a través de la parabólica, se observa que la máxima potencia recibida se corresponde con la frecuencia 1422 MHz, con un valor de aproximadamente -67.7 dBm. Por lo tanto, con este valor de frecuencia se programará el analizador de espectro *low-cost* para monitorizar los valores de RSL de forma continua.

Etapa 2: instalación del sistema de adquisición de datos meteorológicos.

Tras realizar la instalación de los soportes para el pluviómetro y la estación meteorológica de Vaisala, tal y como se aprecia en la Figura 12, es necesario conectar el pluviómetro RG13 a la estación WXT536 a través del cable que integra el propio pluviómetro, que se conecta por una de las entradas analógicas de la estación.

Usando el cable de configuración de Vaisala, se fijan los tiempos de adquisición de datos de la estación a 1 segundo, es decir, cada segundo los sensores enviarán un dato capturado.

Los mensajes compuestos propios de Vaisala siguen una configuración NMEA0183 [20].

Al conectarse la estación meteorológica a uno de los puertos serie de la Raspberry Pi, se reciben estos mensajes y extrae los valores indicados anteriormente y los almacena en un fichero CSV (*Comma-Separated Values*).

Etapa 3: extracción de valores de potencia recibida.

La salida del LNB va conectada, a través de un cable coaxial, hasta un amplificador lineal de 25 dB, ubicado en la caja de registro. La salida del amplificador va conectada a la entrada de un *splitter* de una entrada y dos salidas (1x2), donde una de ellas integra internamente un filtro de señal continua. Por las salidas del *splitter* se conecta, en el terminal sin el bloqueador de DC, una fuente de alimentación regulable que se ajusta a 13 V de tensión continua, permitiendo alimentar el LNB. La salida restante bloquea los 13 V de continua y deja pasar la señal alterna de la entrada del *splitter* (atenuada unos 4.5 dB) y pasa hasta el módulo del analizador de espectros, donde será registrada.

Etapa 4: envío de datos a un servidor conectado a Ethernet.

Las dos unidades Raspberry Pi se conectan vía Ethernet a un mini-switch instalado en la caja de registro donde se aloja el prototipo. A su vez, este mini-switch se conecta a un switch instalado en el Laboratorio 202 del Pabellón B de la EITE, que está conectado al servidor que almacena los datos meteorológicos y radioeléctricos capturados por el prototipo, tal y como se muestra en la Figura 14 [19].

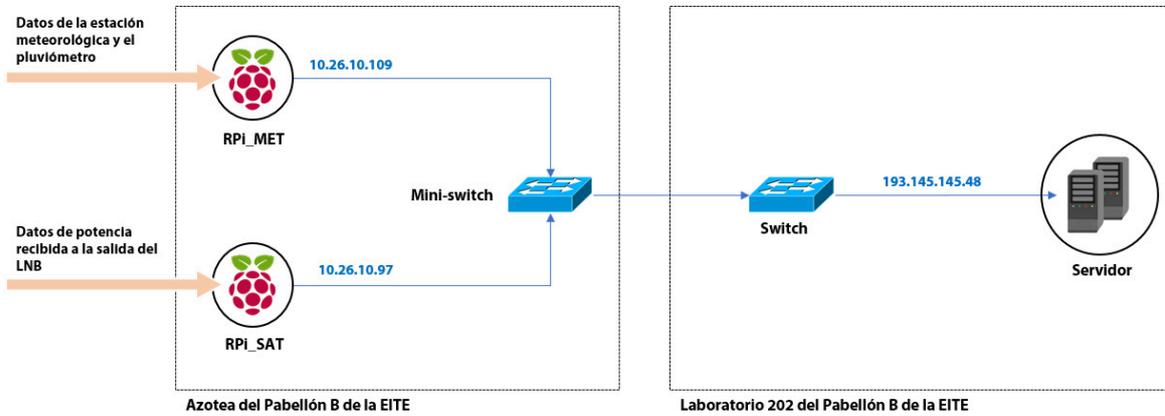


Figura 14. Conexiones de red para el prototipo.

Nota: Adaptado de "Documentación de resultados del proyecto VIMETRI-MAC/MAC 3.5b/065", por N.Molina, 2019.

A continuación, en la Figura 15, se muestra el diagrama de flujo del algoritmo de obtención de los parámetros meteorológicos.

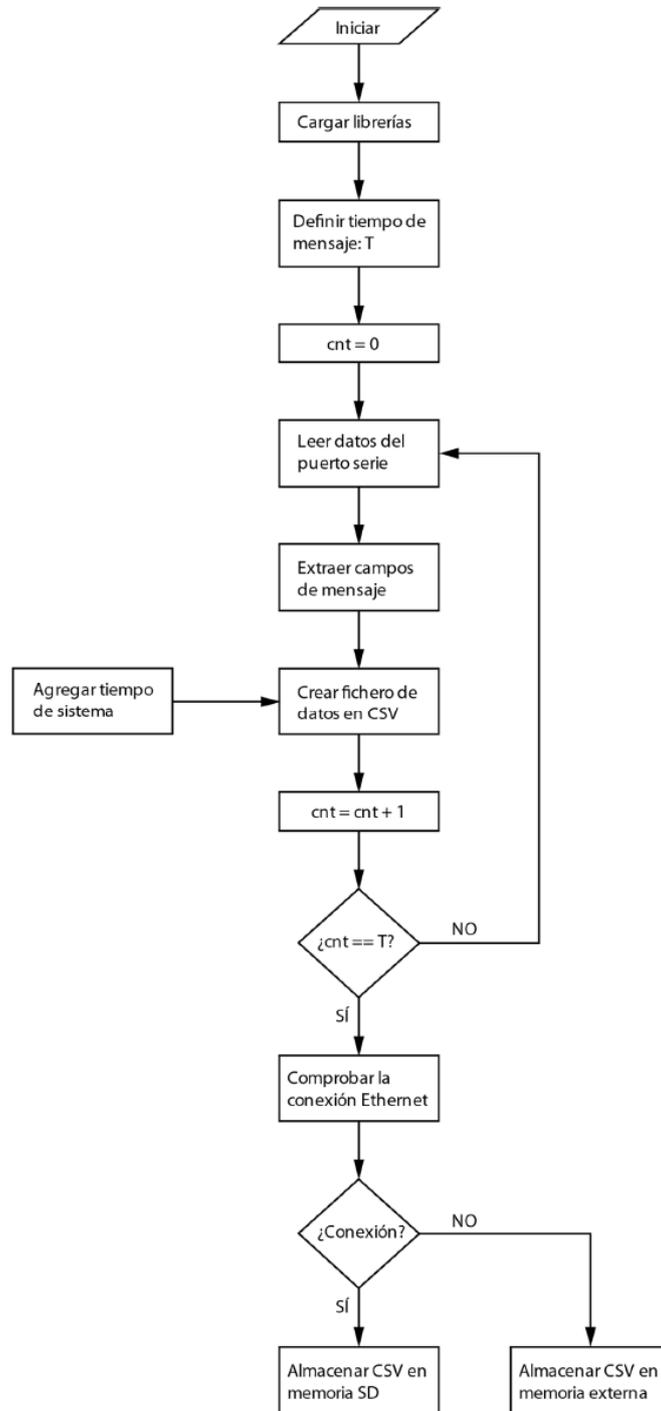


Figura 15. Diagrama de flujo del algoritmo de obtención de parámetros meteorológicos.
 Nota: Adaptado de “Documentación de resultados del proyecto VIMETRI-MAC/MAC 3.5b/065”, por N.Molina, 2019.

El primer paso será cargar las librerías que se van a utilizar, y a su vez, definir un objeto Serial para poder abrir el puerto serie donde se recibirán los datos de la Vaisala WXT536.

Seguidamente, se definen las funciones que serán utilizadas a lo largo del programa. Estas funciones son las siguientes:

- a) *Ethernet_Test*: sirve para comprobar si existe conexión Ethernet. Esta función se combina con la de almacenamiento de datos, de modo que si hay conexión Ethernet se almacenarán los datos sobre la Raspberry Pi y se borrarán una vez se almacenen en la base de datos. En cambio, si no hay conexión Ethernet, los datos pasan a almacenarse en una memoria USB conectada a la Raspberry Pi.
- b) *Generate_CSV_Ethernet*: sirve para crear un fichero CSV donde se almacenan los datos con el formato que se les aplica tras haber sido recibidos. Esta función depende del estado de la red Ethernet para decidir en qué directorio almacenar el fichero: si hay conexión en la propia Raspberry Pi (se borran automáticamente cuando se leen en la base de datos, a través de un fichero *bash*) y si no, en una memoria USB (*Universal Serial Bus*) externa a la Raspberry Pi.

El siguiente paso será ejecutar el programa principal. Durante todo el tiempo se leen los valores del puerto serie, y de los mensajes recibidos se extraen los parámetros que son de utilidad para conocer la meteorología de la zona. Estos parámetros se almacenan en variables y se añade el tiempo de sistema sobre un mensaje compuesto.

Finalmente, un contador añade una unidad cada segundo, y al llegar a valor 15 (es decir, cada 15 segundos) almacena el mensaje en un fichero CSV, sobre el directorio que determina la conexión Ethernet.

3.1.2 Estación meteorológica en el pabellón B de la EITE

En la azotea del Pabellón B de la EITE se encuentra instalado un pluviómetro Vaisala Rain Gauge RG13 como el que se observa en la Figura 16. Este sistema utiliza un mecanismo de tipo balancín, que produce un cierre de contacto cada vez que recibe una cantidad pequeña

predeterminada de lluvia (0.1 mm/ 0.2 mm/ 0.5 mm). Este mecanismo consiste en un cubo dividido pivotado a su centro, la lluvia se almacena en la mitad superior. Cuando este se llena, el mecanismo se inclina y descarga el agua almacenada, permitiendo a la otra mitad del cubo comenzar a llenarse [21].



*Figura 16. Vaisala Rain Gauge RG13.
Nota: Adaptado de "VAISALA - Rain Gauge RG13 , RG13H", por Vaisala, 2017.*

3.1.2 Estación meteorológica de Tafira – El Zurbarán

La estación Tafira (Zurbarán) es una Thies Klima. Fue instalada desde 2008 y está equipada con sensores de temperatura, humedad relativa, anemo-veleta y pluviómetro. Son estaciones modulares, que admiten distintas configuraciones de operación y de conexión de

sensores. Se montan en torno a un mástil de acero de unos 10 metros en el que se ubica la cruceta que soporta el anemo y la veleta (en el tope), así como antena GPRS (*General Packet Radio Service*), placa fotovoltaica de unos 10W y, al pie, armario con clasificación de intemperie. El mástil va fijado a tierra y cimentado con pernos embebidos en un dado de hormigón de $0.6 \times 0.6 \times 0.6$ metros. El núcleo al que se conectan los distintos componentes o sensores es el *DATALOGGER DLx Radiation 9.1755.00.013*, que va recabando información de los sensores de manera continua. Estas estaciones están preparadas para operar con alimentación de batería que se recarga mediante una placa fotovoltaica y un regulador de carga. También, pueden funcionar conectadas a la red eléctrica.

3.2 Adquisición de datos

Los datos con los que se ha llevado a cabo este Trabajo de Fin de Grado han sido recogidos de diferentes fuentes, como son un radioenlace satelital y dos estaciones meteorológicas. A continuación, se explica en mayor detalle cada una de estas fuentes de datos.

3.2.1 Adquisición de datos de los radioenlaces

Los datos de los radioenlaces utilizados en este trabajo han sido obtenidos a través de la plataforma web proporcionada por el propio proyecto VIMETRI-MAC. Los datos adquiridos son niveles de potencia, permitiéndonos seleccionar la franja de tiempo (inicio y fin) en la que queremos obtener estos niveles RSL. Permite tanto descargar los datos seleccionados, como obtener una representación gráfica de dichos niveles. A continuación, se muestra un ejemplo de la gráfica obtenida para el mes de junio de 2020, donde se puede apreciar la variación que ha sufrido los niveles de potencia de la señal (Figura 17).

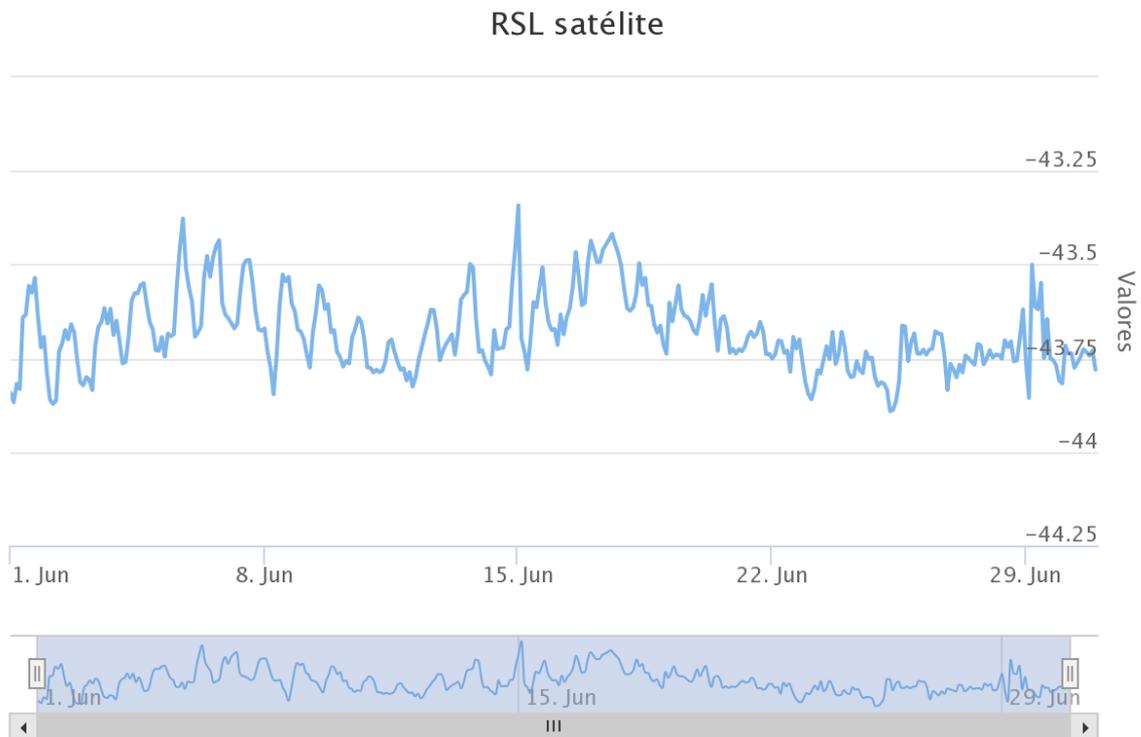


Figura 17. Niveles de potencia de la señal satelital. VIMETRIMAC 2019.

3.2.2 Adquisición de los datos de estaciones

3.2.2.1 Estaciones meteorológicas propias – VAISALA

Estos datos son los proporcionado por el pluviómetro instalado en la azotea del pabellón B de la EITE. Pueden ser obtenido, al igual que los datos RSL, a través de una página web. En ella, seleccionando las fechas correspondientes, se puede proceder a descargar lo datos deseados. Otra opción disponible es la de dibujar una gráfica con los datos de precipitación en mm/h correspondientes al mes de marzo de 2020, como se puede observar en la Figura 18.

Datos de precipitación medidos con pluviómetro

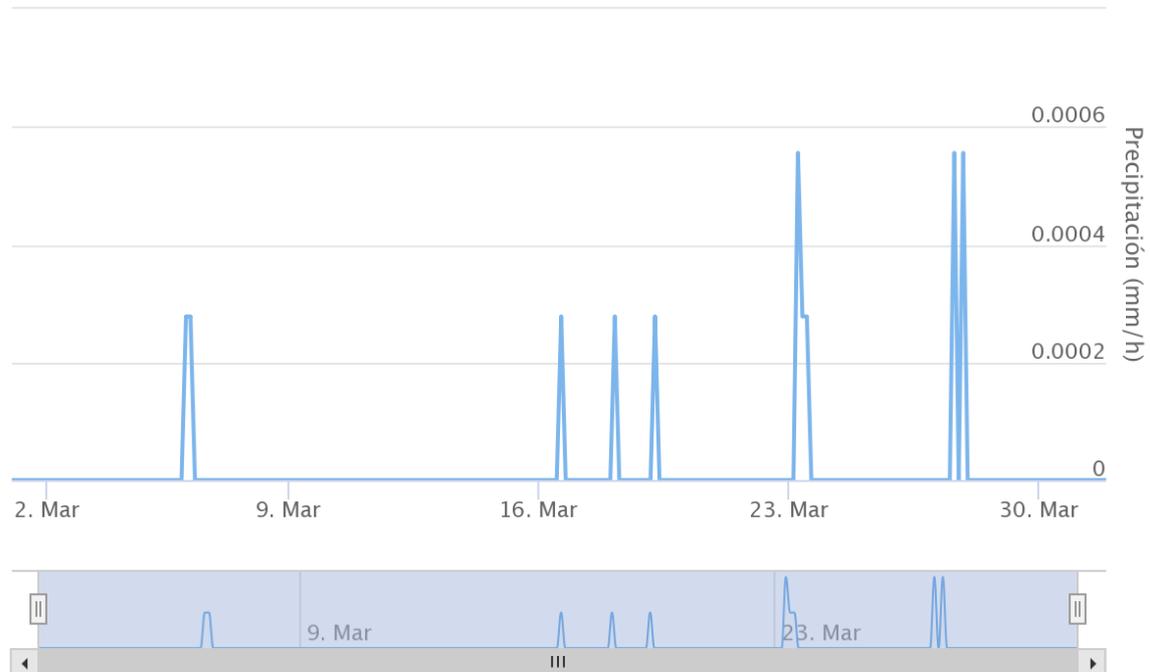


Figura 18. Datos de precipitación medidos con pluviómetro. VIMETRIMAC 2020.

3.2.2.2 Estación meteorológica AEMET – Pluviómetro

Esta estación recoge datos mediante un pluviómetro, los cuales son proporcionados por la AEMET (*Agencia Estatal de Meteorología*). Entre los valores generados se encuentran los correspondientes a la latitud, longitud y altitud de la estación, así como la velocidad máxima del viento, la temperatura y la cantidad de precipitación, entre otros.

3.2.3 Región de trabajo

Este trabajo se ha realizado en la isla de Gran Canaria, empleando datos meteorológicos correspondientes a la zona de Tafira, en la que se encuentran las diferentes estaciones meteorológicas con las que se trabaja en este proyecto.

3.3 Creación de la base de datos

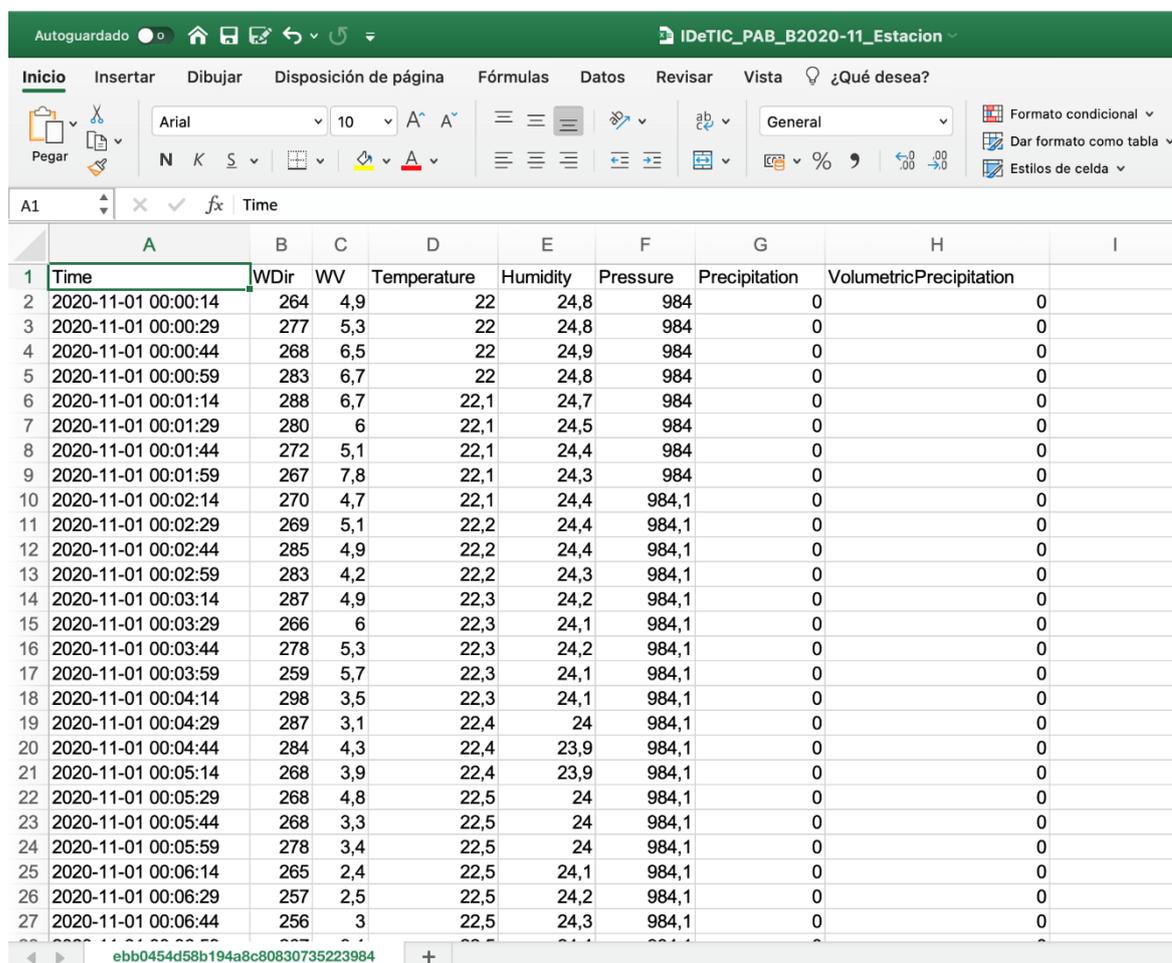
Se ha desarrollado la base de datos en el entorno de Matlab partir de ficheros en formato ‘.xlsx’, recibidos por la AEMET a través de protocolo FTP (*File Transfer Protocol*) para ser alojados en uno de los servidores disponibles en el proyecto VIMETRI-MAC. Dependiendo de la fuente de procedencia, los datos han sido proporcionados con ciertas diferencias, por lo que se debe proceder a realizar un sincronismo de toda esta información. A continuación, se muestran los distintos tipos de datos a tratar.

3.3.1 Tipos de datos

3.3.1.1 Datos meteorológicos

Son los datos recogidos por las estaciones de meteorología. Se trata de archivos ‘.csv’ que incluyen una serie de parámetros: fecha y hora, temperatura (°C), velocidad del viento (km/h), dirección del viento, racha (km/h), dirección de racha, precipitación (mm), presión (hPa), tendencia (hPa) y humedad (%). De todos estos datos, los que nos interesan son: fecha, hora y precipitación, puesto que son los que necesitamos para la realización de este proyecto.

Los datos proporcionados por la estación de VAISALA son muestras tomadas cada quince segundos. Mientras que los datos proporcionados por la estación de El Zurbarán son muestras diezminutales, comenzando por la hora en punto.



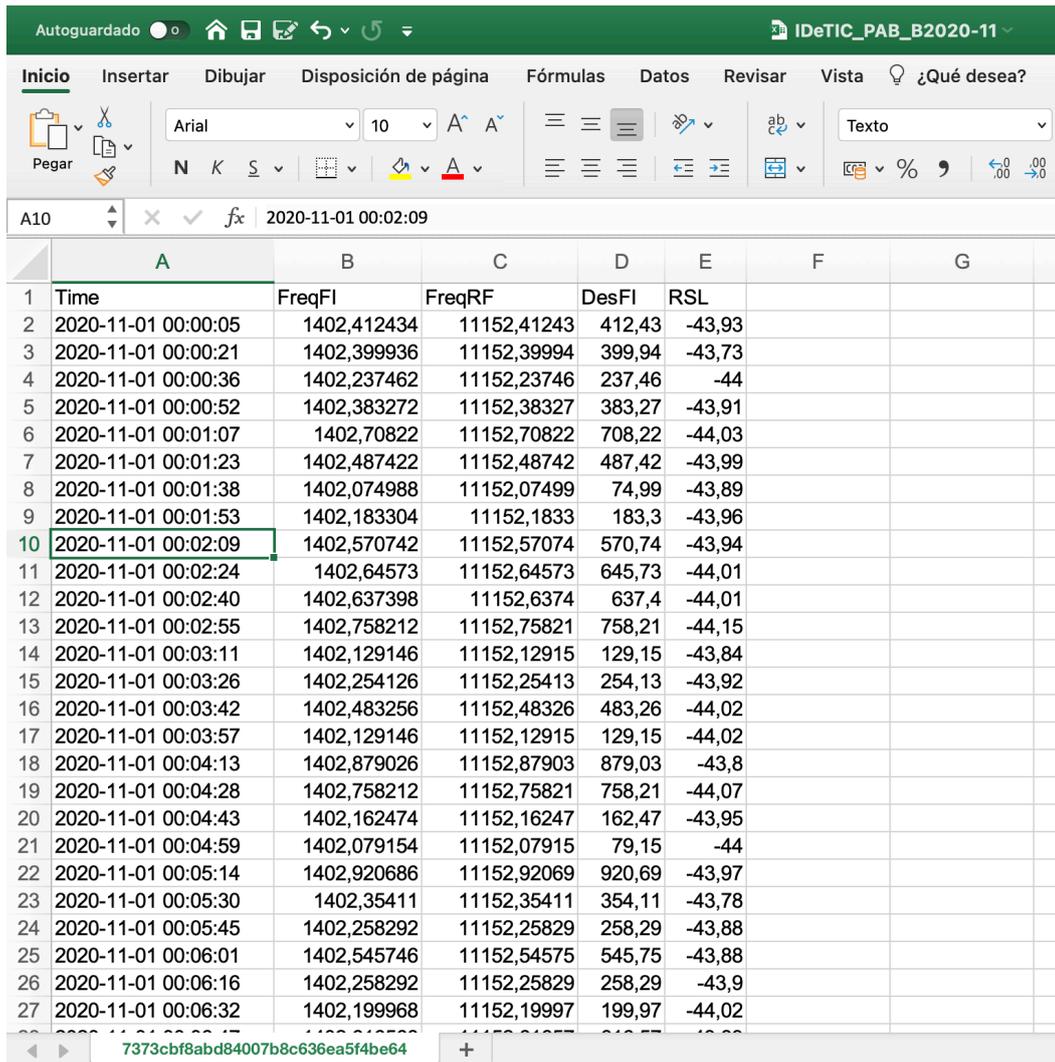
	A	B	C	D	E	F	G	H	I
1	Time	WDir	WV	Temperature	Humidity	Pressure	Precipitation	VolumetricPrecipitation	
2	2020-11-01 00:00:14	264	4,9	22	24,8	984	0	0	
3	2020-11-01 00:00:29	277	5,3	22	24,8	984	0	0	
4	2020-11-01 00:00:44	268	6,5	22	24,9	984	0	0	
5	2020-11-01 00:00:59	283	6,7	22	24,8	984	0	0	
6	2020-11-01 00:01:14	288	6,7	22,1	24,7	984	0	0	
7	2020-11-01 00:01:29	280	6	22,1	24,5	984	0	0	
8	2020-11-01 00:01:44	272	5,1	22,1	24,4	984	0	0	
9	2020-11-01 00:01:59	267	7,8	22,1	24,3	984	0	0	
10	2020-11-01 00:02:14	270	4,7	22,1	24,4	984,1	0	0	
11	2020-11-01 00:02:29	269	5,1	22,2	24,4	984,1	0	0	
12	2020-11-01 00:02:44	285	4,9	22,2	24,4	984,1	0	0	
13	2020-11-01 00:02:59	283	4,2	22,2	24,3	984,1	0	0	
14	2020-11-01 00:03:14	287	4,9	22,3	24,2	984,1	0	0	
15	2020-11-01 00:03:29	266	6	22,3	24,1	984,1	0	0	
16	2020-11-01 00:03:44	278	5,3	22,3	24,2	984,1	0	0	
17	2020-11-01 00:03:59	259	5,7	22,3	24,1	984,1	0	0	
18	2020-11-01 00:04:14	298	3,5	22,3	24,1	984,1	0	0	
19	2020-11-01 00:04:29	287	3,1	22,4	24	984,1	0	0	
20	2020-11-01 00:04:44	284	4,3	22,4	23,9	984,1	0	0	
21	2020-11-01 00:05:14	268	3,9	22,4	23,9	984,1	0	0	
22	2020-11-01 00:05:29	268	4,8	22,5	24	984,1	0	0	
23	2020-11-01 00:05:44	268	3,3	22,5	24	984,1	0	0	
24	2020-11-01 00:05:59	278	3,4	22,5	24	984,1	0	0	
25	2020-11-01 00:06:14	265	2,4	22,5	24,1	984,1	0	0	
26	2020-11-01 00:06:29	257	2,5	22,5	24,2	984,1	0	0	
27	2020-11-01 00:06:44	256	3	22,5	24,3	984,1	0	0	

Figura 19. Archivo de la base de datos de la estación meteorológica correspondiente al mes de noviembre de 2020.

3.3.1.1 Datos del radioenlace satelital

Son los datos recogidos por el equipo de antenas receptoras. Son archivos ‘.csv’ que contienen los siguientes parámetros: fecha y hora, frecuencia FI (*Frecuencia Intermedia*), frecuencia RF (*Radio Frecuencia*), desfase de FI y nivel de señal RSL (*Radio Signal Level*). Esta fuente proporciona los datos cada, aproximadamente, 15 segundos; es decir, la periodicidad de recogida de datos varía entre 14, 15 y 16 segundos.

Un ejemplo de esta base de datos es la Figura 20, donde se muestra una parte de los datos recogidos para el mes de noviembre de 2020.



	A	B	C	D	E	F	G
1	Time	FreqFI	FreqRF	DesFI	RSL		
2	2020-11-01 00:00:05	1402,412434	11152,41243	412,43	-43,93		
3	2020-11-01 00:00:21	1402,399936	11152,39994	399,94	-43,73		
4	2020-11-01 00:00:36	1402,237462	11152,23746	237,46	-44		
5	2020-11-01 00:00:52	1402,383272	11152,38327	383,27	-43,91		
6	2020-11-01 00:01:07	1402,70822	11152,70822	708,22	-44,03		
7	2020-11-01 00:01:23	1402,487422	11152,48742	487,42	-43,99		
8	2020-11-01 00:01:38	1402,074988	11152,07499	74,99	-43,89		
9	2020-11-01 00:01:53	1402,183304	11152,1833	183,3	-43,96		
10	2020-11-01 00:02:09	1402,570742	11152,57074	570,74	-43,94		
11	2020-11-01 00:02:24	1402,64573	11152,64573	645,73	-44,01		
12	2020-11-01 00:02:40	1402,637398	11152,6374	637,4	-44,01		
13	2020-11-01 00:02:55	1402,758212	11152,75821	758,21	-44,15		
14	2020-11-01 00:03:11	1402,129146	11152,12915	129,15	-43,84		
15	2020-11-01 00:03:26	1402,254126	11152,25413	254,13	-43,92		
16	2020-11-01 00:03:42	1402,483256	11152,48326	483,26	-44,02		
17	2020-11-01 00:03:57	1402,129146	11152,12915	129,15	-44,02		
18	2020-11-01 00:04:13	1402,879026	11152,87903	879,03	-43,8		
19	2020-11-01 00:04:28	1402,758212	11152,75821	758,21	-44,07		
20	2020-11-01 00:04:43	1402,162474	11152,16247	162,47	-43,95		
21	2020-11-01 00:04:59	1402,079154	11152,07915	79,15	-44		
22	2020-11-01 00:05:14	1402,920686	11152,92069	920,69	-43,97		
23	2020-11-01 00:05:30	1402,35411	11152,35411	354,11	-43,78		
24	2020-11-01 00:05:45	1402,258292	11152,25829	258,29	-43,88		
25	2020-11-01 00:06:01	1402,545746	11152,54575	545,75	-43,88		
26	2020-11-01 00:06:16	1402,258292	11152,25829	258,29	-43,9		
27	2020-11-01 00:06:32	1402,199968	11152,19997	199,97	-44,02		

Figura 20. Archivo de la base de datos RSL correspondiente al mes de noviembre de 2020.

CAPÍTULO 4.

PRE-PROCESADO Y PROCESADO DE LOS DATOS

En este capítulo describimos el proceso de transformación de los distintos tipos de datos con los que hemos trabajado, es decir, el pre-procesado de los datos para poder utilizarlos. Según el caso, por ejemplo, con los datos satelitales, describimos elementos del procesado. Se describirá la metodología empleada para la obtención de los mismos, tanto en las estaciones meteorológicas como en el radioenlace. Esta parte de nuestro trabajo es fundamental pues de la calidad de los datos dependen todos los elementos.

4.1 Tratamiento de los datos meteorológicos de VIMETRI

Una vez se han obtenido correctamente todos los datos de las diferentes fuentes, el siguiente paso es el pre-procesado, es decir, el tratamiento de dichos datos meteorológicos para que puedan ser usados para estudios e investigación.

En primer lugar, se realiza la lectura de los datos a partir de la base de datos ya generada. Seguidamente se realiza un estudio de los datos de precipitación con el fin de detectar los intervalos temporales en los que no hay datos. A esta ausencia de datos se la denomina GAP (en español, brecha), que pueden ser motivada por diferentes razones, las cuales se explicarán más adelante.

Para facilitar la lectura de los datos, tanto a nivel visual como a la hora de tratarlos en Matlab, éstos han sido transformados de un formato ‘.csv’ a un formato ‘.xlsx’.

Como resultado, se obtendrá un fichero de extensión ‘.mat’, con los datos correspondientes a cada mes debidamente procesados.

4.1.1 Tipos de GAPS

Entre los datos presentes en las bases se pueden observar dos tipos de GAPS diferentes:

- Ausencia de registro de un periodo de tiempo: este tipo de GAP se refiere a la ausencia de registro de un periodo de tiempo, que es variable, desde unos segundos hasta horas e incluso días. Para localizar con mayor facilidad en nuestro programa estos vacíos, se ha incluido un valor aberrante en dichas posiciones, en este caso el valor ‘-1’.
- Ausencia de datos: este tipo de GAP hace referencia a la ausencia de datos en un determinado instante. Para facilitar su distinción se incluye el valor ‘-1’ en estas posiciones, valor aberrante.

El valor aberrante que hemos es una precipitación nunca se dará. Hemos de resaltar que los gaps son muy inusuales y aparecen de forma aislada, por lo que un dato esporádicamente perdido no tiene efecto pernicioso en etapas posteriores.

4.1.2 Posibles causas de aparición de GAPS

A continuación, se exponen los diferentes factores que pueden influir en la aparición de los GAPS anteriormente mencionados:

- 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.

4.2 Tratamiento de los datos de radioenlaces de VIMETRI

Los datos de niveles de potencia provenientes de los radioenlaces vienen dados en dBm. A estos datos se les debe realizar una serie de procesos, como son: lectura de los datos de la base de datos, localización de los GAPS, transformación del periodo de muestreo de la señal, eliminación de bajas frecuencias y estudio de sincronismo. A continuación, se explican con más detalle los primeros elementos, y en el apartado 4.3 explicamos cómo trabajamos el sincronismo entre datos meteorológicos y satelitales.

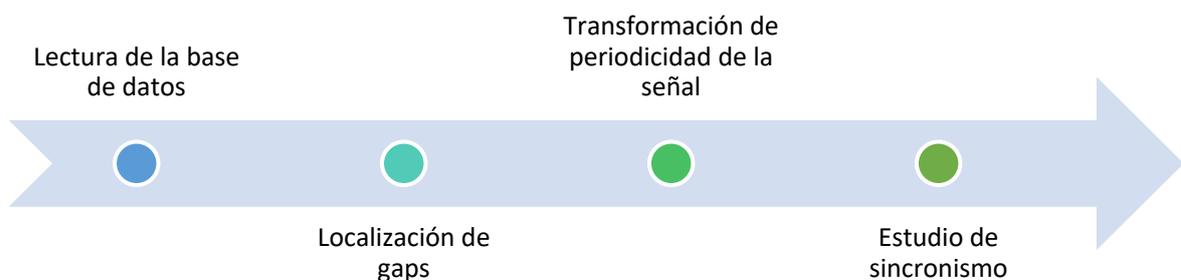


Figura 21. Etapas del pre-procesado de los datos de radioenlaces.

4.2.1 Lectura de la base de datos

La base de datos de los radioenlaces ha sido organizada manualmente generando un archivo ‘.xlsx’ correspondiente a cada mes, en nuestro caso disponemos de 16 meses, de agosto de 2019 a noviembre de 2020. Como resultado de esta etapa obtenemos una señal cuyos valores representan niveles de potencia (RSL) dado en dBm, y ordenamos temporalmente.

4.2.2 Localización de GAPS

Esta segunda etapa del tratamiento de datos de radioenlaces se corresponde con el estudio de la continuidad temporal de los archivos. Se emplea la misma metodología que la utilizada anteriormente para los datos de las estaciones meteorológicas. En este caso, se ha incluido un valor aberrante en dichas posiciones, en este caso el valor ‘0’. Es decir, cuando haya ausencia de datos se insertará el valor ‘0’ en dicha posición. Este se considera un valor imposible dado que un nivel de potencia del radioenlace nunca será igual a 0 dBm. Al igual que con los datos meteorológicos, ocurre que los gaps aquí también son muy inusuales y aparecen de forma aislada, por lo que un dato esporádicamente perdido no tiene efecto pernicioso en etapas posteriores.

4.2.2.1 Tipos de GAPS

Entre los datos presentes en la base se puede ver un solo tipo de GAP. Debido a que en este caso los datos de fecha y niveles de potencia del radioenlace vienen unidos, se entiende por GAP como la ausencia de datos, tanto de fecha como de nivel RSL. Para localizar con mayor facilidad esta ausencia de datos, se han incluido en los instantes donde hay un gap un valor de señal determinado en las posiciones de estos GAPS. Con esto se consigue que coincida el número de datos con los que deberían generarse teóricamente en un día, esto es, no se pierde la periodicidad vista como intervalo entre muestra y muestra, y sincronismo entre la información proveniente de las dos fuentes principales. El valor ficticio de señal

RSL que se inserta en este caso es el '0', considerado valor aberrante, ya que este nivel de potencia nunca será igual a cero.

4.2.2.2 Posibles causas de aparición de GAPs

Por norma general, se considera que el radioenlace no se detiene en ningún momento, por lo que cualquier ausencia de datos se supondrá producida debido a un fallo en el sistema de recolección de los datos. Este hecho se considera muy inusual.

4.2.3 Transformación del periodo de muestreo de la señal

Como ya se comentó anteriormente, los datos proporcionados por las diferentes fuentes poseen se muestrean con distinta periodicidad: estación meteorológica en el pabellón B cada quince segundos, estación meteorológica en El Zurbarán cada diez minutos y radioenlace satelital cada, aproximadamente, quince segundos.

Debido a esto, se generan archivos con un número de datos significativamente diferentes. La solución ha sido crear un programa de sincronismo, cuya función es sincronizar las muestras procedentes de las diferentes fuentes para que coincidan en número, permitiendo realizar, en procesos posteriores, la correcta comparación de los datos. En este caso se ha optado por modificar la periodicidad a una precisión de un minuto, dado que este valor entra dentro los límites razonables para la detección de un periodo de lluvia, una lluvia de duración inferior a un minuto se considera muy poco significativa. Además, podemos hacer un procesado minutil de los valores obtenidos en el Prototipo VIMETRI-MAC, y luego transformarlos a diez minutales para comparar con la estación de Zurbarán, en caso de que se desee.

Cabe destacar que, en general, la pérdida de datos de la estación de Zurbarán puede ir de minutos a horas, incluso días. Por este motivo, en términos de fiabilidad en la recogida de datos, consideramos que el sistema VIMETRI-MAC es mucho más fiable.

4.2.4 Eliminación de bajas frecuencias de la señal

En esta etapa se procesa la señal para eliminar los saltos de potencia y la baja frecuencia. Esto se debe a que se pueden encontrar atenuaciones sobre las que se detecten picos de lluvia. Dicho de otra manera, las variaciones debidas a las atenuaciones se manifiestan como componente de alta frecuencia sobre pedestales o sobre componentes de baja, por ejemplo, debidos a variaciones día-noche.

El procesado se ha realizado empleando un umbral adaptativo. Esta función utiliza una ventana, cuya longitud es prefijada. Cada ventana se toma a partir de una muestra, se cogen x datos anteriores y realiza un promedio, guardando el resultado en la misma posición en la que estaba centrado.

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 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 el 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.

En la Figura 22 vemos un ejemplo de señal RSL sobre la que hemos aplicado una apertura y un cierre a lo largo de este intervalo. En el eje vertical se representan los niveles de potencia dados en dBm y en el eje horizontal las muestras (4 muestras corresponden a un minuto, una cada 15 segundos).

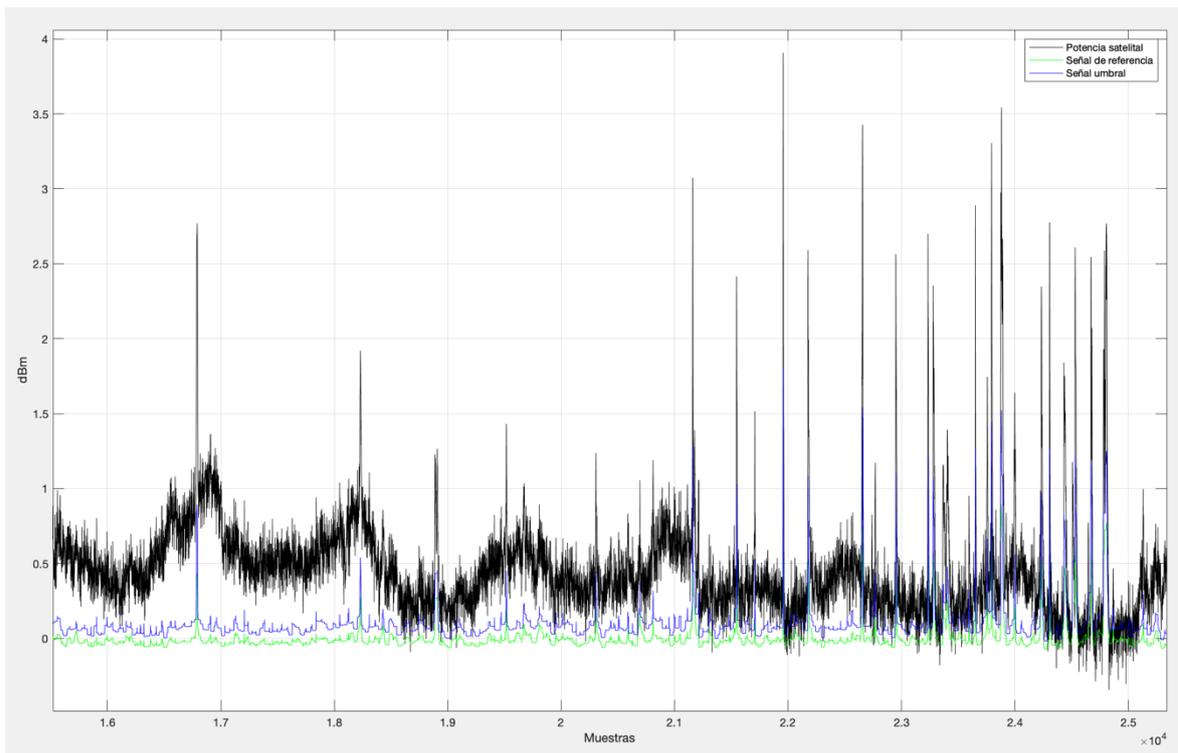


Figura 22. Comparativa de señal la señal original con la señal de referencia y la señal umbral.

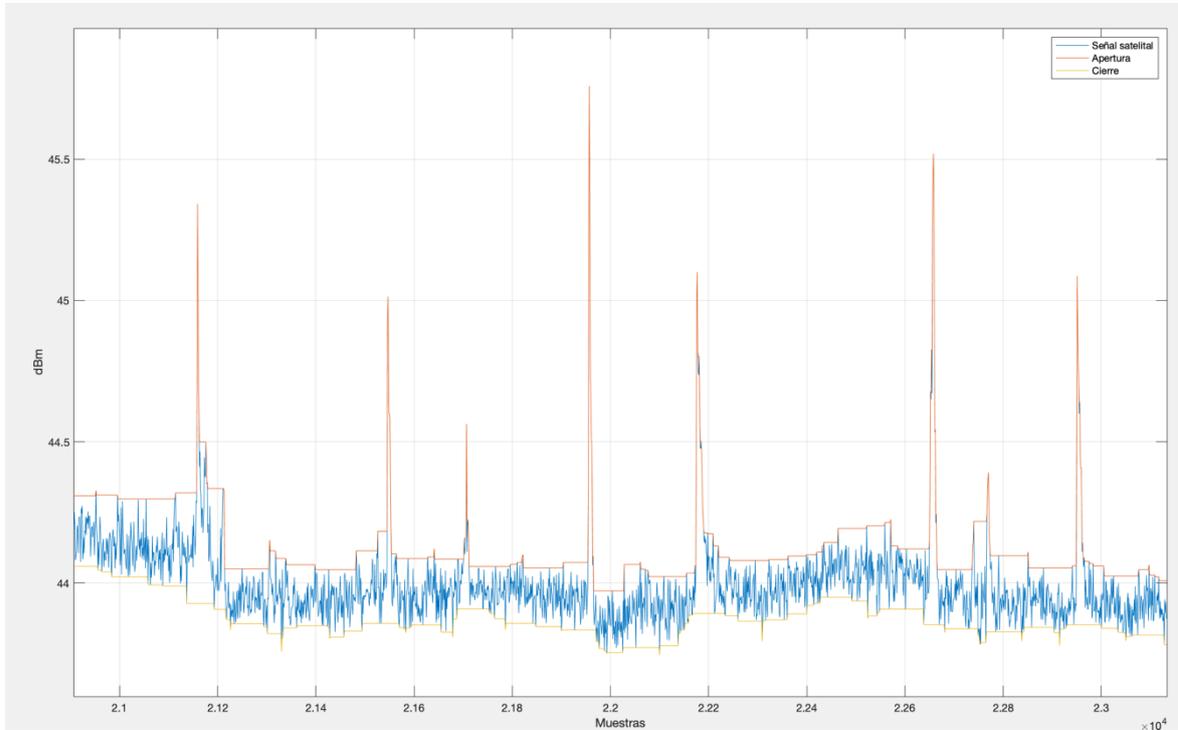


Figura 23. Señal RSL a la que se le ha aplicado una apertura y un cierre morfológico.

4.3 Sincronismo entre señales de radioenlaces y de estaciones meteorológicas

El proceso de sincronismo entre las diferentes bases de datos es muy importante. Los datos procedentes de las estaciones meteorológicas serán utilizados como referencia para buscar indicadores de precipitaciones.

Es importante además que se tengan en cuenta y se corrijan los saltos producidos por ausencia de datos producidos tanto en las estaciones meteorológicas como en el radioenlace. Estas interrupciones temporales son una importante consideración en el estudio del sincronismo entre las señales de radioenlaces y de meteorología, ya que estas no pueden ser ignoradas o de lo contrario no se podría mantener la base temporal.

Por otro lado, sobre la señal de alta frecuencia, se ha procedido a transformar los datos del radioenlace de negativo a positivo, facilitando así muchos cálculos.

4.4 Procesado de los datos meteorológicos y de radioenlaces

En las figuras siguientes se pueden ver representados los registros de los valores de lluvia durante la última parte del mes de noviembre de 2020, concretamente del 25 al 30 de noviembre. Es fácilmente detectable a nivel visual dónde han tenido lugar los eventos de lluvia registrados tanto por el pluviómetro como por el radioenlace satelital. Además, se puede ver que estos eventos coinciden en el eje temporal, es decir, las atenuaciones en la potencia del radioenlace satelital ocurren en el momento en que se ha registrado niveles de lluvia en el pluviómetro.

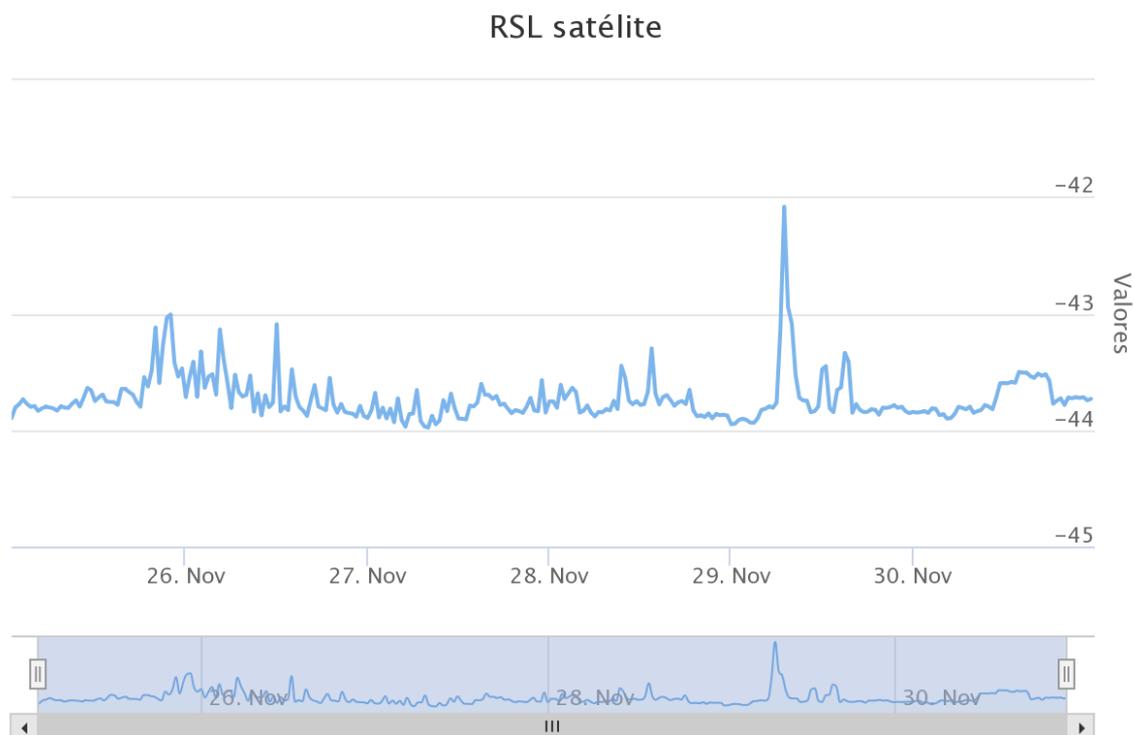


Figura 24. RSL satélite: 25 a 30 de noviembre de 2020.

Datos de precipitación medidos con pluviómetro

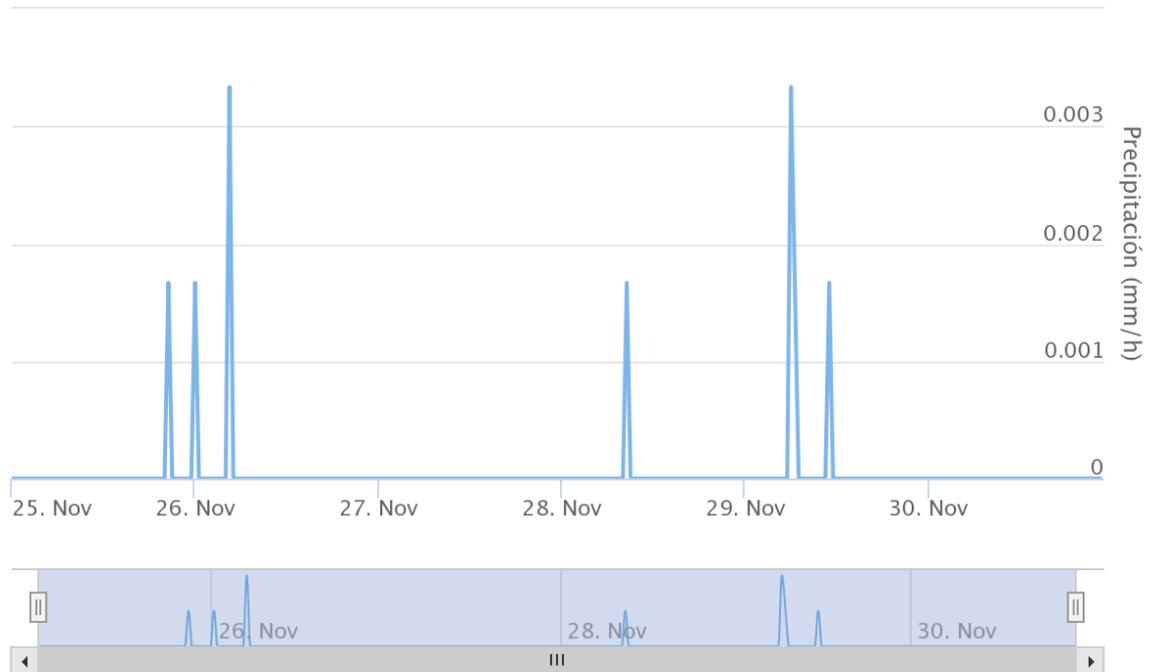


Figura 25. Datos de precipitación medidos por el pluviómetro: 25 a 30 de noviembre de 2020.

Estas figuras anteriores son una representación gráfica de los datos de los que parte este procesado. Una vez descargados los datos correspondientes a los meses con los que se pretende trabajar, y realizado en sincronismo entre ambas fuentes, como se explica en el apartado anterior, comienza el procesado de datos meteorológicos y satelitales.

CAPÍTULO 5.

ASPECTOS TEÓRICOS SOBRE ALGORITMOS DE DETECCIÓN DE PRECIPITACIÓN

5.1 Introducción

La primera cuestión filosófica que se debería plantear al hablar de inteligencia artificial es la siguiente: “¿Es posible que una máquina adquiriera una capacidad de abstracción similar a la inteligencia humana?”. Para responder a esta pregunta se ha de tener en cuenta que el cerebro del ser humano es considerado el sistema de reconocimiento de patrones más complejo y eficiente conocido hasta el momento. La neurociencia actual se basa en el hecho de que en la gran parte de los procesos cognitivos existe una separación de escalas entre la dinámica a nivel neuronal y la aparición de actividad mental abstracta. Esta separación supone la ruptura del vínculo existente entre el hardware (neuronas) y el software de nuestro cerebro (operaciones abstractas, estados mentales), y constituye la hipótesis de partida para que los símbolos abstractos puedan ser manipulados por sistemas artificiales que no requieran un substrato fisiológico natural. La posibilidad de manipular expresiones lógicas y esquemas abstractos mediante sistemas artificiales es la que permite la existencia de lo que se conoce como ‘inteligencia artificial’.

Al analizar el funcionamiento del cerebro humano, otro aspecto importante a tener en cuenta es la función que desempeña la experiencia y el aprendizaje. El cerebro humano

actual no solo es resultado de una evolución biológica basada en alteraciones genéricas, sino también del conjunto de técnicas y conocimientos que la humanidad ha ido acumulando con el tiempo, como pueden ser la cultura o el lenguaje. A nivel biológico, existen múltiples mecanismos que permite la existencia de procesos de activación neuronal dependientes de la experiencia previa y del entrenamiento. El mecanismo principal es conocido como ‘plasticidad sináptica’, un fenómeno por el que las conexiones sinápticas entre neuronas modulan su intensidad en función de la actividad que hayan experimentado previamente. De esta manera, cuantas más veces se active un cierto canal de activación neuronal, más fácil resultará activarlo en el futuro e integrarlo a nuevos procesos cognitivos de mayor complejidad [22].

La plasticidad neuronal es la base de la mayoría de los procesos de aprendizaje y memoria. A este paradigma se le conoce como ‘aprendizaje reforzado’, ya que la actividad sináptica se refuerza en función del número de veces que se establece una conexión entre neuronas. Esta regla se le conoce como ‘regla de Hebb’ por los trabajos del neuropsicólogo canadiense Donald O. Hebb. Algunas técnicas de inteligencia artificial como los métodos de aprendizaje supervisado también se basan en reglas similares que permiten modificar de forma adaptativa la forma en que el sistema artificial procesa la información [23].

La inteligencia artificial (IA) es una disciplina académica relacionada con la teoría de la computación cuyo objetivo es emular algunas de las facultades intelectuales humanas en sistemas artificiales. Con inteligencia humana nos referimos típicamente a procesos de percepción sensorial (visión, audición, etc.) y a sus consiguientes procesos de reconocimiento de patrones, por lo que las aplicaciones más habituales de la IA son el tratamiento de datos y la identificación de sistemas. Eso no excluye que la IA, desde sus inicios en la década del 1960, haya resuelto problemas de carácter más abstracto como la demostración de teoremas matemáticos, la adquisición del lenguaje, el jugar a ajedrez o la traducción automática de textos. El diseño de un sistema de inteligencia artificial normalmente requiere la utilización de herramientas de disciplinas muy diferentes como el cálculo numérico, la estadística, la informática, el procesado de señales, la electrónica sensorial, el control automático, la robótica o la neurociencia. Por este motivo, pese a que la

inteligencia artificial se considera una rama de la informática teórica, es una disciplina en la que contribuyen de forma activa numerosos científicos, ingenieros y matemáticos. En algunos aspectos, además, se beneficia de investigaciones en áreas tan diversas como la psicología, la sociología o la filosofía.

5.1.1 Breve historia de la inteligencia artificial

El nacimiento de la IA (*Inteligencia Artificial*) como disciplina de investigación tiene lugar en 1956, durante una conferencia sobre informática teórica celebrada en el Dartmouth College (Estados Unidos) [20]. Entre los asistentes de esta conferencia se encontraban John McCarthy, Marvin Minsky, Allen Newell y Herbert Simon. Estos dos últimos nombrados, A. Newell y H. Simon, presentaron un trabajo sobre demostración automática de teoremas al que denominaron *Logic Theorist*. El *Logic Theorist* fue el primer programa de ordenador que emulaba características propias del cerebro humano, por lo que es considerado el primer sistema de inteligencia artificial de la historia. El sistema era capaz de demostrar gran parte de los teoremas sobre lógica matemática que se presentaban en los tres volúmenes de los Principia Mathematica de Alfred N. Whitehead y Bertrand Russell (1910-1913). Minsky y McCarthy fundaron más tarde el laboratorio de inteligencia artificial del *Massachusetts Institute of Technology* (MIT), uno de los grupos pioneros en el ámbito. La actividad de los años cincuenta es consecuencia de trabajos teóricos de investigadores anteriores como Charles Babbage (autor de la Máquina analítica, 1842), Kurt Gödel (teorema de incompletitud, 1930), Alan Turing (máquina universal, 1936), Norbert Wiener (cibernética, 1943) y John von Neumann (arquitectura del computador, 1950). La arquitectura de von Neumann consta de una unidad central de proceso (CPU - *Central Processing Unit*) y de un sistema de almacenamiento de datos (memoria), y fue utilizada en 1954 por *RAND Corporation* para construir JOHNIAC (*John v. Neumann Numerical Integrator and Automatic Computer*), una de las primeras computadoras en las que más tarde se implementaron sistemas de inteligencia artificial como el *Logic Theorist* de Newell y Simon. En 1954 también apareció el IBM (*International Business Machines Corporation*) 704, la

primera computadora de producción en cadena, y con ella se desarrollaron numerosos lenguajes de programación específicamente diseñados para implementar sistemas de inteligencia artificial como el LISP (*LISt Processor*). Junto con estos avances, se produjeron los primeros intentos para determinar la presencia de comportamiento inteligente en una máquina. El más relevante desde el punto de vista histórico fue propuesto por Alan Turing en un artículo de 1950 publicado en la revista *Mind* y titulado *Computing Machinery and Intelligence*.

En este trabajo se propone un test de inteligencia para máquinas según el cual una máquina presentaría un comportamiento inteligente en la medida en que fuese capaz de mantener una conversación con un humano sin que otra persona pueda distinguir quién es el humano y quién el ordenador. Aunque el test de Turing ha sufrido innumerables adaptaciones, correcciones y controversias, pone de manifiesto los primeros intentos de alcanzar una definición objetiva de la inteligencia.

En este contexto, es de especial relevancia el Teorema de incompletitud de Gödel de 1931, un conjunto de teoremas de lógica matemática que establecen las limitaciones inherentes a un sistema basado en reglas y procedimientos lógicos (como lo son todos los sistemas de IA). Tras los primeros trabajos en IA de los años cincuenta, en la década de los sesenta se produjo un gran esfuerzo de formalización matemática de los métodos utilizados por los sistemas de IA. Los años setenta, en parte como respuesta al test de Turing, se produjo el nacimiento de un área conocida como procesamiento del lenguaje natural (NLP, *Natural Language Processing*), una disciplina dedicada a sistemas artificiales capaces de generar frases inteligentes y de mantener conversaciones con humanos. El NLP ha dado lugar a diversas áreas de investigación en el campo de la lingüística computacional, incluyendo aspectos como la desambiguación semántica o la comunicación con datos incompletos o erróneos.

En los años ochenta empezaron a desarrollarse las primeras aplicaciones comerciales de la IA, fundamentalmente dirigidas a problemas de producción, control de procesos o contabilidad. Con estas aplicaciones aparecieron los primeros sistemas expertos, que

permitían realizar tareas de diagnóstico y toma de decisiones a partir de información aportada por profesionales expertos.

En torno a 1990, IBM construyó el ordenador ajedrecista Deep Blue, capaz de plantarle cara a un gran maestro de ajedrez utilizando algoritmos de búsqueda y análisis que le permitían valorar cientos de miles de posiciones por segundo. Más allá del intento de diseñar robots humanoides y sistemas que rivalicen con el cerebro humano en funcionalidad y rendimiento, el interés hoy en día es diseñar e implementar sistemas que permitan analizar grandes cantidades de datos de forma rápida y eficiente.

A pesar de los grandes avances en este ámbito, sigue sin existir una máquina que pueda pasar el test de Turing tal y como se planteó en el artículo original. Esto no es tanto debido a un fracaso de la IA como a que los intereses del área se han ido redefiniendo a lo largo de la historia.

Actualmente, se ha consolidado como una disciplina que permite diseñar aplicaciones de gran utilidad práctica en numerosos campos. Existe una enorme lista de ámbitos de conocimiento en los que se utilizan sistemas de IA, entre los que son de especial relevancia la minería de datos, el diagnóstico médico, la robótica, la visión artificial, el análisis de datos bursátiles o la planificación y logística. El campo de la observación meteorológica en que se enmarca este TFG es otro ejemplo, sí como la predicción aplicada al mismo campo.

5.2 Introducción al aprendizaje máquina

El aprendizaje máquina pretende 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.

Dicha búsqueda se lleva a cabo a partir de pasos individuales los cuales son llevados a cabo por operadores de búsqueda que han sido especificados previamente por el programador.

El objetivo de este trabajo no es reclasificar objetos cuyas clases ya son conocidas, lo que realmente se pretende es etiquetar ejemplos futuros, aquellos cuyas clases aún se ignoran.

Por ello no se puede 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.

Tabla 8. Ejemplos disponibles.

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 solución para evitar problemas consiste en 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, se ha de resaltar el hecho de que las tasas de acierto son medidas de calidad del clasificador y hay variantes que se utilizarán convenientemente. Por ejemplo, si lo importante es la exactitud, también lo es la sensibilidad a la presencia de lluvia.

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 [24].

5.2.1 Formas de aprendizaje

Existen dos tipos comunes de aprendizaje. 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 adecuada para aplicar a nuestro proyecto.

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 [25].

En nuestro caso no se ha podido realizar un entrenamiento completamente supervisado porque no disponemos de estaciones meteorológicas a lo largo de toda la trayectoria que cubre el radioenlace utilizado, por lo que el aprendizaje llevado a cabo ha sido el “aprendizaje semi-supervisado”.

A diferencia de los dos enfoques más tradicionales (entrenamiento supervisado y entrenamiento no supervisado), los algoritmos semi-supervisados emplean pocos datos etiquetados y muchos datos no etiquetados como parte del conjunto de entrenamiento. Dichos algoritmos tratan de explorar la información estructural que contienen los datos no etiquetados con el objetivo de generar modelos predictivos que funcionen mejor que los que sólo utilizan datos etiquetados [26].

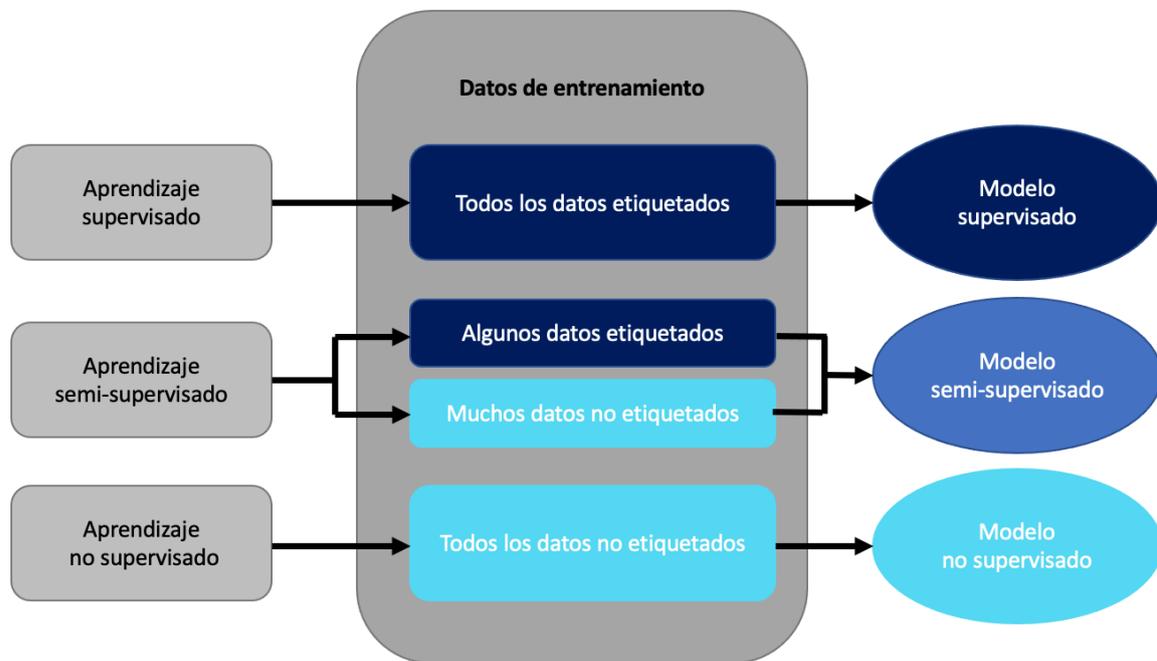


Figura 26. Esquema de los distintos enfoques de aprendizaje automático.

Los modelos de aprendizaje semi-supervisados son cada vez más utilizados hoy en día. Un ejemplo clásico en el que se observa el valor que aportan dichos modelos es el análisis de las conversaciones grabadas en un *call center*. Con el objetivo de inferir automáticamente características de los interlocutores (género, edad, geografía, etc.) o los motivos de la llamada (error en la factura, nivel de servicio, problemas de calidad, etc.), entre otros, es necesario disponer de un volumen elevado de casos ya etiquetados sobre los cuales aprender los patrones de cada tipología de llamada. El etiquetado de estos casos es una tarea ardua de conseguir, ya que etiquetar archivos de audio, por lo general, es una tarea que requiere tiempo y mucha intervención humana. En estas situaciones en las que el etiquetado de los casos es escaso, ya sea porque es costoso, necesita mucho tiempo de recopilación, requiere de mucha intervención humana o simplemente porque es completamente desconocido, los algoritmos de aprendizaje semi-supervisado son de gran utilidad gracias a sus características de funcionamiento. Sin embargo, no todos los problemas pueden abordarse directamente con

estas técnicas, ya que existen algunas características esenciales que deben estar presentes en los problemas para poder solucionarlos, de manera efectiva, utilizando esta tipología de algoritmos.

Probablemente la primera aproximación sobre el uso de datos no etiquetados para construir un modelo de clasificación es el método *Self-Learning*. *Self-learning* es un método *wrapper* muy simple y uno de los métodos más utilizados en la práctica. La primera fase de dicho algoritmo consiste en aprender un clasificador con los pocos datos etiquetados. Posteriormente, el clasificador es usado para predecir los datos no etiquetados y sus predicciones de mayor fiabilidad son añadidas al conjunto de entrenamiento. Finalmente, el clasificador es reentrenado con el nuevo conjunto de entrenamiento. Este proceso (Figura 27) se repite hasta que no se puedan añadir nuevos datos al conjunto de entrenamiento [26].

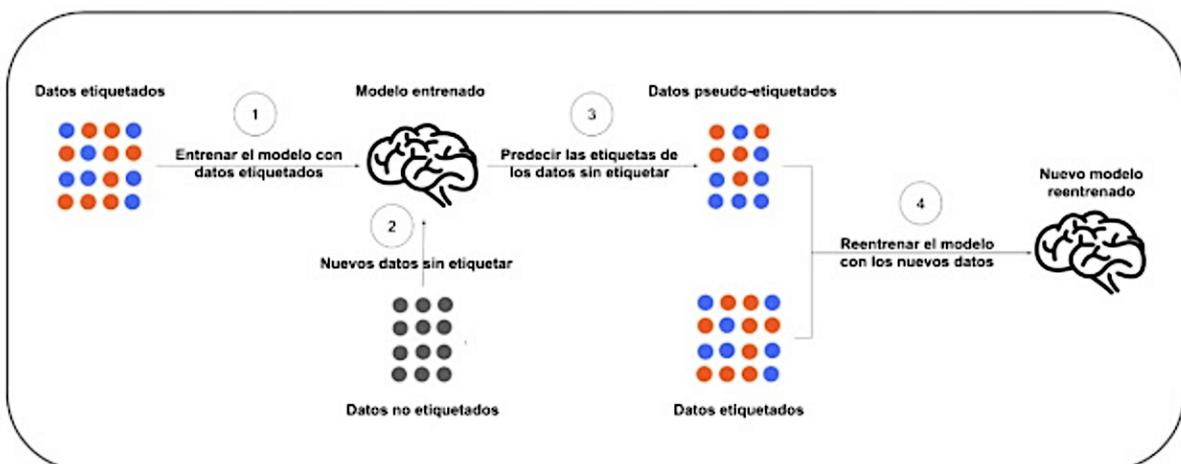


Figura 27. Esquema del funcionamiento del algoritmo Self-Learning.

Nota: Adaptado de <https://empresas.blogthinkbig.com/semi-supervised-learning-el-gran-desconocido/> [26].

En el enfoque semi-supervisado se asume cierta estructura en la distribución subyacente de los datos, es decir, los datos más próximos entre sí se supone que tienen la misma etiqueta. La Figura 28 refleja cómo los algoritmos semi-supervisados van ajustando, iteración tras iteración, la frontera de decisión entre las etiquetas. Si solo se dispone de datos etiquetados

la frontera de decisión es muy distinta a la frontera aprendida al incorporar la información de la estructura subyacente de todos los datos no etiquetados.

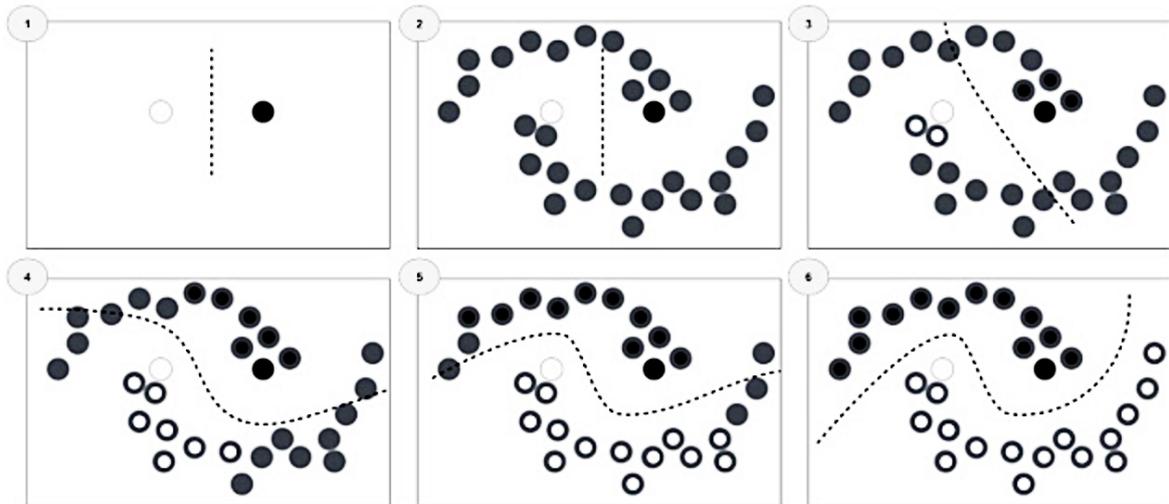


Figura 28. Ajuste iterativo de la frontera de decisión en el aprendizaje semi-supervisado.

Nota: Adaptado de <https://empresas.blogthinkbig.com/semi-supervised-learning-el-gran-desconocido/> [26].

Otra situación en la que los datos semi-etiquetados son de utilidad es en la detección de anomalías, ya que se trata de un problema típico en el que es difícil disponer de una gran cantidad de datos etiquetados. Este tipo de problema puede abordarse con un enfoque no supervisado. El objetivo de dicho enfoque es el de identificar, en base a las características de los datos, aquellos casos que difieren mucho del patrón habitual de comportamiento. En este contexto, el subconjunto de datos etiquetados puede ayudar a evaluar las distintas iteraciones del algoritmo, y así, guiar la búsqueda de los parámetros óptimos del algoritmo analizado.

Finalmente, con los ejemplos anteriormente expuestos, queda demostrado que el uso de datos no etiquetados junto con una pequeña cantidad de datos etiquetados puede mejorar considerablemente la precisión tanto de los modelos supervisados como la de los no supervisados.

En nuestro proyecto, esto se verá reflejado en la obtención de resultados, donde la sensibilidad no será la máxima, que sí podría obtenerse al disponer de más estaciones meteorológicas.

5.2.2 Bases de datos balanceadas 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 este TFG, debido a que se dispone de muchos instantes de muestreo, pero muy pocos de ellos corresponden a lluvia, que es el fenómeno que se pretende detectar.

Lo descrito en el párrafo anterior presenta varios problemas:

- Las clases minoritarias pueden quedar mal representadas en los experimentos.
- Los datos de calidad (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 [27]. 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 [28].

Centrándonos en este TFG, el balanceo de las bases de datos se realiza mediante una estrategia de muestreo mediante submuestreo de la clase mayoritaria (la de no lluvia). En la siguiente pieza de código se muestra cómo se ha hecho. Básicamente, la idea es tomar como referencia los instantes de lluvia y tomar intervalos de no lluvia anteriores generando una secuencia continua hasta varios minutos después de acabado el episodio de lluvia. De esta forma recogemos con realismo las secuencias no lluvia-lluvia-no lluvia.

```

% Balanceo de datos
[Iout,Jout] = size(Marcs_LSTM2);
for aa=1:Jout % Forzamos secuencias de longitud múltiplo entero de sequenceLength
    difSeq = Marcs_LSTM2(2,aa) - Marcs_LSTM2(1,aa) + 1;
    if (difSeq<sequenceLength)
        Marcs_LSTM2(1,aa) = Marcs_LSTM2(1,aa) - sequenceLength-1;
        Marcs_LSTM2(2,aa) = Marcs_LSTM2(2,aa) + (sequenceLength-difSeq);
    elseif ((difSeq>sequenceLength))
        Marcs_LSTM2(1,aa) = Marcs_LSTM2(1,aa) - sequenceLength-1;
        %      coci      = fix(difSeq,sequenceLength);
        coci      = fix(difSeq/sequenceLength);
        resto      = rem(difSeq,sequenceLength);
        Marcs_LSTM2(2,aa) = Marcs_LSTM2(1,aa) + (sequenceLength + difSeq) + (coci-1)*sequenceLength-resto;
    end
end

```

Figura 29. Balanceo de datos.

El proceso que se ha llevado a cabo para capturar los datos con que se entrenará o testeará ha sido establecer una ventana de 180 minutos (3 horas). Este parámetro se introduce en el código con el nombre `sequenceLength = 180`. Cada ventana queda separada de la siguiente por un 25% de las muestras (minutos), es decir, cada ventana se encontrará separada de las que le rodean por 45 minutos. El objetivo de esta forma de tomar los datos es organizar la obtención de los eventos de lluvia y no-lluvia ordenados secuencialmente, a la vez que se hace de forma lo más balanceada posible.

5.3 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) [29].

5.3.1 Perceptrón multicapa

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 30 se puede observar un esquema sencillo de una neurona artificial.

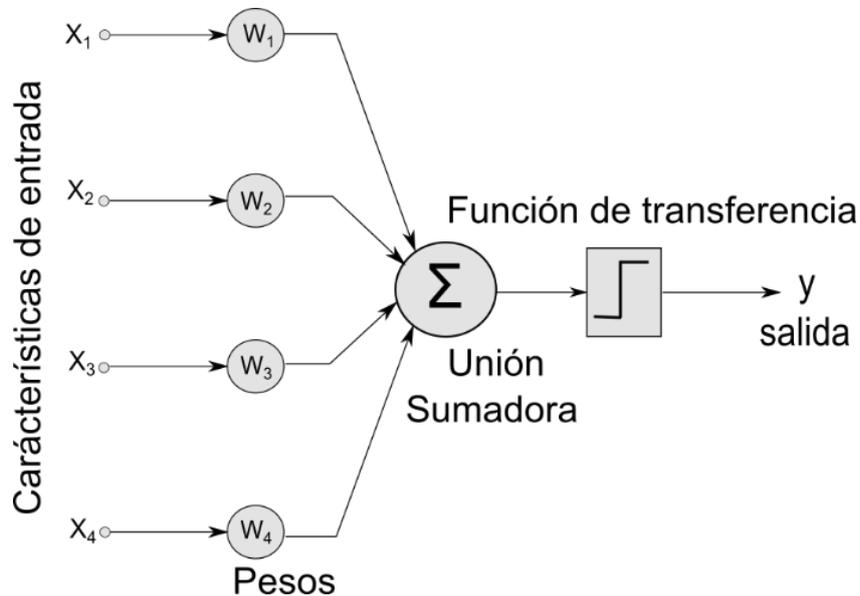


Figura 30. Esquema básico de una neurona artificial con cuatro características de entrada.
 Nota: Adaptado de "Técnicas de aprendizaje máquina para detección y cuantificación de lluvia mediante radioenlaces terrenos a frecuencia de gigahercios" por S. Fuente González, 2020.

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 [29]. 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 provienen 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 delante [29].

Como se puede observar en la Figura 31 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.

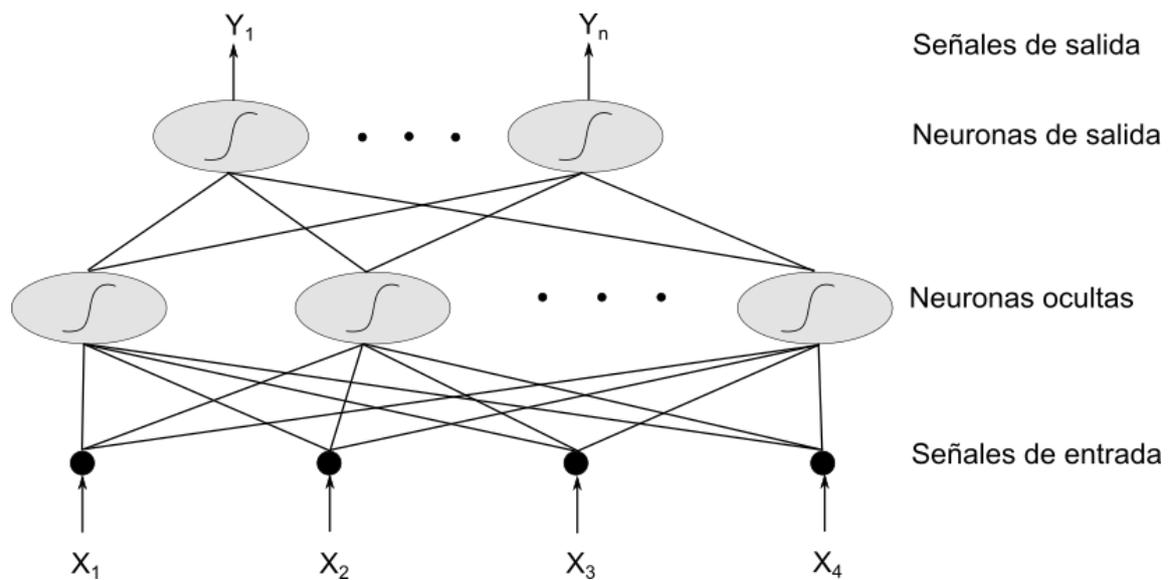


Figura 31. Representación gráfica de la estructura de un perceptrón multicapa de dos niveles.
 Nota: Adaptado de "Técnicas de aprendizaje máquina para detección y cuantificación de lluvia mediante radioenlaces terrenos a frecuencia de gigahercios" por S. Fuente González, 2020.

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.

- **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.

5.3.2 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* [29].

5.3.3 Redes Neuronales Recurrentes

Las redes neuronales recurrentes (RNN) fueron concebidas en la década de 1980. Estas redes han sido muy difíciles de entrenar por sus requerimientos en computación y hasta la

llegada de los avances de estos últimos años, no se había conseguido que fueran más accesibles y popularizar su uso por la industria.

Una red RNN tiene una característica que la diferencia de las otras redes, y es que incluye conexiones que apuntan “hacia atrás”, es decir, “*backpropagation*”.

Suponiendo la RNN más simple posible, compuesta por una sola neurona que recibe una entrada, produciendo una salida, y enviando esa salida a sí misma, como se muestra en la siguiente figura:

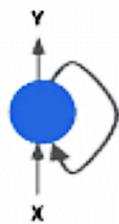


Figura 32. RNN de una neurona.

Nota: Adaptado de <https://torres.ai/redes-neuronales-recurrentes/> [30].

En cada instante de tiempo (también llamado *timestep* en este contexto), esta neurona recurrente recibe la entrada x de la capa anterior, así como su propia salida del instante de tiempo anterior para generar su salida y . Podemos representar visualmente esta pequeña red desplegada en el eje del tiempo como se muestra en la figura:

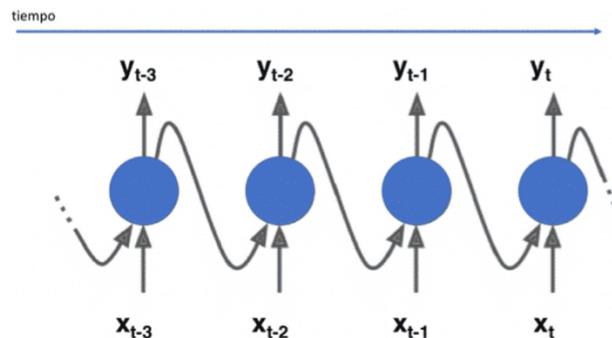


Figura 33. RNN desplegada en el eje del tiempo.

Nota: Adaptado de <https://torres.ai/redes-neuronales-recurrentes/> [30].

Siguiendo esta misma idea, una capa de neuronas recurrentes se puede implementar de tal manera que, en cada instante de tiempo, cada neurona recibe dos entradas, la entrada correspondiente de la capa anterior y a su vez la salida del instante anterior de la misma capa.

Ahora cada neurona recurrente tiene dos conjuntos de parámetros, uno que lo aplica a la entrada de datos que recibe de la capa anterior y otro conjunto que lo aplica a la entrada de datos correspondiente al vector salida del instante anterior. Sin entrar demasiado en formulación, y siguiendo la notación explicada en la primera parte del libro, podríamos expresarlo de la siguiente manera:

$$y_t = f(Wx_t + Uy_{t-1} + b) \quad (5.1)$$

Donde $x = (x_1, \dots, x_T)$ representa la secuencia de entrada proveniente de la capa anterior, W los pesos de la matriz y b la *bias* vistos ya en las anteriores capas. Las RNN extienden esta función con una conexión recurrente en el tiempo donde U es la matriz de pesos que opera sobre el estado de la red en el instante de tiempo anterior (y_{t-1}) anterior. Ahora, en la fase de entrenamiento a través del *backpropagation* también se actualizan los pesos de esta matriz [30].

5.3.4 Long-Short Term Memory

Las redes de memoria a corto y largo plazo, o *Long-Short Term Memory* (LSTM) son una extensión de las redes neuronales recurrentes, que, básicamente, amplían su memoria para aprender de experiencias importantes que han pasado hace mucho tiempo. Fueron introducidas por Hochreiter y Schmidhuber (1997), y varias personas lo han seguido perfeccionando en sus trabajos posteriores haciéndose bastante popular. Funcionan bastante bien en una gran variedad de problemas y son ampliamente utilizadas.

Las LSTM permiten a las RNN recordar sus entradas durante un largo período de tiempo. Esto se debe a que LSTM contiene su información en la memoria, que puede considerarse

similar a la memoria de un ordenador, en el sentido que una neurona de una LSTM puede leer, escribir y borrar información de su memoria.

Esta memoria se puede ver como una “celda” bloqueada, donde “bloqueada” significa que la célula decide si almacenar o eliminar información dentro (abriendo la puerta o no para almacenar), en función de la importancia que asigna a la información que está recibiendo. La asignación de importancia se decide a través de los pesos, que también se aprenden mediante el algoritmo. Esto se puede ver como que aprende con el tiempo qué información es importante y cuál no.

Los LSTM tienen la forma de una cadena de módulos repetidos, donde el módulo repetido consta de cuatro capas de red neuronal que interactúan de una manera muy especial.

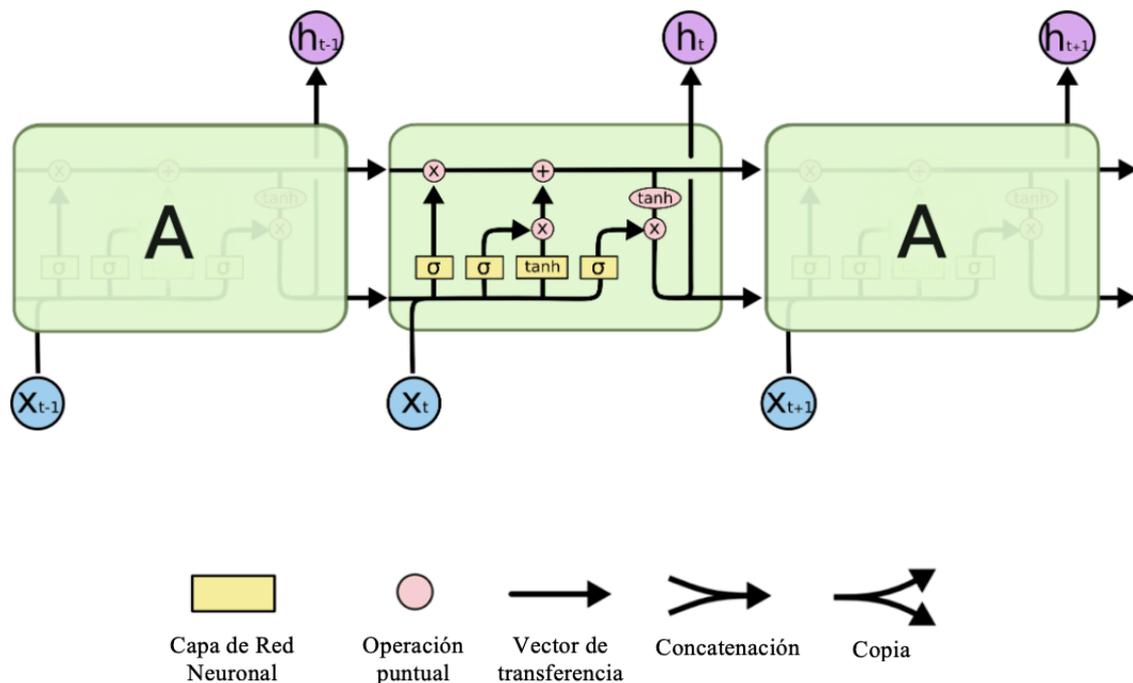


Figura 34. Diagrama de una LSTM.
 Nota: Adaptado de <http://colah.github.io/posts/2015-08-Understanding-LSTMs/> [31].

En el diagrama anterior, cada línea lleva un vector completo, desde la salida de un nodo hasta las entradas de otros. Los círculos rosas representan operaciones puntuales, como la suma de vectores, mientras que los cuadros amarillos son capas de redes neuronales aprendidas. Las líneas que se fusionan denotan concatenación, mientras que una bifurcación de líneas indica que su contenido se está copiando y las copias van a diferentes ubicaciones.

La clave de los LSTM es el estado de la celda, la línea horizontal que atraviesa la parte superior del diagrama. El estado de la celda es como una cinta transportadora. Corre directamente por toda la cadena, con solo algunas interacciones lineales menores. Es muy fácil que la información fluya sin cambios [31].

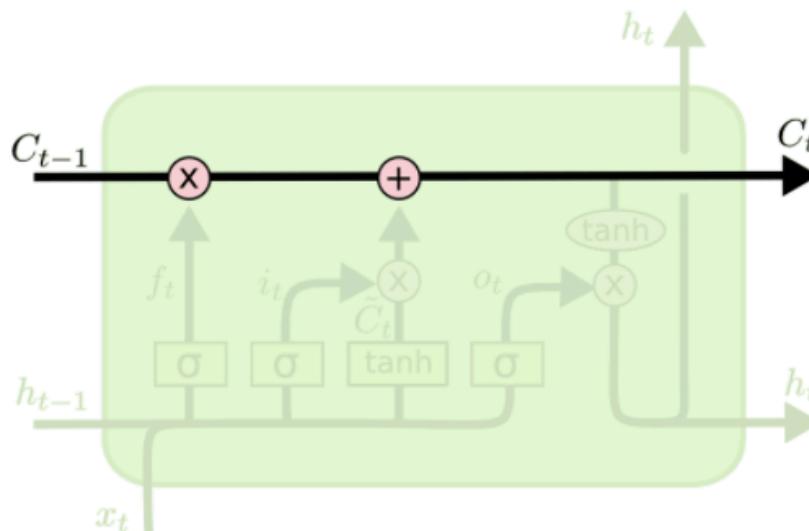


Figura 35. Diagrama de una LSTM – celda.

Nota: Adaptado de <http://colah.github.io/posts/2015-08-Understanding-LSTMs/> [31].

El LSTM tiene la capacidad de eliminar o agregar información al estado de la celda, regulada cuidadosamente por estructuras llamadas puertas.

Las puertas son una forma de dejar pasar información opcionalmente. Están compuestos por una capa de red neuronal sigmoidea y una operación de multiplicación puntual.

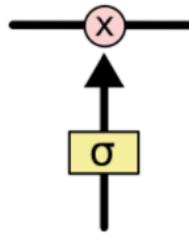


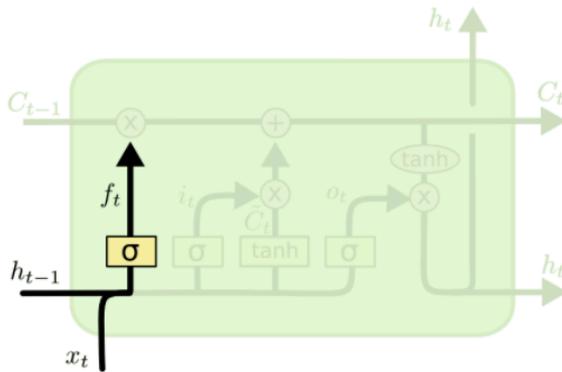
Figura 36. Diagrama de una LSTM - capa sigmoidea.
Nota: Adaptado de <http://colah.github.io/posts/2015-08-Understanding-LSTMs/> [31].

La capa sigmoidea genera números entre cero y uno, que describen cuánto de cada componente se debe dejar pasar. Un valor de cero significa "no dejar pasar nada", mientras que un valor de uno significa "dejar pasar todo".

En una neurona LSTM hay tres puertas a estas "celdas" de información: puerta de entrada (*input gate*), puerta de olvido (*forget gate*) y puerta de salida (*output gate*). Estas puertas determinan si se permite o no una nueva entrada, se elimina la información porque no es importante o se deja que afecte a la salida en el paso de tiempo actual [30].

5.3.4.1 LSTM paso a paso

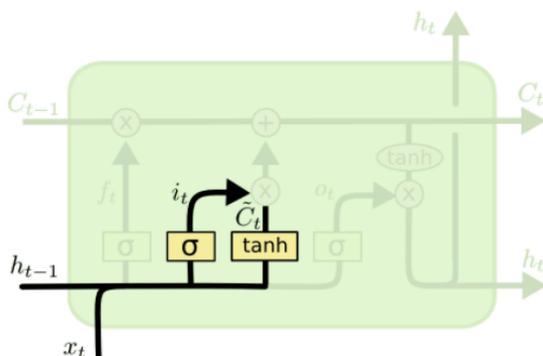
El primer paso en nuestro LSTM es decidir qué información vamos a eliminar del estado de la celda. Esta decisión la toma una capa sigmoidea llamada "capa de puerta de olvido". Observa h_{t-1} y x_t , y genera un número entre 0 y 1 para cada número en el estado de celda C_{t-1} . Un 1 representa "mantener esto completamente", mientras que un 0 representa "deshacerse de esto por completo".



$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

Figura 37. Funcionamiento de una LSTM - Paso 1.
 Nota: Adaptado de <http://colah.github.io/posts/2015-08-Understanding-LSTMs/> [31].

El siguiente paso es decidir qué nueva información vamos a almacenar en el estado de la celda. Esto tiene dos partes. Primero, una capa sigmoidea llamada "capa de puerta de entrada" decide qué valores actualizaremos. A continuación, una capa \tanh crea un vector de nuevos valores candidatos, \tilde{C}_t , que podrían agregarse al estado. En el siguiente paso, combinaremos estos dos para crear una actualización del estado [31].



$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

Figura 38. Funcionamiento de una LSTM - Paso 2.
 Nota: Adaptado de <http://colah.github.io/posts/2015-08-Understanding-LSTMs/> [31].

Ahora es el momento de actualizar el estado de la celda anterior, C_{t-1} , al nuevo estado de la celda C_t . Los pasos anteriores ya decidieron qué hacer, solo tenemos que hacerlo realmente.

Multiplicamos el estado anterior por f_t , olvidando las cosas que decidimos olvidar antes. Luego sumamos $i \cdot \tilde{C}_t$. Estos son los nuevos valores candidatos, escalados según cuánto decidimos actualizar cada valor de estado [31].

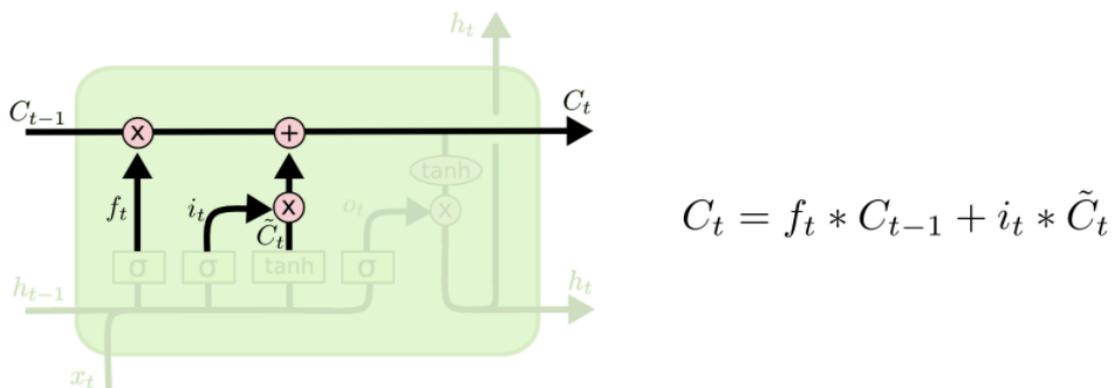


Figura 39. Funcionamiento de una LSTM - Paso 3.

Nota: Adaptado de <http://colah.github.io/posts/2015-08-Understanding-LSTMs/> [31].

Finalmente, necesitamos decidir qué vamos a producir. Esta salida se basará en el estado de nuestra celda, pero será una versión filtrada. Primero, ejecutamos una capa sigmoidea que decide qué partes del estado celular vamos a generar. Luego, colocamos el estado de la celda a través de \tanh (para presionar los valores para que estén entre -1 y 1) y lo multiplicamos por la salida de la puerta sigmoidea, de modo que solo produzcamos las partes que decidimos [31].

5.3.5 Cómo entrenar una red neuronal

Para que una red pueda ser utilizada debe ser entrenada. Partiendo de un conjunto de pesos sinápticos aleatorios, el proceso de aprendizaje busca un conjunto de pesos que permitan a la red desarrollar correctamente una determinada tarea. Durante el proceso de

aprendizaje se va refinando iterativamente la solución hasta alcanzar un nivel de operación suficientemente bueno. El proceso de aprendizaje se puede dividir, como ya se nombró anteriormente en el apartado 5.2, en dos grandes grupos de acuerdo con sus características: aprendizaje supervisado y aprendizaje no supervisado. En este caso estamos trabajando con una red de aprendizaje supervisado, es decir, se presenta a la red un conjunto de patrones de entrada junto con la salida esperada. Los pesos se van modificando de manera proporcional al error que se produce entre la salida real de la red y la salida esperada.

En el contexto de las redes neuronales el aprendizaje puede ser visto como el proceso de ajuste de los parámetros libres de la red. Partiendo de un conjunto de pesos sinápticos aleatorio, el proceso de aprendizaje busca un conjunto de pesos que permitan a la red desarrollar correctamente una determinada tarea. El proceso de aprendizaje es un proceso iterativo, en el cual se va refinando la solución hasta alcanzar un nivel de operación suficientemente bueno. La mayoría de los métodos de entrenamiento utilizados en las redes neuronales con conexión hacia delante consisten en proponer una función de error que mida el rendimiento actual de la red en función de los pesos sinápticos. El objetivo del método de entrenamiento es encontrar el conjunto de pesos sinápticos que minimizan (o maximizan) la función. El método de optimización proporciona una regla de actualización de los pesos que en función de los patrones de entrada modifica iterativamente los pesos hasta alcanzar el punto óptimo de la red neuronal [32].

CAPÍTULO 6.

ALGORITMO DE DETECCIÓN DE PRECIPITACIÓN COMBINANDO ESTACIONES METEOROLÓGICAS Y RADIOENLACES SATELITALES

En este capítulo se pretende describir paso a paso las distintas fases por las que han pasado los datos de las estaciones meteorológicas y el radioenlace satelital hasta llegar al paso final, donde se realizará el entrenamiento y test empleando las redes LSTM (*Long-Short Term Memory*).

6.1 Diagrama de flujo

A continuación, se muestran las distintas fases que constituyen el diagrama de flujo seguido por el algoritmo de detección.



Figura 40. Diagrama de flujo del algoritmo de detección.

6.2 Lectura de datos y sincronismo de los mismos

El primer paso que se ha llevado a cabo es el sincronismo entre los datos de las diferentes fuentes, es decir, entre los datos de potencia RSL (*Radio Signal Level*), obtenidos a partir del radioenlace satelital, y los datos de lluvia recogidos por la estación meteorológica.

Para ello, lo primero que se debe comprobar es si el archivo de sincronismo de un mes x se ha generado previamente para ambas fuentes de datos o si, por el contrario, se debe proceder a generarlo. En caso de que el archivo exista, se procederá al análisis del mismo.

Los archivos de datos de los que partimos se encuentran en formato ‘.xlsx’. Este proceso incluye todos los apartados descritos anteriormente en el Capítulo 4, donde tiene lugar la localización de gaps, la transformación del periodo de muestreo y la eliminación de las bajas frecuencias.

El siguiente paso es el sincronismo, ya nombrado en el apartado 4.3, cuyo objetivo es mantener una base temporal conjunta para los datos satelitales y los procedentes del pluviómetro. Las tablas de datos, previamente procesadas, serán recorridas con el fin de acumular los datos minuto a minuto. Se debe tener en cuenta, a la hora de realizar el algoritmo, la posibilidad de la presencia de GAPS, pudiendo ser de larga o corta duración.

Existen numerosas casuísticas cuando se encuentra una ausencia de datos, ante las cuales el código actúa de forma determinada según sea el caso. Se han encontrado casos en los que existe ausencia de un solo dato, lo cual no afecta en gran medida a los resultados. Pero también se han registrados ausencia de datos de varios minutos, siendo este caso más común, o incluso de horas o días, pudiendo deberse a múltiples factores (p.e., parada del sistema por cortes de luz).

Para ambas fuentes se lleva a cabo el mismo proceso. Mediante un bucle, se va recorriendo la columna de tiempo proporcionada por las tablas de datos. Se pretende ir acumulando los datos de minuto en minuto, por lo que se recorre la matriz de datos teniendo en cuenta los cambios de minutos, horas y días. En el caso de la existencia de gaps, estos pueden ser cortos o largos, y actuaremos de forma diferente ante cada uno de ellos. En el caso de un gap corto, siempre y cuando existan más datos en el mismo minuto, no afectará

a la acumulación de lluvia, ya que se realiza el promedio del restos de datos de los que sí se dispone. En el caso de gaps con mayor amplitud en el tiempo, tal que no permita obtener un valor promedio para un minuto dado, quedarán rellenados con un valor “NaN (*Not-a-Number*)”, con lo cual serán fácilmente reconocibles.

6.3 Detección de gaps

Una vez se han generado correctamente los archivos .mat correspondientes a los meses que se desean analizar, se procede la detección de los gaps.

Como ya se indicó en el apartado 4.2.2, estos gaps han sido previamente señalados con valores aberrantes para cada una de las fuentes, haciéndolos así fácilmente detectables. En el caso de la señal de potencia satelital, quedarán marcados con un valor de “0”, ya que la potencia de la señal nunca tomará este valor. Para el caso de la estación meteorológica, se ha empleado el valor de “-1”, dado que nunca se registrará un valor de lluvia igual a un valor negativo.

El proceso llevado a cabo consiste en localizar las posiciones de estos gaps, sacando factor común entre ambas fuentes, por lo que albergaremos en una única matriz todas las posiciones en la que se ha dado ausencia de datos, ya sea debido a una de las fuentes o a ambas. Ya conocidas las posiciones de los gaps, se generará una nueva matriz con tantos datos como minutos haya en el mes. En ella identificaremos con un valor de “1” las posiciones en las que disponemos de datos y con el valor “0” marcamos las posiciones de gaps.

La siguiente figura representa claramente un intervalo de gaps que se produjo en el mes de abril por un corte de luz en la sistema de recepción de la señal satelital, donde estuvo 19 horas sin recoger datos.

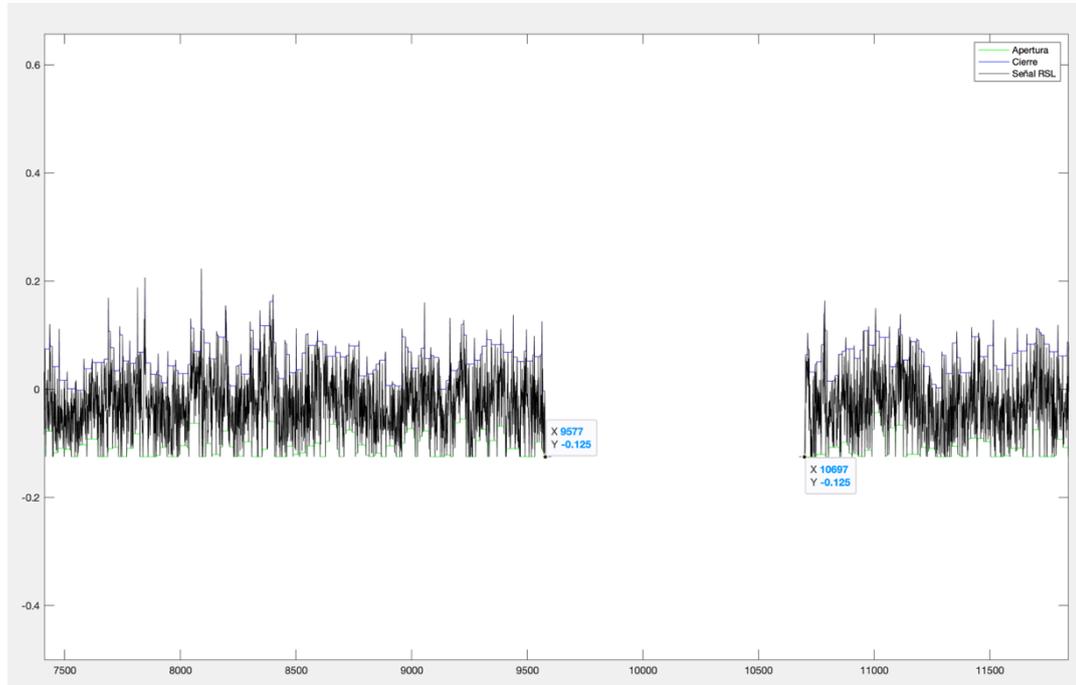


Figura 41. Gap RSL - abril de 2020.

El siguiente paso de análisis consiste en suprimir las componentes de muy baja frecuencia (p.e., evolución día-noche) con la finalidad de aislar los cambios abruptos asociados a la lluvia. Esto se ha hecho mediante una morfología matemática. La morfología matemática es una teoría basada en conceptos de geometría, álgebra, topología y teoría de conjuntos, creada para caracterizar propiedades físicas y estructurales de diversos materiales. La idea central de esta teoría es examinar las estructuras geométricas de una imagen por superposición con pequeños patrones localizados (denominados elementos estructurantes) en distintas partes de la misma [31] [32]. Así, aplicando apertura y cierre morfológico a nuestra señal satelital se puede obtener la componente de baja frecuencia. Restadas estas de la señal RSL original, quedaría como se muestra en la Figura 42. En esta figura se puede apreciar cómo queda una señal de ruido que fluctúa $\pm 0'125 \text{ dB}$. Esto es debido al ruido de cuantificación al digitalizar los valores de RSL. En sí mismo este valor de $0'125 \text{ dB}$ es una especie de valor mínimo de variación detectable debida a un evento de lluvia. Esto es, podrían ser detectables los eventos que causen atenuaciones superiores a este valor.

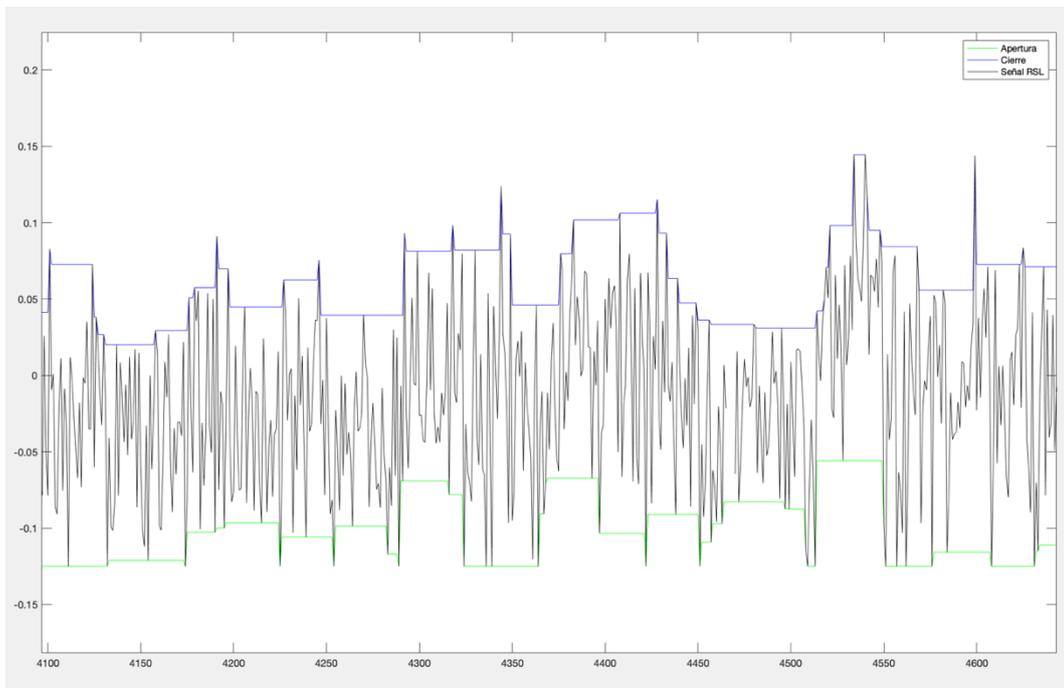


Figura 42. Señal RSL a la que se le ha quitado la componente de baja aplicado una apertura y un cierre morfológico.

6.4 Detector óptimo de umbral

El umbral empleado en este proceso es un umbral adaptativo, es decir, haciendo uso de una ventana, cuya longitud es prefijada, se cogen x datos anteriores y realiza un promedio, guardando el resultado en la misma posición en la que estaba centrada.

Para cada mes se obtiene un valor óptimo de umbral. Cuando se quiere realizar el análisis de varios meses u otros periodos, es necesario encontrar el valor de umbral óptimo que permita obtener los mejores resultados posibles para todo el conjunto.

Las siguientes figuras representan un ejemplo de este proceso, donde se pueden ver representados, mediante líneas discontinuas, tres de los umbrales analizados.

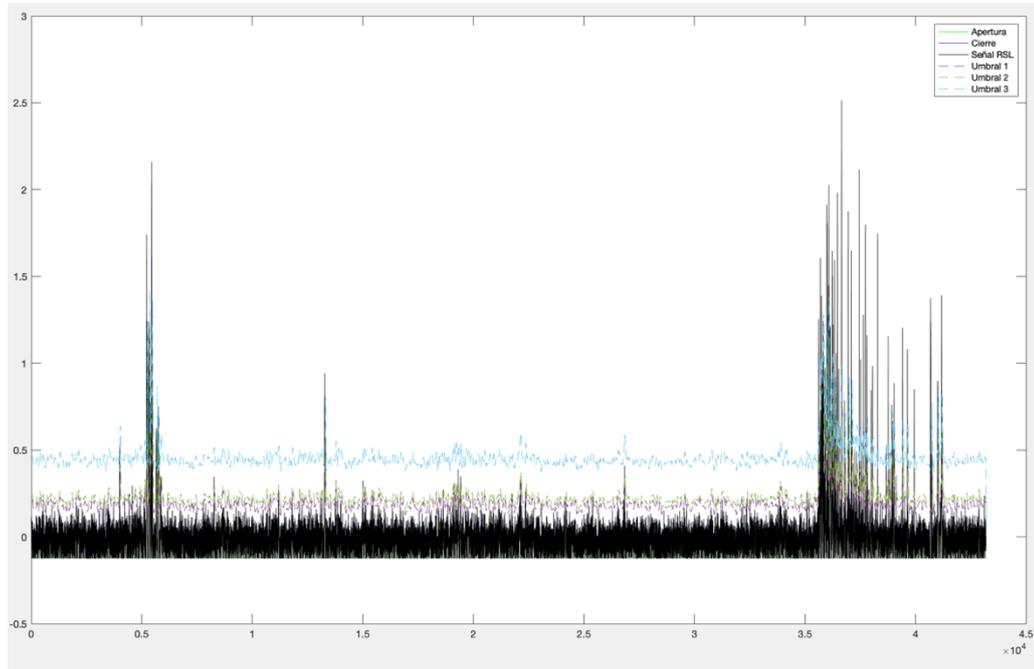


Figura 43. Aplicación de umbrales a la señal RSL.

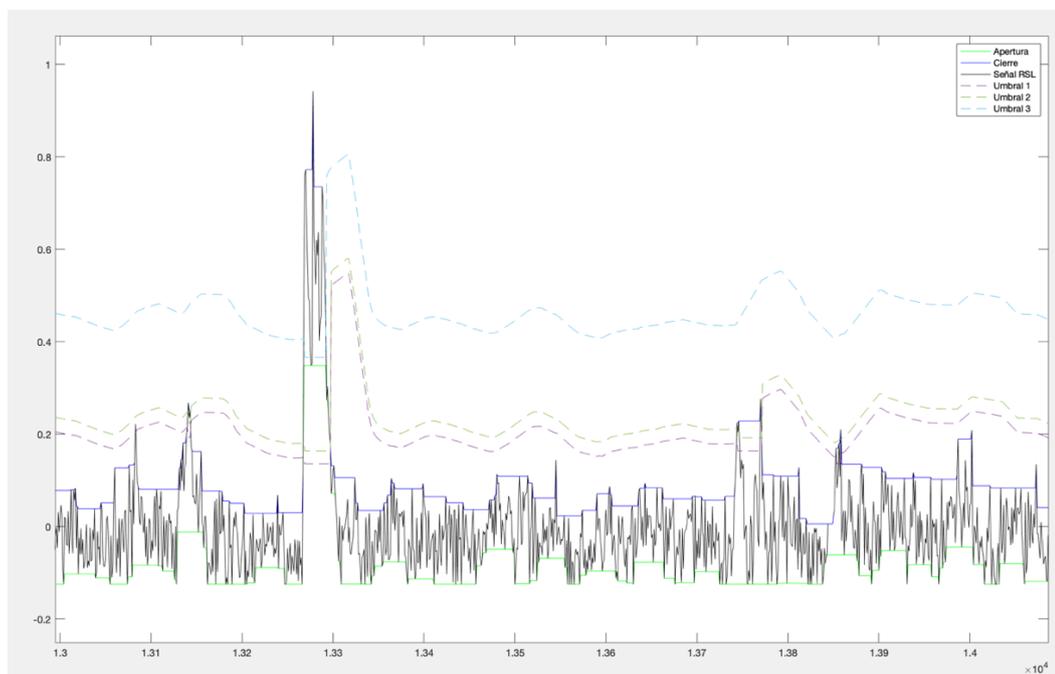


Figura 44. Aplicación de los umbrales a señal RSL – Zoom.

```
Comienza el bucle de análisis sobre los coeficientes de umbral
Analizado el coeficiente de umbral nº 1 de 10
Analizado el coeficiente de umbral nº 2 de 10
Analizado el coeficiente de umbral nº 3 de 10
Analizado el coeficiente de umbral nº 4 de 10
Analizado el coeficiente de umbral nº 5 de 10
Analizado el coeficiente de umbral nº 6 de 10
Analizado el coeficiente de umbral nº 7 de 10
Analizado el coeficiente de umbral nº 8 de 10
Analizado el coeficiente de umbral nº 9 de 10
Analizado el coeficiente de umbral nº 10 de 10
Coeficiente k óptimo: 2.25
```

Figura 45. Bucle para el cálculo del umbral óptimo.

El cálculo del umbral óptimo, $k0_{opt}$, se realiza mediante un bucle donde el valor de umbral ($k0$) cambiará en cada iteración dentro de unos valores prefijados, $k0 = [k0_{ini}:k0_{delta}:k0_{fin}]$. En nuestros experimentos, $k0_{ini} = 0'95$, $k0_{fin} = 3$ y $k0_{delta} = 0'05$. La fórmula de aplicar el umbral es $umbral = (mean(RSL(ind)) + k0 * q_Error)$, donde el umbral depende la media de valores de la señal RSL una serie de minutos antes (p.e., 25) expresados en 'ind'. Se entiende que el nivel mínimo de lluvia detectable es $q_Error = 0'125\text{ dB}$. Este valor depende del error de cuantificación al digitalizar los niveles de RSL medidos. Entonces, un nivel que sobresalga lo suficiente se supone debido a la lluvia. El valor se determina de forma iterativa en experimentos al efecto.

Con los resultados obtenidos se generarán unos valores de sensibilidad y especificidad dados por las curvas ROC (*Receiver Operating Characteristic*), como se puede observar en la Figura 46. La zona rosa de la figura representa la de un clasificador muy malo, nada fiable. La verde se obtiene de un suavizado (línea roja) hecho a partir de los puntos (en azul) dados por cada valor de umbral. A partir de estos datos se generará una estimación del umbral óptimo haciendo uso de la medida de Sensibilidad (Sen) y Exhaustividad (Exh), representadas en la Figura 47 para el conjunto de meses seleccionado. El criterio elegido es maximizar el F1-score = $2 * Exh * Sen / (Exh + Sen)$ pues es un parámetro ampliamente usado como métrica de calidad, y representa un compromiso entre ambos. Estas métricas las veremos con más detalle en el Capítulo 7.

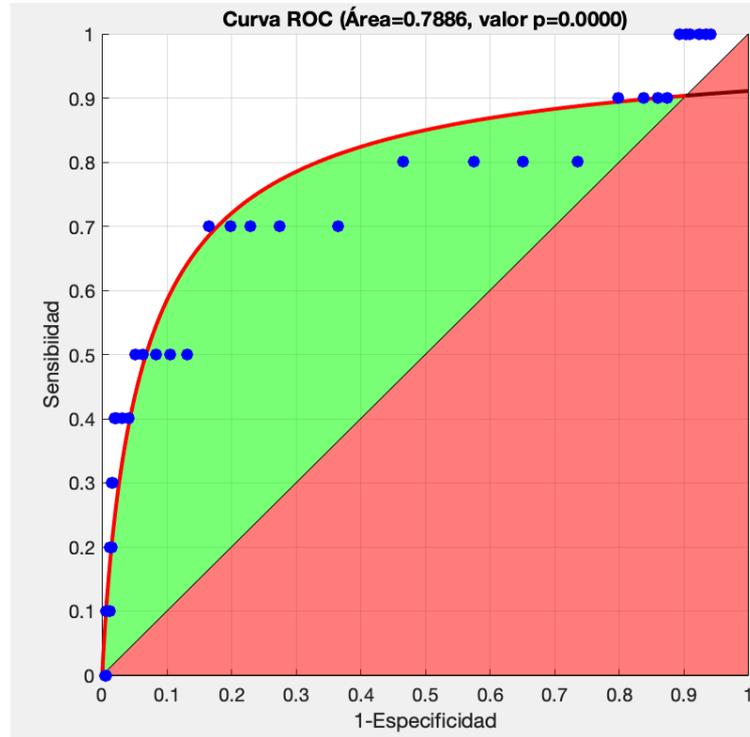


Figura 46. Curva ROC.

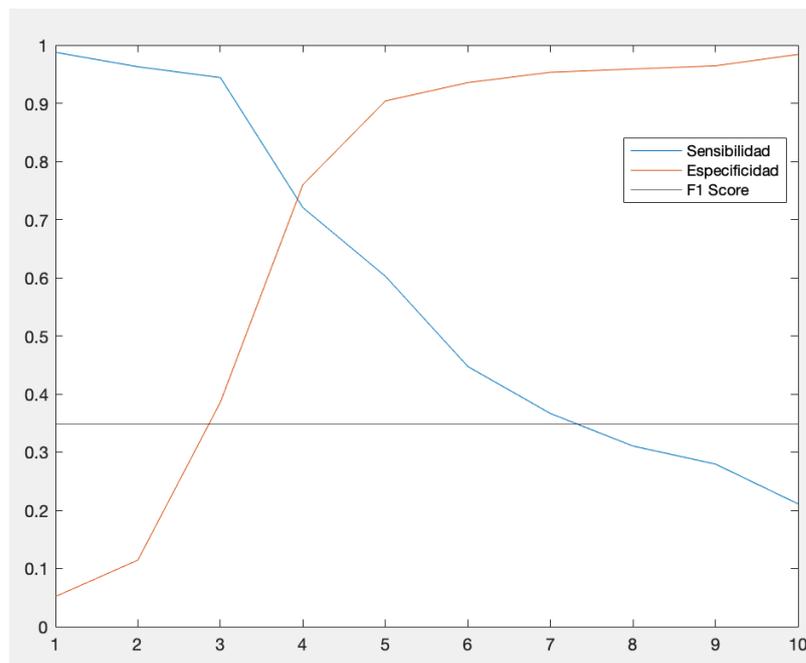


Figura 47. Sensibilidad vs Especificidad.

6.5 Detector de umbral y márgenes de confianza

En el paso anterior se ha determinado el valor de umbral óptimo, como el representado en la Figura 48, el siguiente paso será, una vez estimado, aplicar dicho umbral a nuestra señal para la detección de eventos de lluvia.

El proceso a seguir consiste en recorrer la señal, ya preparada previamente, detectando pendientes mayores que el umbral. Una pendiente superior al umbral establecido indica que se ha producido un evento de lluvia. En este caso, se considera evento de lluvia a aquellos que duren más de un minuto.

Una vez detectado un valor superior al umbral, se añadirá una marca de inicio de evento, y se seguirá avanzando sobre la base de datos, comprobando siempre que se siga cumpliendo dicha condición. En cuanto se detecte una pendiente menor que el umbral, se considerará fin del evento de lluvia, lo cual quedará señalizado con una marca de fin de evento de lluvia.

En este apartado se ha tenido en la cuenta la presencia de posibles gaps, para lo que se ha empleado una ventana de análisis de 20 muestras, es decir, 20 minutos, la gestión del gap pues significa ausencia de datos.

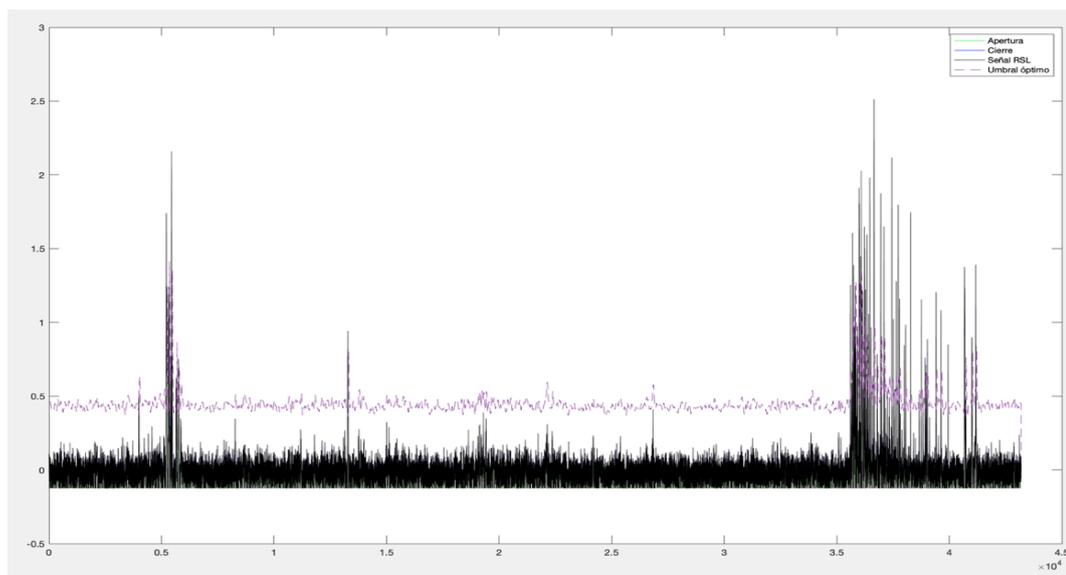


Figura 48. Aplicación del umbral óptimo a la señal RSL.

A la hora de establecer las marcas de principio y fin de la lluvia, tanto real como detectada, con el objetivo de determinar los intervalos de lluvia, se ha establecido un margen de confianza para evitar solapes entre eventos. El valor de este margen, especificado en minutos, lo iremos variando según nuestros intereses. Dispone de dos parámetros: el límite superior y el límite inferior.

La siguiente imagen representa un ejemplo sencillo donde se puede observar fácilmente un evento de lluvia, el cual queda marcado con las marcas de principio (“Inicio Lluvia”, en azul) y fin (“Fin Lluvia”, en verde).

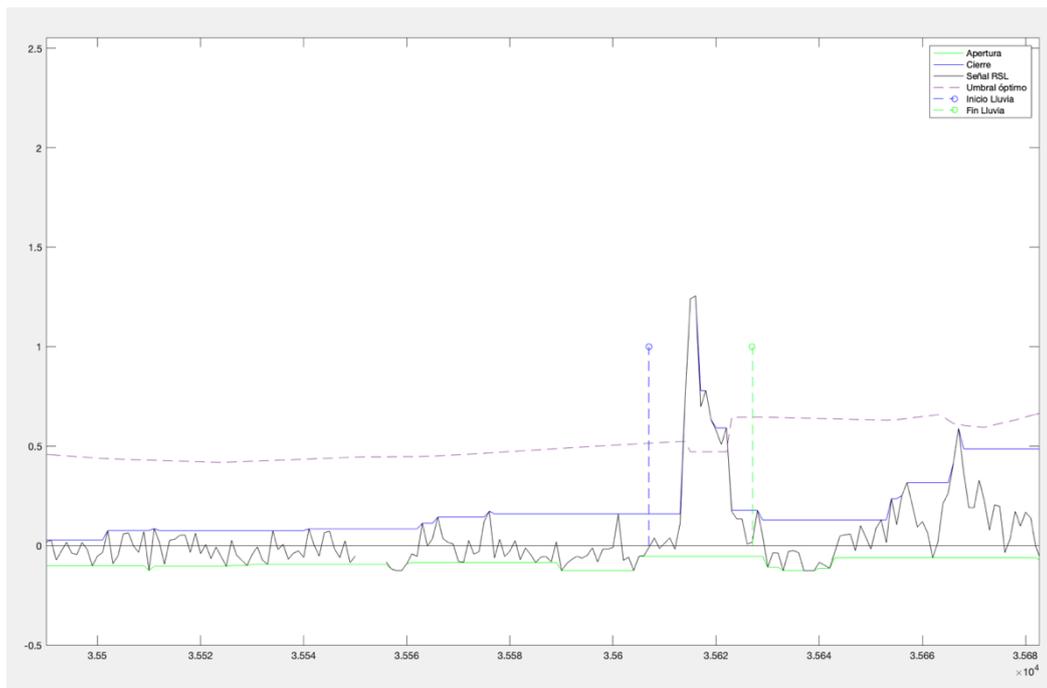


Figura 49. Evento de lluvia representado con las marcas de principio-fin.

Así, al final de este proceso tendremos una matriz organizada por pares, donde el primer valor de cada pareja se corresponderá con el minuto de comienzo del evento de lluvia, y el segundo valor representará el minuto en el que ha finalizado dicho evento de lluvia.

6.6 Proceso de selección de características de entrenamiento y test

A partir de las marcas de principio y fin de los intervalos de precipitación ya determinadas previamente, se genera un nuevo conjunto de datos adaptado a los requisitos de las redes LSTM, con las cuales vamos a entrenar.

Los eventos de lluvia y no lluvia se componen de características propias, haciéndolos unos totalmente diferente de los otros. Esto se entiende fácilmente con la figura que se muestra a continuación. Los eventos de lluvia son esas formas de señal que sobresalen del plato de ruido, presentando un patrón totalmente distinguible. Sin embargo, los eventos de no lluvia presentan las características de ruido blanco.

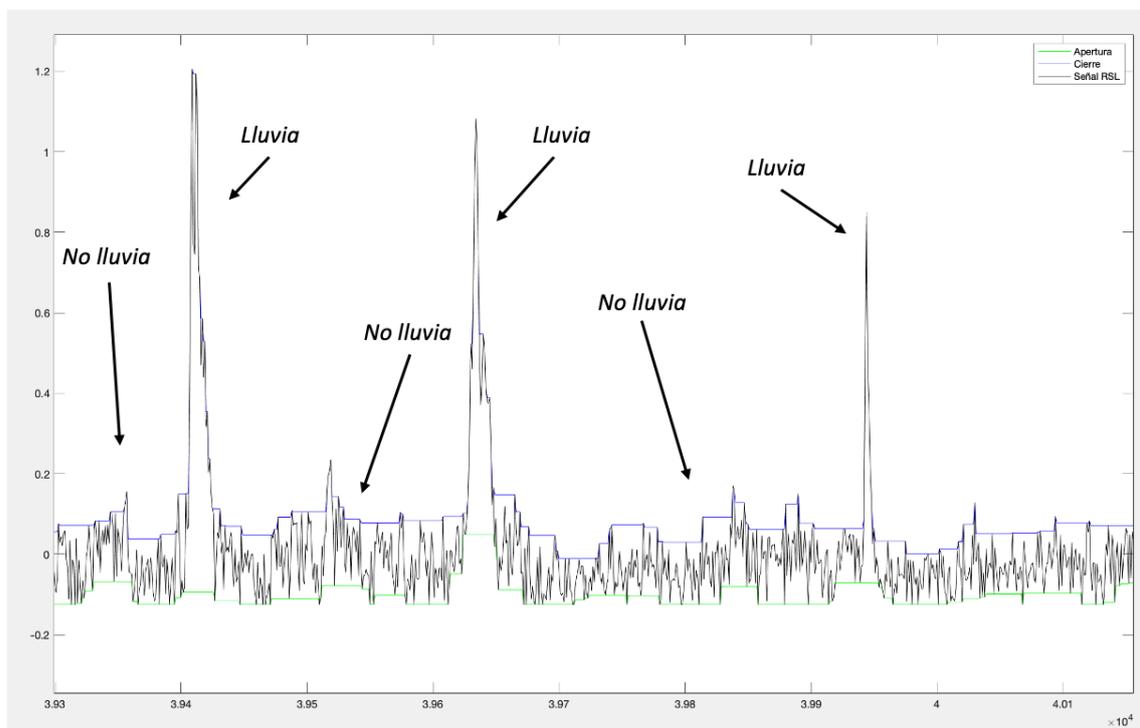


Figura 50. Características de los eventos de lluvia y no lluvia.

En una primera aproximación, lo que hemos hecho es tomar las señales procesadas según se explicó en el apartado 4.2, esto es, una vez que hemos suprimido el efecto de las bajas frecuencias. A partir de ahí, tomamos la señal muestra a muestra y se alimenta la Red LSTM.

6.7 Red LSTM

El algoritmo de detección se va a llevar a cabo mediante redes LSTM. Se trata de un entrenamiento supervisado, como ya se nombró en el Capítulo 5. Se presenta a la red un conjunto de patrones de entrada junto con la salida esperada. Los pesos se van modificando de manera proporcional al error que se produce entre la salida real de la red y la salida esperada.

En nuestro caso, el entrenamiento de la red para un mes x , se realizará utilizando dos meses anteriores al actual. Para ello, el primer paso será configurar las capas y las opciones correspondientes, las cuales se muestran en la siguiente figura.

```
% Train the LSTM Network
if doTraining
    % CAPAS Y OPCIONES
    layers = [ ...
        sequenceInputLayer( size(trainFeatureCell,1) )
        lstmLayer(200,"OutputMode","sequence")
        lstmLayer(200,"OutputMode","sequence")
        fullyConnectedLayer(2)
        softmaxLayer
        classificationLayer
    ];
    maxEpochs = 6;
    miniBatchSize = 16;
    options = trainingOptions("adam", ...
        "MaxEpochs",maxEpochs, ...
        "MiniBatchSize",miniBatchSize, ...
        "Shuffle","every-epoch", ...
        "Verbose",0, ...
        "SequenceLength",sequenceLength, ...
        "ValidationFrequency",10, ...
        "ValidationData",{trainFeatureCell.',trainLabelCell.'}, ...
        "Plots","training-progress", ...
        "LearnRateSchedule","piecewise", ...
        "LearnRateDropFactor",0.1, ...
        "LearnRateDropPeriod",5);
```

Figura 51. Capas y opciones para el entrenamiento de la red.

A continuación, vamos a hablar de las **capas** que utilizamos, las cuales se componen de 6 niveles:

- **1ª capa:** A la capa de entrada se la denomina “*sequenceInputLayer*”, y es donde se introducen los datos en forma de vectorial con sus correspondientes etiquetas.
- **2ª capa:** Es la capa LSTM. En su forma más habitual, se usa una “*lstmLayer*”. Esta capa procesará la salida de neuronas que están conectadas en «regiones locales» de entrada (es decir píxeles cercanos), calculando el producto escalar entre sus pesos (valor de píxel) y una pequeña región a la que están conectados en el volumen de entrada. En nuestro caso, utilizaremos 2 filtros y ese será el volumen de salida. Por otro lado, será la encargada también de recordar las características de todos los análisis para realizar una correcta evaluación. Una capa bidireccional LSTM (*BiLSTM*) es una capa RNN que aprende las dependencias bidireccionales a largo plazo entre los pasos de tiempo. Estas dependencias pueden ser útiles cuando desea que la red aprenda de la serie temporal. Es nuestro caso y ha dado mejores resultados que una “*lstmLayer*” unidireccional.
- **3ª capa:** Es la capa en la que se almacena el resultado del análisis realizado por la capa anterior. Se denomina “*dropoutLayer*”.
- **4ª capa:** Es la que diferencia a otras redes (p.e., CNN) de una red LSTM, la cual se denomina “*fullyConnectedLayer*”, ya que necesitan muchas menos conexiones necesarias, pues realmente los pesos que se ajustan serán los de los filtros que hemos establecido en la capa LSTM.
- **5ª capa:** Es la penúltima capa de la red, que se denomina “*softmaxLayer*”. Conecta con la capa de salida final que tendrá la cantidad de neuronas correspondientes con las clases que estamos clasificando.
- **6ª capa:** Es la última capa de la red, que se denomina “*classificationLayer*”. Esta capa se encarga de realizar la clasificación de los datos obtenidos de las capas intermedias.

```
layers = [ ...  
    sequenceInputLayer(numFeatures, 'Name', 'sequence')  
    lstmLayer(20, 'OutputMode', 'last', 'Name', 'lstm')  
    dropoutLayer(0.5, 'Name', 'drop')  
    fullyConnectedLayer(numClasses, 'Name', 'fc')  
    softmaxLayer('Name', 'softmax')  
    classificationLayer('Name', 'classification')  
];
```

Figura 52. Capas de entrenamiento de la RNN.

En segundo lugar, vamos a introducir las **opciones de entrenamiento** que hemos establecido para el funcionamiento de la red. Antes de proceder, hemos de precisar que usamos un entrenamiento supervisado, tal como se explicó en el capítulo 5, en el apartado 5.2.1.

- **Epoch:** Este es el número de veces que se van a pasar cada ejemplo de entrenamiento por la red. Nosotros establecemos un valor de 6, ya que hemos comprobado que es suficiente para que la red aprenda.
- **Batch Size:** Este es el número de ejemplos que se introducen en la red para que entrene de cada vez. Si el número es pequeño, significa que la red tiene en memoria poca cantidad de datos, y entrena más rápido. Sin embargo, es posible que no aprenda las características y detalles que pueden ser significativos en la predicción. En cambio, si es grande es más probable que tenga en cuenta los casos más importantes a la hora de aprender, pero entrena más lento. En nuestro caso, hemos utilizado un número elevado, ya que no nos afecta demasiado el tiempo de ejecución y, además, nos sirve para que la red aprenda numerosos detalles.
- **InitialLearnRate:** Este parámetro controla la tasa o velocidad a la que aprende el modelo. Específicamente, controla la cantidad de error repartido con el que se actualizan los pesos del modelo cada vez que se actualizan, como al final de cada lote de ejemplos de entrenamiento. En el caso de nuestra red, le hemos dado un valor de $1 \cdot 10^{-4}$, Esto significa que una tasa de aprendizaje de $1 \cdot 10^{-4}$, significaría que los pesos en la red se actualizan $1 \cdot 10^{-4} \cdot (\text{error de peso estimado})$ o 0.01 % del error de peso estimado cada vez que se actualizan los pesos.

- **GradientThreshold:** El recorte de degradado implica forzar los valores del degradado (por elementos) a un valor mínimo o máximo específico si el degradado excedió un rango esperado. En nuestro caso, hemos puesto un valor de 2, lo que significa que si un valor de gradiente fuera menor que -2 , se establece en -2 y si es mayor que 2, entonces se configurará en 2.
- **ValidationData:** Este parámetro recoge los niveles de entrenamiento que se van a evaluar y las diferentes secuencias que se han recogido especificando sus tamaños.
- **ValidationFrequency:** Este parámetro define el número de iteraciones que se van a realizar por cada *Epoch*. Dicho valor se establece a partir del cociente entre el número de observaciones dividido por el *Batch Size*.

```
options = trainingOptions('adam', ...  
    'MaxEpochs',maxEpochs, ...  
    'MiniBatchSize',miniBatchSize, ...  
    'InitialLearnRate',1e-4, ...  
    'GradientThreshold',2, ...  
    'Shuffle','every-epoch', ...  
    'ValidationData',{sequencesTrain,labelsTrain}, ...  
    'ValidationFrequency',numIterationsPerEpoch, ...  
    'Plots','training-progress', ...  
    'Verbose',false);
```

Figura 53. Opciones de entrenamiento de la RNN.

Seguidamente se ejecutará el test sobre el mes x . Lo primero que se comprueba es que han existido eventos de lluvia. En caso de no haber registrado marcas de lluvia, nos saltaremos este paso, pues no habrá datos válidos como consecuencia de la falta de eventos de test.

En caso de que sí se detecten eventos de lluvia, se procede a ejecutar el test. Se tendrán en cuenta una serie de métricas de calidad, a partir de las cuales obtenemos los valores de sensibilidad y especificidad, entre otros, y se generarán las curvas ROC correspondientes. De esta forma conseguimos evaluar el grado de acierto de nuestra red.

CAPÍTULO 7.

EXPERIMENTOS Y RESULTADOS DE ESTIMACIÓN DE LLUVIA

7.1 Parámetros de calidad

Para definir los parámetros de calidad tendremos en cuenta que estamos haciendo una detección. A tal efecto, utilizamos unos parámetros calidad que se definen a partir de las detecciones, según el tipo de acierto o confusión. Además, se hace un estudio de la fiabilidad de los modelos obtenidos.

Antes de proceder hemos de precisar que los eventos de lluvia de referencia para el cálculo de los umbrales óptimos y los modelos de la red LSTM son los que muestran las estaciones meteorológicas que disponemos en el Pabellón B. Además, hay otros eventos que se producen entre Tafira Baja y las cumbres de la isla de los que no tenemos datos. En estos casos el sistema basado en el receptor satelital sí que muestra eventos de lluvia.

7.1.1 Basados en la matriz de confusión

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 9 se muestran estos conceptos visualmente:

Tabla 9. Definición gráfica de True Positive, True Negative, False Positive y False Negative.

		Predicción	
		Positiva	Negativa
Observación	Positiva	Verdadero Positivo <i>True Positive (TP)</i>	Falso Negativo <i>False Negative (FN)</i>
	Negativa	Falso Positivo <i>False Positive (FP)</i>	Verdadero Negativo <i>True Negative (TN)</i>

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 10 se presenta nuestro modelo casos de lluvia con una matriz de confusión de 2x2 que muestra los cuatro resultados posibles.

Tabla 10. Matriz de confusión del modelo.

		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).

Para evaluar el rendimiento de la clasificación hemos utilizado los parámetros: sensibilidad, especificidad, exactitud 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 (7.1) y (7.2).

$$\text{Sensibilidad} = \frac{\text{n}^{\circ} \text{ eventos detectados}}{\text{n}^{\circ} \text{ eventos meteorología}} = \frac{TP}{TP + FN} \quad (7.1)$$

$$\text{Especificidad} = \frac{\text{n}^{\circ} \text{ falsas alarmas}}{\text{n}^{\circ} \text{ totales de muestras sin evento}} = \frac{TN}{TN + FP} \quad (7.2)$$

La exactitud mide la proporción del total de elementos clasificados correctamente, ecuación (7.3).

$$\text{Exactitud} = \frac{TP + TN}{TP + TN + FP + FN} \quad (7.3)$$

El valor F1 mide la precisión del modelo. Se calcula haciendo la media armónica entre la sensibilidad y la Exhaustividad. Estos se definen en las siguientes ecuaciones:

$$\text{Exhaustividad} = \frac{TP}{TP + FN} \quad (7.4)$$

$$F_1 = 2 \cdot \frac{\text{Sensibilidad} \cdot \text{Exhaustividad}}{\text{Sensibilidad} + \text{Exhaustividad}} \quad (7.4)$$

7.1.2 Basados en la media y desviación sobre múltiples experimentos

Es importante saber cuánto 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:

- Definir un número total de repeticiones (N), en nuestro caso N=100.
- En cada iteración hemos seleccionado de la base de datos un cierto porcentaje para y el resto para testear los modelos, en el bien entendido de que los ejemplos de test no participan en el entrenamiento.
- La selección entre entrenamiento y test se hace de forma aleatoria y de tal forma que en cada repetición los datos que entrenan no participan en el test.
- La sección de los porcentajes óptimos será objeto de estudio.

Una vez obtenidos los valores de detección en cada uno de los N 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.

7.2 Experimentos para la estimación de umbral óptimo

Con el fin de obtener el umbral “k óptimo”, el cual se utilizará en el posterior cálculo del umbral una vez que estemos en los experimentos de detección propiamente dicha, se lleva a cabo una serie de experimentos. A continuación, se muestra el diagrama de flujo seguido para el cálculo de dicho coeficiente. Antes de explicar cómo se determina el óptimo, recordemos del apartado 6.4 cómo se calcula éste. La fórmula para aplicar el umbral es:

$$umbral = (mean(RSL(ind)) + k_{opt} * q_{Error}) \quad (7.5)$$

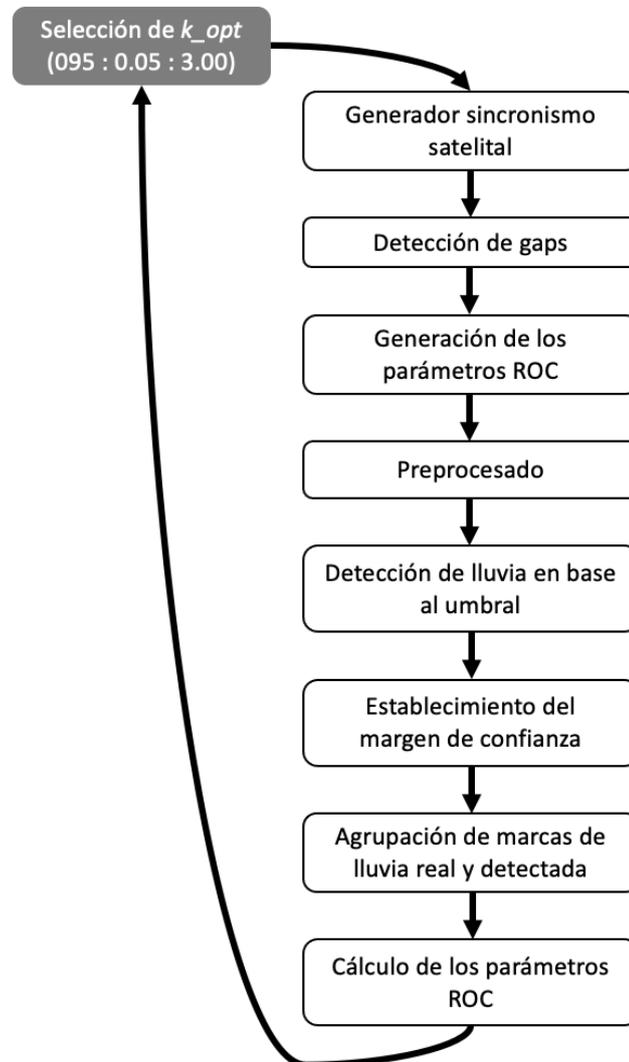


Figura 54. Diagrama de flujo seguido para el cálculo del coeficiente k_{opt} .

Los pasos a seguir para el cálculo de los resultados en base a un k óptimo determinado, comienza por realizar el sincronismo de las bases de datos con las que se van a trabajar, procedentes de las estaciones meteorológicas y el radioenlace satelital. Una vez hecho esto, se procederá a la localización y tratamiento de los gaps.

En el siguiente paso se generarán los parámetros que permitirán generar las curvas ROC (*Receiver Operating Characteristic*). Realizado el preprocesado de los datos, se procederá a

la detección de eventos de lluvia en base al umbral seleccionado. Con el fin de mejorar estas detecciones, se establece un margen de confianza seleccionable a ambos lados de cada evento de precipitación registrado.

A cada evento de lluvia se le asignará unas marcas de principio y fin, las cuales se utilizarán para obtener una nueva curva ROC.

Mediante un bucle, se irá variando el valor de k_0 entre 0.95 y 3 con saltos de 0.05, repitiendo en cada una de las iteraciones los pasos anteriormente comentado. Para cada uno de los k_0 se obtiene una matriz con los parámetros TP, TN, FP y FN. A través de éstos se puede calcular la sensibilidad y especificidad.

Representando los valores de sensibilidad y especificidad podremos obtener la curva ROC en cada una de las iteraciones, como la que se muestra en la Figura 55 (mes de febrero de 2020).

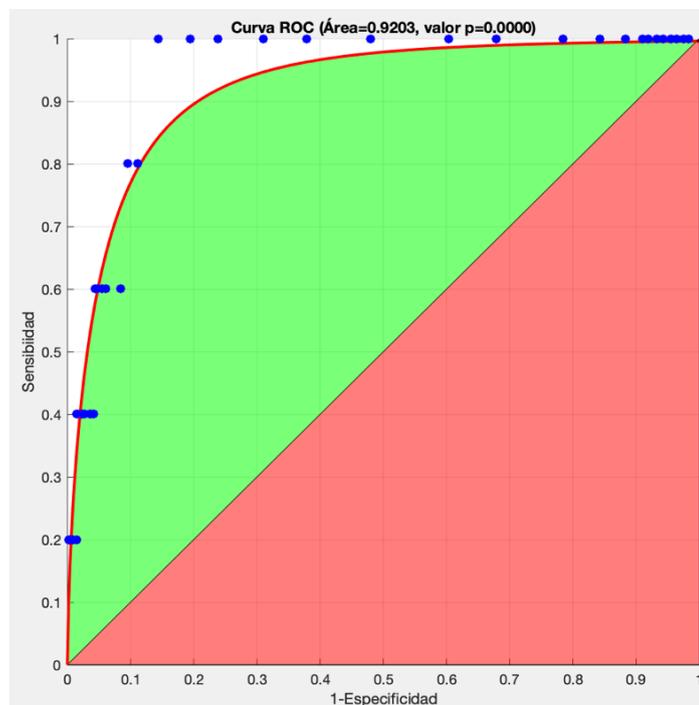


Figura 55. Curva ROC - 02/2020.

Esta curva nos permite obtener el valor de k óptimo para los meses seleccionados. En este caso, el umbral óptimo es que da Sensibilidad ≈ 1 y falsa alarma $\approx 0,12$.

En este TFG estamos trabajando con 16 meses, como ya se comentó en el apartado 4.2.1. El cálculo de dicho parámetro se realiza de forma diferente dependiendo del mes de estudio:

- Para el primer mes, el cálculo del k óptimo se realiza con el propio mes seleccionado.
- A partir del segundo mes, el k óptimo se calcula de dos en dos meses.

Este proceso se muestra en el trozo de código de la imagen siguiente, donde el mes inicial será siempre igual a 1 ($Mes_Ini=1$) y el mes final será igual número de meses que queramos analizar. Estamos realizando el estudio de 16 meses (de agosto de 2019 a noviembre de 2020), por lo que en este caso el valor del mes final tomará el valor de 16 ($Mes_Fin=16$).

```
for nn=Mes_Ini:Mes_Fin    % Recorreremos los meses de interés

    % Entrenamiento del k_opt
    if (nn==1)
        Mes_1 = 1;
        Mes_2 = 1;
    elseif (nn==2)
        Mes_1 = 1;
        Mes_2 = 2;
    else
        Mes_1 = nn-2;
        Mes_2 = nn-1;
    end
```

Figura 56. Entrenamiento para el cálculo del k óptimo.

Siguiendo estas pautas, obtenemos todos los valores de k óptimo para cada par de meses, como se observa en la Figura 57.

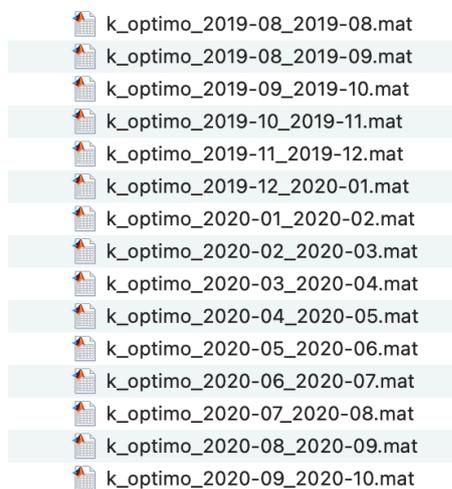


Figura 57. Archivos de k_{opt} por meses.

Los datos quedan recogidos en la tabla siguiente:

Tabla 11. Datos de k_{opt} de los meses de estudio.

Mes	k_{opt}	Nombre del archivo
<i>Agosto de 2019</i>	2.80	k_optimo_2019-08_2019-08.mat
<i>Agosto y septiembre de 2019</i>	2.80	k_optimo_2019-08_2019-09.mat
<i>Septiembre y octubre de 2019</i>	3.00	k_optimo_2019-09_2019-10.mat
<i>Octubre y noviembre de 2019</i>	3.00	k_optimo_2019-10_2019-11.mat
<i>Noviembre y diciembre de 2019</i>	2.75	k_optimo_2019-11_2019-12.mat
<i>Diciembre de 2019 y enero de 2020</i>	2.20	k_optimo_2019-12_2020-01.mat
<i>Enero y febrero de 2020</i>	1.95	k_optimo_2020-01_2020-02.mat
<i>Febrero y marzo de 2020</i>	2.80	k_optimo_2020-02_2020-03.mat
<i>Marzo y abril de 2020</i>	2.80	k_optimo_2020-03_2020-04.mat
<i>Abril y mayo de 2020</i>	2.95	k_optimo_2020-04_2020-05.mat
<i>Mayo y junio de 2020</i>	2.95	k_optimo_2020-05_2020-06.mat
<i>Junio y julio de 2020</i>	2.45	k_optimo_2020-06_2020-07.mat
<i>Julio y agosto de 2020</i>	3.00	k_optimo_2020-07_2020-08.mat
<i>Agosto y septiembre de 2020</i>	2.55	k_optimo_2020-08_2020-09.mat
<i>Septiembre y octubre de 2020</i>	2.95	k_optimo_2020-09_2020-10.mat

En este caso solo se ha realizado el cálculo hasta los meses septiembre y octubre, puesto que, a la hora de testear los resultados de la red con respecto a un mes, realizaremos el entrenamiento con los dos meses anteriores a este. Debido a que nuestros datos llegan hasta noviembre de 2020, realizaremos el entrenamiento con septiembre y octubre, por lo que sólo será necesario tener los valores de k óptimo hasta este último par.

7.3 Experimentos de detección mes a mes

Teniendo en cuenta los umbrales previamente calculados, el siguiente paso consiste en realizar la detección de lluvia mes a mes. Un ejemplo de los datos obtenidos es lo que se puede observar en la Figura 58.

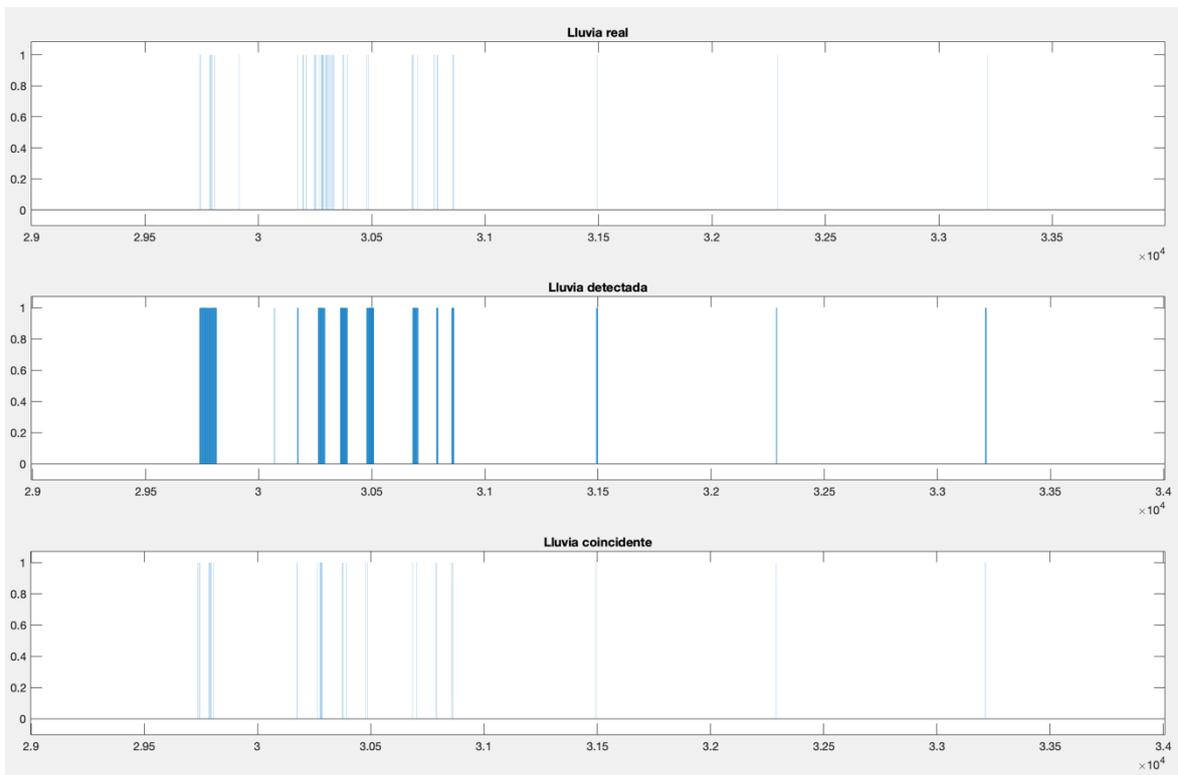


Figura 58. Lluvia real vs. Lluvia detectada - octubre de 2020.

Esta figura, correspondiente a la última parte del mes de octubre de 2020, es el resultado obtenido de aplicar el umbral óptimo correspondiente los meses de agosto y septiembre de 2020, es decir, los dos meses anteriores, como ya se indicó en el apartado anterior.

En la imagen están representadas con barras de valor 1 las marcas de lluvia detectadas por el pluviómetro, llamada “Lluvia real”, en la primera gráfica. La segunda gráfica corresponde a la lluvia detectada por el radioenlace satelital, llamada “Lluvia detectada”. La tercera gráfica representa los eventos de lluvia que han sido detectadas por ambas fuentes.

El término “Lluvia real”, en nuestro caso, representa los eventos de lluvia que han sido detectados por la estación meteorológica o el pluviómetro y de los cuales tenemos registros. Son los que tomamos como referencia para entrenar la red o estimar los umbrales óptimos.

Una observación que consideramos importante viene del hecho de que alguno de estos dispositivos no registre un evento de lluvia. Esto no significa que no lo haya habido. Es decir, como ya se explicó en apartados anteriores, por ejemplo, en casos en los que la lluvia pueda ser muy tenue, puede no haber ninguna marca en nuestras bases de datos que lo indique. Así, uno de los hallazgos más interesantes de este TFG es que el detector satelital es muy sensible y fiable detectando lluvia. En consecuencia, los datos proporcionados por este son altamente fiables. Esto ha sido constatado por nosotros en numerosas ocasiones de forma presencial. Por ejemplo, el radioenlace detectará esa lluvia tenue que el pluviómetro no fue capaz de contabilizar. Además, también comentado anteriormente en el apartado 5, el radioenlace recoge datos a lo largo de toda su trayectoria, mientras que la estación meteorológica que estamos empleando sólo lo hace en un punto. Con esto se pretende argumentar que, los datos representados en la Figura 58 en la gráfica central, llamada “Lluvia detectada” es realmente lluvia, pero pudo haber sido gran parte de ella detectada fuera de la zona del Campus de Tafira o pudo haber sido tan débil que el pluviómetro no lo detectara.

Por ello, cuando se trate de dar datos de lluvia en Tafira Baja, trabajaremos con los datos coincidentes para ambos tipos de fuente, pluviómetro-estación y receptor satelital.

Se puede apreciar visualmente que el radioenlace satelital presenta una mayor sensibilidad a los efectos de la lluvia, ya que detecta una mayor cantidad de eventos.

Este proceso de detección y recopilación de lluvia detectada por ambas fuentes se lleva a cabo para cada uno de los meses aplicando los umbrales correspondientes. Es más, aprovechando que el prototipo satelital está en el Pabellón B, hemos podido comprobar en numerosas ocasiones en que hay lluvia en Tafira, que su sensibilidad es mayor que las estaciones meteorológicas que tenemos a disposición. Particularmente, existen eventos de bajos niveles de lluvia en los que las estaciones no registran nada mientras que el prototipo satelital sí permite detectarla con nuestro procesado.

7.4 Principales resultados

Llegados a este punto, ya se encuentran almacenados los datos de lluvia - no lluvia registrados por ambas fuentes. A partir de estos datos, se procede al entrenamiento y test de la red neuronal.

En la siguiente tabla se muestran los porcentajes de exactitud de validación obtenidos a partir del entrenamiento (en este caso 70% para entrenar) con cada uno de los meses. Se puede observar que durante la validación en el entrenamiento la red neuronal genera unos resultados de exactitud (*accuracy*) muy buenos, pues los valores de exactitud se encuentran siempre por encima del 90%. Esto es un buen indicador de que los modelos están yendo bien. Falta testear para conocer bien cómo son de fiables nuestros modelos.

Tabla 12. Validation Accuracy.

Mes	<i>k_{opt}</i>	<i>Validation Accuracy (%)</i>
08-2019	2.80	90.89 %
09-2019	2.80	93.97 %
10-2019	2.80	90.89 %
11-2019	3.00	96.16 %
12-2019	3.00	97.75%
01-2020	2.75	97.52 %

Mes	k_{opt}	Validation Accuracy (%)
02-2020	2.20	96.35 %
03-2020	1.95	96.39 %
04-2020	2.80	97.27 %
05-2020	2.80	97.34 %
06-2020	2.95	95.82 %
07-2020	2.95	94.58 %
08-2020	2.45	93.07 %
09-2020	3.00	92.82 %
10-2020	2.55	93.83 %
11-2020	2.95	96.54 %

Las siguientes figuras representan los valores de sensibilidad (*Sen*), especificidad (*Spe*), exactitud (*Acc*) y la media armónica entre sensibilidad y especificidad (*F1 Score*), extraídos de las matrices de confusión (TN, FP, FN, TP).

Los experimentos que mostraremos están en función de la cantidad de datos, visto en términos del porcentaje usado sobre el total, que se usan para entrenar. El objetivo de esto es conocer la fiabilidad de nuestro método de detección frente a la cantidad de datos disponibles, pues los eventos de lluvia son poco frecuentes. En las figuras se muestran, para cada mes, los parámetros de calidad obtenidos a partir de modelos entrenados con los dos meses anteriores. Superpuestos están los márgenes de desviación respecto de la media.

En la primera figura que se muestra a continuación, se ha utilizado un 50 % de los datos para entrenar, y el otro 50% para testear. Los resultados obtenidos para cada uno de los meses son los siguientes:

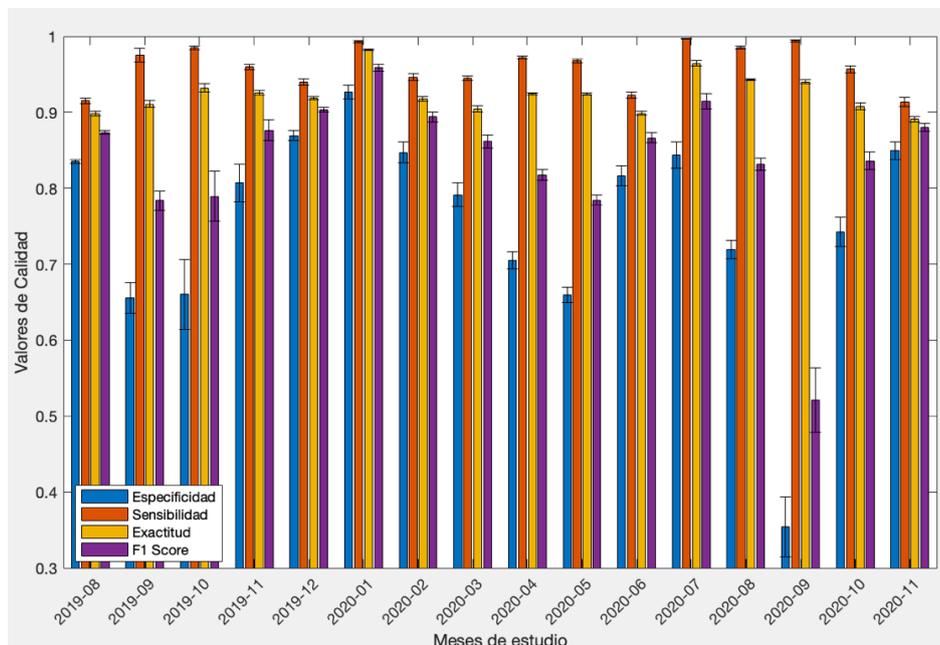


Figura 59. Parámetros de calidad obtenidos para cada uno de los meses de estudio utilizando un 50% de los datos.

En esta segunda figura se ha utilizado un 60 % de los datos para entrenar, y el otro 40% para testear. Los resultados obtenidos son los siguientes:

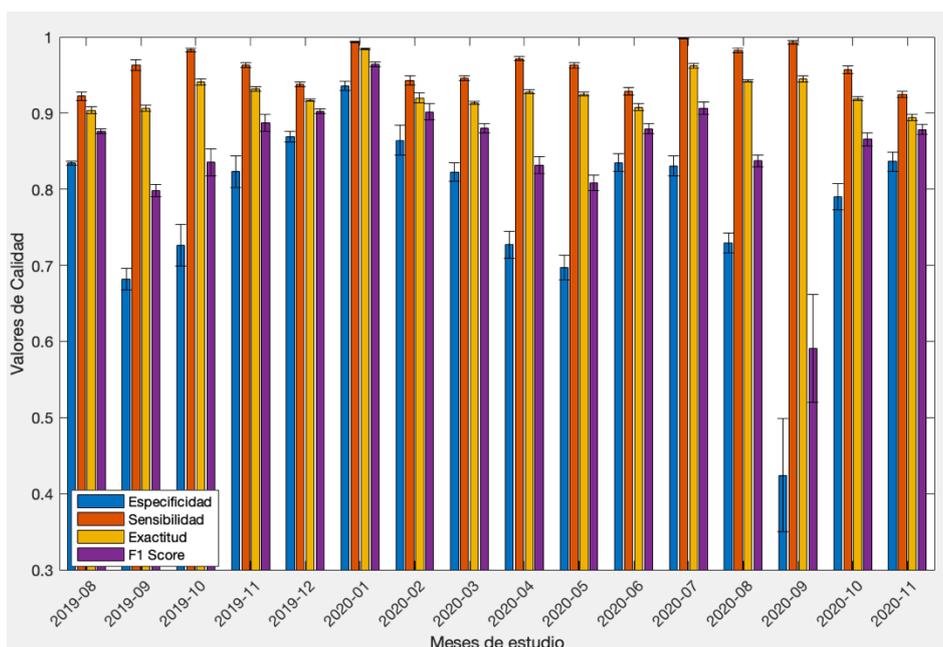


Figura 60. Parámetros de calidad obtenidos para cada uno de los meses de estudio utilizando un 60% de los datos.

En esta tercera figura se ha utilizado un 70 % de los datos para entrenar, y el otro 30% para testear. Los resultados obtenidos son los siguientes:

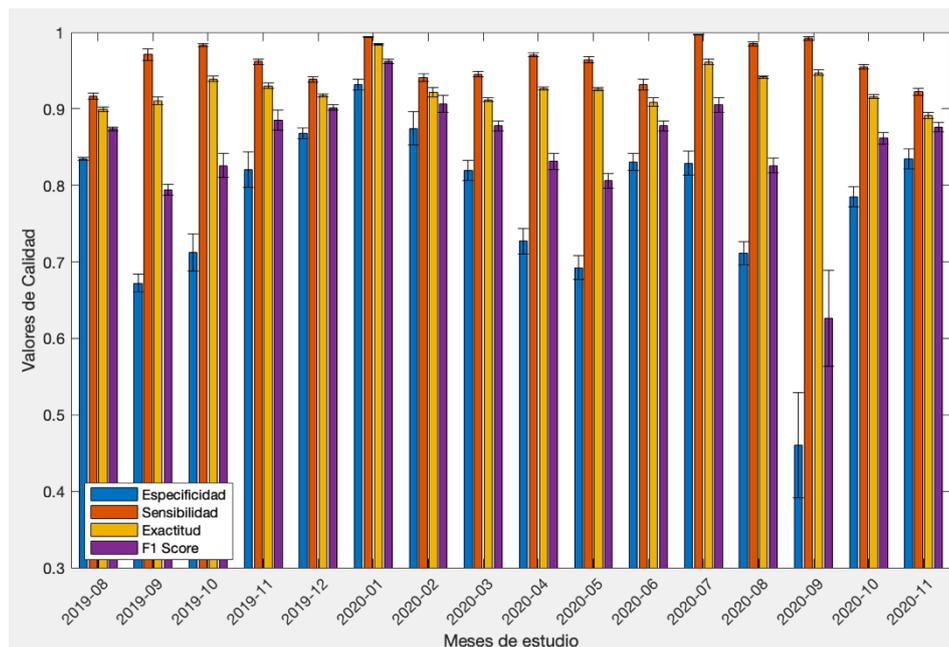


Figura 61. Parámetros de calidad obtenidos para cada uno de los meses de estudio utilizando un 70% de los datos.

Para la cuarta figura se ha utilizado un 80 % de los datos para entrenar, y el otro 20% para testear. Los resultados obtenidos son los siguientes:

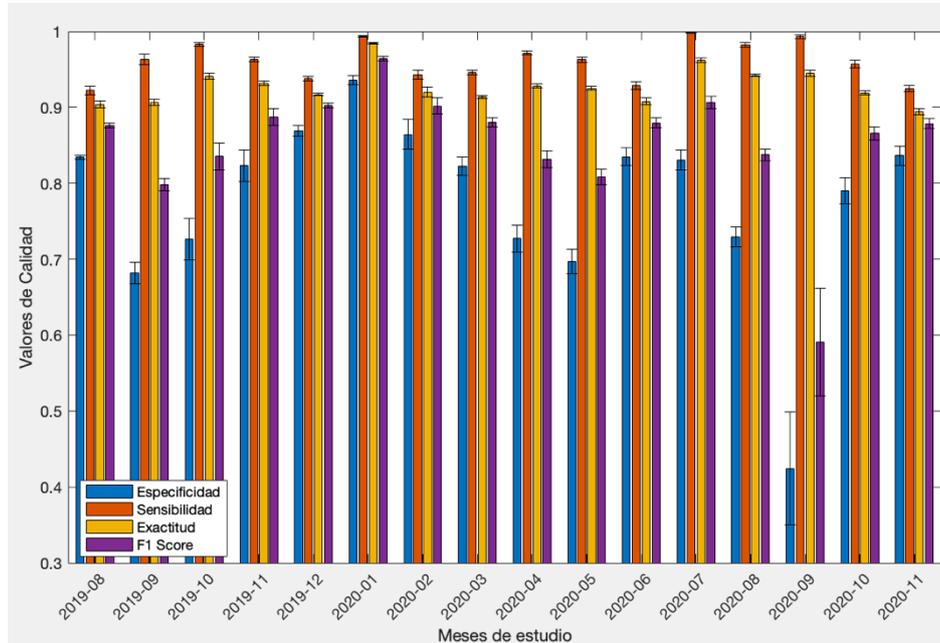


Figura 62. Parámetros de calidad obtenidos para cada uno de los meses de estudio utilizando un 80% de los datos.

Prestando atención al mes de septiembre de 2020 en cada uno de los casos de estudio anteriores, se ha decidido calcular también los valores de emplear un 90% de datos para el entrenamiento, dado que hemos obtenido unos resultados que son mejorables.

El motivo de que estos datos no sean tan óptimos como el resto, se debe a cómo han sido los datos tomados en julio y agosto, meses que usamos para entrenar (umbral y modelos), ya que se disponen de muy pocos eventos de lluvia. Aplicando un 90% de los datos en el entrenamiento los resultados mejoran sensiblemente, si bien no hasta llegar a ser tan buenos como otros meses.

Los resultados obtenidos se muestran en la siguiente figura:

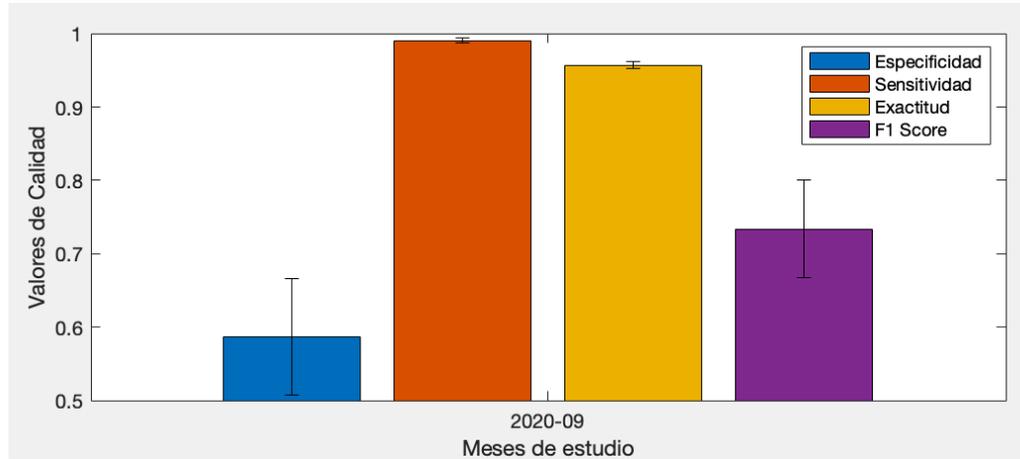


Figura 63. Parámetros de calidad obtenidos para el mes de septiembre de 2020 utilizando un 90% de los datos.

La siguiente figura representa una media de los valores de media y desviación obtenidos a lo largo de todos los meses de estudio, empleando diferentes valores de porcentajes de los meses utilizados en el entrenamiento (50%, 60%, 70% y 80%).

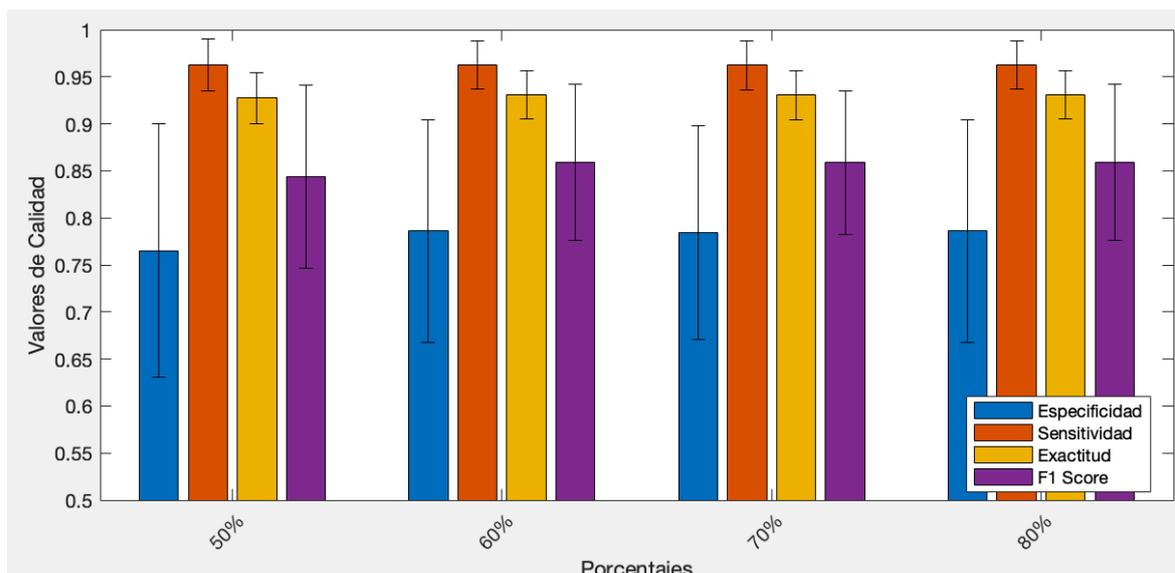


Figura 64. Parámetros de calidad medios atendiendo al porcentaje de train-test utilizado.

Las principales conclusiones que podemos sacar de los resultados son las siguientes:

- No se aprecian grandes diferencias en función del porcentaje de datos usados para entrenar. Esto da a entender que con pocos datos es posible entrenar unos modelos que muestran buena calidad. No obstante, esto tiene un límite (por ejemplo, caso del mes de septiembre de 2020).
- Las desviaciones vistas mes a mes son pequeñas, independientemente de cuántos datos se han usado para entrenar. Esto da a entender que los datos son suficientemente representativos y que los modelos obtenidos son fiables frente a situaciones reales.
- En general los valores de exactitud o f1-score son buenos o muy buenos.
- La sensibilidad a la lluvia resultados por encima del 90%, lo que representa valores muy buenos.
- La especificidad a la no lluvia muestra valores más pequeños, si bien en general altos (por encima del 75%). Hemos constatado que muchos de los eventos que las estaciones meteorológicas muestran como de no lluvia realmente son de lluvia en las medianías o las cumbres. Por otro lado, cuando los niveles de lluvia son muy bajos en Tafira Baja las estaciones no detectan nada, pero el radioenlace sí. Por este motivo la especificidad “aparenta” más baja que los demás parámetros de calidad.

CAPÍTULO 8.

CONCLUSIONES

En este capítulo se presentan las conclusiones obtenidas tras el desarrollo del TFG. Además, se exponen las posibles líneas de investigación para futuros trabajos relacionados con técnicas de aprendizaje máquina para detección de combinando radioenlaces satelitales y estaciones meteorológicas.

8.1 Conclusiones

A lo largo de este Trabajo de Fin de Grado se ha logrado un objetivo principal, la implementación de algoritmos de *Machine Learning (ML)* para la detección de precipitación de lluvia basándonos en los niveles de potencia involucrados en comunicaciones satelitales. 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 satelitales.
- El estudio de los datos para llevar a cabo un pre-procesado 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:

- Dado que las redes LSTM han generado muy buenos resultados, incluso mejor de lo esperado, no ha sido necesario emplear otros mecanismos de detección.

- Hemos visto que los modelos obtenidos son muy estables bajo diferentes condiciones de entrenamiento. Además, son fiables, esto es, en nuevas condiciones (p.e., nuevos meses) dan resultados similares a los expuestos.
- Dejamos sentadas las bases para implementar un detector que trabaje en tiempo real.

8.2 Líneas futuras

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

- Integración de nuevas estaciones meteorológicas a lo largo de la trayectoria radioenlace, 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.
- Aplicación de otros mecanismos de detección. Por ejemplo, basados en máquinas de vectores de soporte (SVM).

BIBLIOGRAFÍA

- [1] I. D. Guerra Moreno, “Algoritmos de Aprendizaje Aplicados a la Monitorización Meteorológica mediante Redes Inalámbricas de Comunicaciones Distribuidas,” 2017.
- [2] H. Messer, A. Zinevich, and P. Alpert, “Environmental monitoring by wireless communication networks,” *Science (80-.)*, vol. 312, no. 5774, p. 713, 2006.
- [3] A. Zinevich, P. Alpert, and H. Messer, “Estimation of rainfall fields using commercial microwave communication networks of variable density,” *Adv. Water Resour.*, vol. 31, no. 11, pp. 1470–1480, 2008.
- [4] A. Rayitsfeld, R. Samuels, A. Zinevich, U. Hadar, and P. Alpert, “Comparison of two methodologies for long term rainfall monitoring using a commercial microwave communication system,” *Atmos. Res.*, vol. 104–105, pp. 119–127, 2012.
- [5] H. Messer, A. Zinevich, and P. Alpert, “Environmental sensor networks using existing wireless communication systems for rainfall and wind velocity measurements,” *IEEE Instrum. Meas. Mag.*, vol. 15, no. 2, pp. 32–38, 2012.
- [6] N. David, P. Alpert, and H. Messer, “Novel method for fog monitoring using cellular networks infrastructures,” *Atmos. Meas. Tech. Discuss.*, vol. 5, no. 4, pp. 5725–5752, 2012.
- [7] G. Amaral *et al.*, “A Proposed Standard Fog Collector for Use in High-Elevation Regions,” *J. Petrol.*, vol. 369, no. 1, pp. 1689–1699, 2013.
- [8] I. Martínez Mateos, “Síntesis de Series Temporales de Atenuación Por Lluvia en Radioenlaces Horizontales,” 2014.
- [9] “Hispasat.” [Online]. Available: <https://www.hispasat.com/es/>. [Accessed: 06-Dec-2020].
- [10] R. Nathaly and O. Molina, “Estudio de los Factores Atmosféricos que Influyen En La Degradación De Calidad De Un Enlace Satelital,” 2010.

- [11] UIT, “UIT-R P, 525 Cálculo de la atenuación en el espacio libre,” vol. 02, no. 514, p. 10813, 2014.
- [12] UIT, “UIT-R P.530-16. Datos de propagación y métodos de predicción necesarios para el diseño de sistemas terrenales con visibilidad directa.,” vol. 14, 2012.
- [13] UIT, *UIT-R-P.618 Datos de propagación y métodos de predicción necesarios para el diseño de sistemas terrenales con visibilidad directa Serie P*, vol. 13. 2009.
- [14] L. Barros and M. A. Chamon, “Cálculo de enlace para sistemas de comunicação digital via satélite.,” 2004.
- [15] UIT, *UIT-R P.838-3 – Modelo de la atenuación específica debida a la lluvia para los métodos de predicción*. 2005, pp. 1–8.
- [16] UIT, “UIT-R P.387-6 Radio-frequency channel and block arrangements for fixed wireless systems operating in the 71-76 and 81-86 GHz bands F Series Fixed service,” vol. 2006, 2012.
- [17] A. Dissanayake, J. Allnutt, and F. Haidara, “A Prediction Model that Combines Rain Attenuation and Other Propagation Impairments Along Earth- Satellite Paths Asoka Dissanayake 1 , Jeremy Allnutt 2 , Fatim Haidara 3,” *Online J. Sp. Commun.*, vol. 45, no. 2, pp. 1546–1558, 1997.
- [18] UIT, “RECOMENDACIÓN UIT-R P . 531 Datos de propagación ionosférica y métodos de predicción requeridos para el diseño de servicios y sistemas de satélites Anexo 1,” vol. 13, pp. 1–21, 2007.
- [19] N. Molina Padrón, “DOCUMENTACIÓN DE RESULTADOS DEL PROYECTO VIMETRI-MAC/MAC 3.5b/065. Prototipo para la obtención de datos meteorológicos y radioeléctricos,” 2019.
- [20] R. B. Langley, “NMEA 0183: A GPS Receiver Interface Standard,” *GPS World*, vol. 6, no. 7. pp. 54–57, 1995.
- [21] “Rain Gauge RG13 , RG13H.”
- [22] R. Benítez, G. Escudero, and S. Kanaan, *Inteligencia Artificial Avanzada*. 2013.
- [23] D. O. Hebb, *The organization of behavior: a neuropsychological theory*. J. Wiley; Chapman & Hall, 1949.

- [24] S. Fuente González, “Técnicas de aprendizaje máquina para detección y cuantificación de lluvia mediante radioenlaces terrenos a frecuencia de gigahercios,” 2020.
- [25] V. Cherkassky and F. M. Mulier, *Learning from data: concepts, theory, and methods*. John Wiley & Sons, 2007.
- [26] “Semi-Supervised Learning...el gran desconocido - Think Big Empresas.” [Online]. Available: <https://empresas.blogthinkbig.com/semi-supervised-learningel-grandesconocido/>. [Accessed: 18-Dec-2020].
- [27] H. He, Y. Bai, E. A. Garcia, and S. Li, “ADASYN: Adaptive synthetic sampling approach for imbalanced learning,” in *2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence)*, 2008, pp. 1322–1328.
- [28] H. He, Y. Bai, E. A. Garcia, and S. Li, “ADASYN: Adaptive synthetic sampling approach for imbalanced learning,” *Proc. Int. Jt. Conf. Neural Networks*, no. 3, pp. 1322–1328, 2008.
- [29] L. M. Núñez Vivero, “Segmentación de imágenes de resonancia magnética cerebral mediante redes neuronales artificiales convolucionales,” 2018.
- [30] “Redes Neuronales Recurrentes - Jordi TORRES.AI.” [Online]. Available: <https://torres.ai/redes-neuronales-recurrentes/>. [Accessed: 06-Dec-2020].
- [31] “Understanding LSTM Networks -- colah’s blog.” [Online]. Available: <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>. [Accessed: 15-Dec-2020].
- [32] X. Yao, “Evolutionary artificial neural networks.,” *International journal of neural systems*, vol. 4, no. 3. World Scientific Publishing Company , pp. 203–222, 21-Nov-1993.
- [33] J. Serra, *Image analysis and mathematical morphology*. Academic Press, Inc., 1983.
- [34] A. Bouchet, M. Brun, and V. Ballarin, “Morfología Matemática Difusa aplicada a la segmentación de angioografías retinales,” *Rev. Argentina Bioingeniería*, vol. 16, no. 1, pp. 7–10, 2010.

ANEXO

Tabla 13. Parámetros de calidad con 50 % de datos para el entrenamiento.

Exactitud	Sensibilidad	Especificidad	F1-Score	TN	FP	FN	TP
0,86	0,92	0,90	0,89	50377	4538	4164	25521
0,85	0,90	0,88	0,87	49347	5568	4439	25246
0,87	0,91	0,89	0,89	49863	5052	3838	25847
0,84	0,92	0,89	0,88	50511	4404	4840	24845
0,84	0,91	0,89	0,87	50143	4772	4757	24928
0,86	0,91	0,89	0,88	50012	4903	4245	25440
0,86	0,91	0,89	0,88	49900	5015	4169	25516
0,85	0,92	0,89	0,88	50345	4570	4461	25224
0,87	0,91	0,89	0,89	49863	5052	3859	25826
0,85	0,90	0,89	0,88	49617	5298	4306	25379
0,83	0,92	0,89	0,87	50433	4482	5079	24606
0,85	0,91	0,89	0,88	50104	4811	4598	25087
0,85	0,91	0,89	0,88	50065	4850	4498	25187
0,87	0,91	0,90	0,89	50140	4775	3998	25687
0,84	0,91	0,89	0,87	50042	4873	4803	24882
0,85	0,91	0,89	0,88	50156	4759	4499	25186
0,83	0,92	0,89	0,87	50333	4582	4921	24764
0,85	0,91	0,89	0,88	50086	4829	4404	25281
0,84	0,92	0,89	0,88	50557	4358	4622	25063
0,83	0,92	0,89	0,87	50794	4121	5048	24637
0,85	0,91	0,89	0,88	50194	4721	4309	25376
0,87	0,91	0,89	0,89	49727	5188	3775	25910
0,85	0,91	0,89	0,88	49928	4987	4489	25196
0,86	0,91	0,90	0,89	50166	4749	4107	25578
0,85	0,91	0,89	0,88	50210	4705	4593	25092
0,86	0,90	0,88	0,88	49258	5657	4199	25486
0,86	0,91	0,89	0,88	50017	4898	4260	25425
0,82	0,93	0,89	0,87	50925	3990	5222	24463
0,86	0,91	0,89	0,89	50176	4739	4188	25497
0,86	0,91	0,89	0,88	49811	5104	4206	25479
0,85	0,91	0,89	0,88	50159	4756	4458	25227
0,86	0,91	0,89	0,89	49984	4931	4045	25640
0,84	0,92	0,89	0,88	50665	4250	4810	24875
0,86	0,92	0,90	0,89	50361	4554	4204	25481
0,85	0,92	0,89	0,88	50316	4599	4376	25309

Exactitud	Sensibilidad	Especificidad	F1-Score	TN	FP	FN	TP
0,85	0,91	0,89	0,88	50222	4693	4560	25125
0,84	0,92	0,89	0,88	50508	4407	4883	24802
0,85	0,91	0,89	0,88	50109	4806	4440	25245
0,85	0,91	0,89	0,88	49897	5018	4457	25228
0,87	0,90	0,89	0,89	49508	5407	3866	25819
0,85	0,91	0,89	0,88	49836	5079	4366	25319
0,84	0,91	0,89	0,88	50152	4763	4706	24979
0,85	0,91	0,89	0,88	50120	4795	4508	25177
0,82	0,92	0,88	0,87	50277	4638	5222	24463
0,84	0,93	0,90	0,88	50899	4016	4849	24836
0,85	0,92	0,89	0,88	50292	4623	4577	25108
0,85	0,91	0,89	0,88	49813	5102	4569	25116
0,84	0,91	0,89	0,88	50213	4702	4633	25052
0,85	0,91	0,89	0,88	50056	4859	4464	25221
0,85	0,92	0,89	0,88	50280	4635	4376	25309
0,86	0,91	0,89	0,88	50203	4712	4261	25424
0,85	0,92	0,90	0,88	50470	4445	4427	25258
0,86	0,91	0,89	0,88	50088	4827	4281	25404
0,86	0,91	0,89	0,89	50178	4737	4183	25502
0,86	0,91	0,89	0,88	50215	4700	4286	25399
0,85	0,91	0,89	0,88	49816	5099	4457	25228
0,88	0,91	0,90	0,89	50092	4823	3626	26059
0,85	0,91	0,89	0,88	50010	4905	4512	25173
0,86	0,92	0,90	0,89	50312	4603	4102	25583
0,84	0,92	0,89	0,88	50488	4427	4706	24979
0,85	0,92	0,89	0,88	50472	4443	4547	25138
0,83	0,92	0,89	0,88	50724	4191	4984	24701
0,85	0,91	0,89	0,88	50039	4876	4418	25267
0,84	0,92	0,89	0,88	50352	4563	4843	24842
0,87	0,91	0,89	0,89	49723	5192	3924	25761
0,82	0,92	0,89	0,87	50741	4174	5276	24409
0,84	0,92	0,89	0,88	50510	4405	4830	24855
0,85	0,91	0,89	0,88	50057	4858	4409	25276
0,86	0,92	0,90	0,89	50390	4525	4134	25551
0,83	0,92	0,89	0,87	50434	4481	5120	24565
0,86	0,91	0,89	0,89	50008	4907	4103	25582
0,85	0,91	0,89	0,88	50164	4751	4329	25356

Exactitud	Sensibilidad	Especificidad	F1-Score	TN	FP	FN	TP
0,84	0,92	0,89	0,88	50437	4478	4613	25072
0,85	0,91	0,89	0,88	50203	4712	4576	25109
0,83	0,91	0,88	0,87	50035	4880	5081	24604
0,86	0,91	0,89	0,88	50091	4824	4250	25435
0,84	0,92	0,89	0,88	50430	4485	4808	24877
0,84	0,92	0,89	0,88	50464	4451	4750	24935
0,85	0,91	0,89	0,88	49901	5014	4336	25349
0,84	0,92	0,90	0,88	50790	4125	4673	25012
0,86	0,91	0,89	0,88	49838	5077	4078	25607
0,85	0,91	0,89	0,88	49986	4929	4413	25272
0,86	0,91	0,89	0,88	50071	4844	4244	25441
0,86	0,91	0,89	0,88	49974	4941	4183	25502
0,86	0,91	0,89	0,88	49832	5083	4250	25435
0,85	0,91	0,89	0,88	50051	4864	4538	25147
0,85	0,91	0,89	0,88	49952	4963	4397	25288
0,86	0,92	0,90	0,89	50252	4663	4050	25635
0,84	0,91	0,89	0,88	50086	4829	4676	25009
0,84	0,92	0,89	0,88	50556	4359	4684	25001
0,84	0,92	0,89	0,88	50255	4660	4618	25067
0,86	0,90	0,88	0,88	49326	5589	4282	25403
0,86	0,90	0,89	0,88	49553	5362	4294	25391
0,85	0,90	0,89	0,88	49639	5276	4328	25357
0,84	0,91	0,89	0,88	50210	4705	4714	24971
0,84	0,92	0,89	0,88	50568	4347	4723	24962
0,85	0,92	0,89	0,88	50367	4548	4369	25316
0,87	0,91	0,89	0,89	49760	5155	3897	25788
0,85	0,91	0,89	0,88	50110	4805	4594	25091
0,82	0,93	0,89	0,87	50832	4083	5451	24234

Tabla 14. Parámetros de calidad con 60 % de datos para el entrenamiento.

Exactitud	Sensibilidad	Especificidad	F1-Score	TN	FP	FN	TP
0,85	0,92	0,89	0,88	50317	4598	4560	25125
0,85	0,92	0,90	0,89	50765	4150	4353	25332
0,82	0,92	0,89	0,87	50670	4245	5288	24397
0,84	0,93	0,89	0,88	50850	4065	4895	24790
0,84	0,93	0,90	0,88	50804	4111	4664	25021
0,85	0,92	0,90	0,88	50703	4212	4529	25156
0,84	0,93	0,90	0,88	50933	3982	4879	24806
0,84	0,93	0,90	0,88	50887	4028	4827	24858
0,84	0,93	0,90	0,88	50900	4015	4856	24829
0,83	0,93	0,90	0,88	51044	3871	4984	24701
0,86	0,92	0,90	0,89	50661	4254	4256	25429
0,83	0,92	0,89	0,87	50519	4396	4954	24731
0,83	0,92	0,89	0,88	50714	4201	4959	24726
0,84	0,93	0,90	0,88	51077	3838	4676	25009
0,86	0,93	0,90	0,89	50838	4077	4209	25476
0,85	0,92	0,90	0,89	50670	4245	4326	25359
0,84	0,93	0,90	0,88	51020	3895	4836	24849
0,82	0,93	0,89	0,87	50819	4096	5429	24256
0,83	0,93	0,90	0,88	51206	3709	5111	24574
0,85	0,93	0,90	0,89	50895	4020	4346	25339
0,84	0,93	0,90	0,88	51102	3813	4767	24918
0,84	0,93	0,90	0,88	51205	3710	4850	24835
0,84	0,92	0,90	0,88	50765	4150	4612	25073
0,84	0,92	0,89	0,88	50741	4174	4841	24844
0,86	0,92	0,90	0,89	50523	4392	4089	25596
0,84	0,93	0,90	0,88	50978	3937	4685	25000
0,84	0,92	0,89	0,88	50369	4546	4820	24865
0,83	0,93	0,89	0,87	50867	4048	5105	24580
0,82	0,93	0,89	0,87	50904	4011	5223	24462
0,85	0,93	0,90	0,89	50887	4028	4449	25236
0,84	0,92	0,89	0,88	50633	4282	4832	24853
0,84	0,92	0,89	0,88	50421	4494	4691	24994
0,81	0,93	0,89	0,86	51153	3762	5721	23964
0,81	0,93	0,89	0,87	51051	3864	5551	24134
0,83	0,93	0,89	0,88	50959	3956	5032	24653
0,81	0,93	0,89	0,87	50926	3989	5591	24094

Exactitud	Sensibilidad	Especificidad	F1-Score	TN	FP	FN	TP
0,84	0,93	0,90	0,88	51116	3799	4755	24930
0,84	0,92	0,89	0,88	50543	4372	4673	25012
0,84	0,92	0,90	0,88	50786	4129	4739	24946
0,83	0,92	0,89	0,87	50414	4501	5107	24578
0,85	0,92	0,90	0,88	50774	4141	4553	25132
0,82	0,92	0,89	0,87	50750	4165	5221	24464
0,80	0,93	0,89	0,86	51081	3834	5800	23885
0,81	0,93	0,89	0,87	51063	3852	5556	24129
0,83	0,92	0,89	0,88	50504	4411	4910	24775
0,82	0,92	0,89	0,87	50716	4199	5250	24435
0,83	0,92	0,89	0,87	50682	4233	5011	24674
0,81	0,93	0,89	0,87	51057	3858	5513	24172
0,82	0,93	0,89	0,87	51172	3743	5298	24387
0,81	0,93	0,89	0,87	51250	3665	5576	24109
0,84	0,92	0,89	0,88	50548	4367	4877	24808
0,84	0,92	0,89	0,88	50763	4152	4866	24819
0,82	0,93	0,89	0,87	50976	3939	5361	24324
0,84	0,93	0,89	0,88	50831	4084	4849	24836
0,85	0,93	0,90	0,89	50966	3949	4374	25311
0,86	0,92	0,90	0,89	50652	4263	4125	25560
0,85	0,93	0,90	0,88	50863	4052	4568	25117
0,85	0,92	0,89	0,88	50492	4423	4540	25145
0,84	0,93	0,90	0,88	50882	4033	4734	24951
0,84	0,93	0,90	0,88	51004	3911	4734	24951
0,85	0,92	0,90	0,89	50682	4233	4392	25293
0,84	0,92	0,89	0,88	50620	4295	4745	24940
0,81	0,93	0,89	0,87	50869	4046	5494	24191
0,84	0,92	0,89	0,88	50543	4372	4710	24975
0,83	0,93	0,89	0,88	50997	3918	5105	24580
0,84	0,92	0,89	0,88	50662	4253	4767	24918
0,83	0,93	0,89	0,87	51019	3896	5154	24531
0,86	0,92	0,90	0,89	50346	4569	4239	25446
0,85	0,92	0,90	0,88	50755	4160	4511	25174
0,84	0,92	0,89	0,88	50721	4194	4898	24787
0,85	0,92	0,89	0,88	50519	4396	4537	25148
0,85	0,93	0,90	0,89	50954	3961	4441	25244
0,83	0,93	0,89	0,87	50804	4111	5049	24636

Exactitud	Sensibilidad	Especificidad	F1-Score	TN	FP	FN	TP
0,82	0,93	0,89	0,87	51016	3899	5242	24443
0,85	0,93	0,90	0,89	50843	4072	4349	25336
0,82	0,93	0,89	0,87	51153	3762	5302	24383
0,84	0,92	0,89	0,88	50650	4265	4876	24809
0,85	0,92	0,90	0,88	50686	4229	4534	25151
0,84	0,92	0,89	0,88	50745	4170	4768	24917
0,83	0,92	0,89	0,87	50348	4567	5019	24666
0,83	0,93	0,89	0,88	50925	3990	5068	24617
0,85	0,92	0,90	0,88	50646	4269	4434	25251
0,85	0,92	0,90	0,88	50610	4305	4528	25157
0,84	0,92	0,89	0,88	50668	4247	4815	24870
0,85	0,92	0,90	0,88	50710	4205	4580	25105
0,83	0,93	0,89	0,88	50896	4019	4923	24762
0,82	0,93	0,89	0,87	50887	4028	5310	24375
0,84	0,92	0,89	0,88	50524	4391	4722	24963
0,83	0,93	0,89	0,88	50893	4022	4908	24777
0,82	0,93	0,89	0,87	51007	3908	5407	24278
0,83	0,93	0,90	0,88	51165	3750	5092	24593
0,84	0,92	0,89	0,88	50564	4351	4748	24937
0,82	0,92	0,89	0,87	50780	4135	5322	24363
0,84	0,92	0,90	0,88	50698	4217	4613	25072
0,84	0,92	0,89	0,88	50610	4305	4775	24910
0,83	0,93	0,89	0,88	51080	3835	5071	24614
0,86	0,92	0,90	0,89	50512	4403	4015	25670
0,85	0,92	0,89	0,88	50439	4476	4475	25210
0,84	0,92	0,89	0,88	50786	4129	4854	24831
0,83	0,93	0,89	0,88	51022	3893	5052	24633

Tabla 15. Parámetros de calidad con 70 % de datos para el entrenamiento.

Exactitud	Sensibilidad	Especificidad	F1-Score	TN	FP	FN	TP
0,84	0,92	0,90	0,88	50779	4136	4652	25033
0,85	0,92	0,89	0,88	50603	4312	4601	25084
0,83	0,91	0,89	0,87	50214	4701	5005	24680
0,82	0,92	0,89	0,87	50741	4174	5290	24395
0,85	0,92	0,89	0,88	50433	4482	4461	25224
0,84	0,92	0,89	0,88	50351	4564	4808	24877
0,85	0,92	0,89	0,88	50485	4430	4587	25098
0,85	0,92	0,90	0,89	50519	4396	4362	25323
0,82	0,91	0,88	0,87	50242	4673	5322	24363
0,83	0,92	0,89	0,87	50715	4200	5140	24545
0,83	0,93	0,89	0,88	50955	3960	5075	24610
0,83	0,93	0,89	0,88	50915	4000	5022	24663
0,84	0,92	0,90	0,88	50714	4201	4676	25009
0,81	0,93	0,89	0,87	51223	3692	5638	24047
0,84	0,93	0,89	0,88	50854	4061	4828	24857
0,85	0,92	0,89	0,88	50447	4468	4536	25149
0,81	0,93	0,89	0,87	51129	3786	5646	24039
0,83	0,91	0,89	0,87	50241	4674	4918	24767
0,85	0,92	0,89	0,88	50535	4380	4566	25119
0,84	0,92	0,89	0,87	50304	4611	4860	24825
0,84	0,92	0,89	0,88	50435	4480	4742	24943
0,85	0,92	0,90	0,88	50712	4203	4568	25117
0,82	0,92	0,88	0,87	50464	4451	5305	24380
0,82	0,92	0,89	0,87	50783	4132	5375	24310
0,85	0,92	0,90	0,88	50677	4238	4574	25111
0,81	0,93	0,88	0,86	50835	4080	5659	24026
0,84	0,92	0,89	0,88	50385	4530	4720	24965
0,84	0,93	0,90	0,88	50886	4029	4799	24886
0,81	0,92	0,88	0,86	50647	4268	5520	24165
0,84	0,92	0,89	0,88	50549	4366	4797	24888
0,86	0,92	0,90	0,89	50329	4586	4123	25562
0,83	0,92	0,89	0,87	50682	4233	5036	24649
0,86	0,91	0,89	0,88	49969	4946	4115	25570
0,82	0,93	0,89	0,87	50804	4111	5237	24448
0,83	0,92	0,89	0,87	50734	4181	5120	24565

Exactitud	Sensibilidad	Especificidad	F1-Score	TN	FP	FN	TP
0,85	0,92	0,90	0,88	50708	4207	4500	25185
0,85	0,92	0,90	0,88	50615	4300	4471	25214
0,83	0,92	0,89	0,88	50755	4160	4916	24769
0,84	0,92	0,89	0,87	50372	4543	4867	24818
0,86	0,92	0,90	0,89	50303	4612	4090	25595
0,86	0,92	0,90	0,89	50283	4632	4049	25636
0,83	0,93	0,89	0,88	51047	3868	5126	24559
0,85	0,92	0,90	0,88	50619	4296	4456	25229
0,83	0,93	0,89	0,88	50892	4023	5009	24676
0,85	0,92	0,89	0,88	50271	4644	4542	25143
0,82	0,93	0,89	0,87	51059	3856	5416	24269
0,84	0,92	0,89	0,88	50290	4625	4773	24912
0,83	0,93	0,89	0,88	50920	3995	5065	24620
0,81	0,92	0,88	0,86	50644	4271	5532	24153
0,84	0,92	0,89	0,88	50592	4323	4725	24960
0,82	0,93	0,89	0,87	50843	4072	5405	24280
0,85	0,92	0,90	0,88	50721	4194	4534	25151
0,83	0,92	0,89	0,87	50456	4459	5025	24660
0,80	0,93	0,88	0,86	50820	4095	5855	23830
0,84	0,92	0,89	0,88	50565	4350	4850	24835
0,83	0,92	0,89	0,87	50583	4332	5146	24539
0,84	0,92	0,89	0,88	50609	4306	4663	25022
0,84	0,91	0,89	0,88	50191	4724	4763	24922
0,85	0,92	0,90	0,88	50713	4202	4593	25092
0,83	0,93	0,89	0,88	50808	4107	5029	24656
0,85	0,93	0,90	0,89	50885	4030	4491	25194
0,85	0,92	0,90	0,89	50726	4189	4463	25222
0,84	0,92	0,89	0,88	50301	4614	4793	24892
0,82	0,93	0,89	0,87	50904	4011	5292	24393
0,84	0,92	0,90	0,88	50792	4123	4716	24969
0,84	0,93	0,89	0,88	50813	4102	4805	24880
0,81	0,93	0,89	0,86	51019	3896	5669	24016
0,82	0,93	0,89	0,87	50948	3967	5280	24405
0,84	0,92	0,89	0,88	50547	4368	4884	24801
0,84	0,91	0,89	0,88	50155	4760	4742	24943
0,82	0,92	0,89	0,87	50656	4259	5293	24392
0,82	0,93	0,89	0,87	50882	4033	5324	24361

Exactitud	Sensibilidad	Especificidad	F1-Score	TN	FP	FN	TP
0,83	0,91	0,88	0,87	50191	4724	5035	24650
0,85	0,92	0,90	0,88	50720	4195	4563	25122
0,83	0,93	0,89	0,88	50869	4046	4997	24688
0,82	0,93	0,89	0,87	50922	3993	5487	24198
0,84	0,93	0,90	0,88	50816	4099	4741	24944
0,85	0,92	0,89	0,88	50629	4286	4600	25085
0,84	0,92	0,89	0,88	50315	4600	4661	25024
0,82	0,93	0,89	0,87	51107	3808	5395	24290
0,84	0,93	0,89	0,88	50833	4082	4855	24830
0,84	0,92	0,89	0,88	50577	4338	4715	24970
0,81	0,93	0,89	0,87	50967	3948	5593	24092
0,82	0,93	0,89	0,87	51079	3836	5343	24342
0,85	0,92	0,89	0,88	50373	4542	4446	25239
0,82	0,93	0,89	0,87	50842	4073	5306	24379
0,84	0,92	0,89	0,88	50508	4407	4898	24787
0,83	0,91	0,88	0,87	50185	4730	5168	24517
0,82	0,93	0,89	0,87	50898	4017	5408	24277
0,82	0,93	0,89	0,87	51254	3661	5445	24240
0,83	0,92	0,89	0,87	50534	4381	4979	24706
0,82	0,92	0,89	0,87	50710	4205	5314	24371
0,85	0,93	0,90	0,89	50831	4084	4376	25309
0,84	0,92	0,89	0,88	50790	4125	4871	24814
0,85	0,92	0,89	0,88	50360	4555	4387	25298
0,85	0,92	0,90	0,88	50576	4339	4502	25183
0,83	0,92	0,89	0,87	50409	4506	5033	24652
0,82	0,91	0,88	0,87	50237	4678	5232	24453
0,85	0,92	0,90	0,88	50647	4268	4536	25149
0,83	0,93	0,89	0,88	50894	4021	4981	24704

Tabla 16. Parámetros de calidad con 80 % de datos para el entrenamiento.

Exactitud	Sensibilidad	Especificidad	F1-Score	TN	FP	FN	TP
0,85	0,92	0,89	0,88	50317	4598	4560	25125
0,85	0,92	0,90	0,89	50765	4150	4353	25332
0,82	0,92	0,89	0,87	50670	4245	5288	24397
0,84	0,93	0,89	0,88	50850	4065	4895	24790
0,84	0,93	0,90	0,88	50804	4111	4664	25021
0,85	0,92	0,90	0,88	50703	4212	4529	25156
0,84	0,93	0,90	0,88	50933	3982	4879	24806
0,84	0,93	0,90	0,88	50887	4028	4827	24858
0,84	0,93	0,90	0,88	50900	4015	4856	24829
0,83	0,93	0,90	0,88	51044	3871	4984	24701
0,86	0,92	0,90	0,89	50661	4254	4256	25429
0,83	0,92	0,89	0,87	50519	4396	4954	24731
0,83	0,92	0,89	0,88	50714	4201	4959	24726
0,84	0,93	0,90	0,88	51077	3838	4676	25009
0,86	0,93	0,90	0,89	50838	4077	4209	25476
0,85	0,92	0,90	0,89	50670	4245	4326	25359
0,84	0,93	0,90	0,88	51020	3895	4836	24849
0,82	0,93	0,89	0,87	50819	4096	5429	24256
0,83	0,93	0,90	0,88	51206	3709	5111	24574
0,85	0,93	0,90	0,89	50895	4020	4346	25339
0,84	0,93	0,90	0,88	51102	3813	4767	24918
0,84	0,93	0,90	0,88	51205	3710	4850	24835
0,84	0,92	0,90	0,88	50765	4150	4612	25073
0,84	0,92	0,89	0,88	50741	4174	4841	24844
0,86	0,92	0,90	0,89	50523	4392	4089	25596
0,84	0,93	0,90	0,88	50978	3937	4685	25000
0,84	0,92	0,89	0,88	50369	4546	4820	24865
0,83	0,93	0,89	0,87	50867	4048	5105	24580
0,82	0,93	0,89	0,87	50904	4011	5223	24462
0,85	0,93	0,90	0,89	50887	4028	4449	25236
0,84	0,92	0,89	0,88	50633	4282	4832	24853
0,84	0,92	0,89	0,88	50421	4494	4691	24994
0,81	0,93	0,89	0,86	51153	3762	5721	23964
0,81	0,93	0,89	0,87	51051	3864	5551	24134
0,83	0,93	0,89	0,88	50959	3956	5032	24653
0,81	0,93	0,89	0,87	50926	3989	5591	24094

Exactitud	Sensibilidad	Especificidad	F1-Score	TN	FP	FN	TP
0,84	0,93	0,90	0,88	51116	3799	4755	24930
0,84	0,92	0,89	0,88	50543	4372	4673	25012
0,84	0,92	0,90	0,88	50786	4129	4739	24946
0,83	0,92	0,89	0,87	50414	4501	5107	24578
0,85	0,92	0,90	0,88	50774	4141	4553	25132
0,82	0,92	0,89	0,87	50750	4165	5221	24464
0,80	0,93	0,89	0,86	51081	3834	5800	23885
0,81	0,93	0,89	0,87	51063	3852	5556	24129
0,83	0,92	0,89	0,88	50504	4411	4910	24775
0,82	0,92	0,89	0,87	50716	4199	5250	24435
0,83	0,92	0,89	0,87	50682	4233	5011	24674
0,81	0,93	0,89	0,87	51057	3858	5513	24172
0,82	0,93	0,89	0,87	51172	3743	5298	24387
0,81	0,93	0,89	0,87	51250	3665	5576	24109
0,84	0,92	0,89	0,88	50548	4367	4877	24808
0,84	0,92	0,89	0,88	50763	4152	4866	24819
0,82	0,93	0,89	0,87	50976	3939	5361	24324
0,84	0,93	0,89	0,88	50831	4084	4849	24836
0,85	0,93	0,90	0,89	50966	3949	4374	25311
0,86	0,92	0,90	0,89	50652	4263	4125	25560
0,85	0,93	0,90	0,88	50863	4052	4568	25117
0,85	0,92	0,89	0,88	50492	4423	4540	25145
0,84	0,93	0,90	0,88	50882	4033	4734	24951
0,84	0,93	0,90	0,88	51004	3911	4734	24951
0,85	0,92	0,90	0,89	50682	4233	4392	25293
0,84	0,92	0,89	0,88	50620	4295	4745	24940
0,81	0,93	0,89	0,87	50869	4046	5494	24191
0,84	0,92	0,89	0,88	50543	4372	4710	24975
0,83	0,93	0,89	0,88	50997	3918	5105	24580
0,84	0,92	0,89	0,88	50662	4253	4767	24918
0,83	0,93	0,89	0,87	51019	3896	5154	24531
0,86	0,92	0,90	0,89	50346	4569	4239	25446
0,85	0,92	0,90	0,88	50755	4160	4511	25174
0,84	0,92	0,89	0,88	50721	4194	4898	24787
0,85	0,92	0,89	0,88	50519	4396	4537	25148
0,85	0,93	0,90	0,89	50954	3961	4441	25244
0,83	0,93	0,89	0,87	50804	4111	5049	24636

Exactitud	Sensibilidad	Especificidad	F1-Score	TN	FP	FN	TP
0,82	0,93	0,89	0,87	51016	3899	5242	24443
0,85	0,93	0,90	0,89	50843	4072	4349	25336
0,82	0,93	0,89	0,87	51153	3762	5302	24383
0,84	0,92	0,89	0,88	50650	4265	4876	24809
0,85	0,92	0,90	0,88	50686	4229	4534	25151
0,84	0,92	0,89	0,88	50745	4170	4768	24917
0,83	0,92	0,89	0,87	50348	4567	5019	24666
0,83	0,93	0,89	0,88	50925	3990	5068	24617
0,85	0,92	0,90	0,88	50646	4269	4434	25251
0,85	0,92	0,90	0,88	50610	4305	4528	25157
0,84	0,92	0,89	0,88	50668	4247	4815	24870
0,85	0,92	0,90	0,88	50710	4205	4580	25105
0,83	0,93	0,89	0,88	50896	4019	4923	24762
0,82	0,93	0,89	0,87	50887	4028	5310	24375
0,84	0,92	0,89	0,88	50524	4391	4722	24963
0,83	0,93	0,89	0,88	50893	4022	4908	24777
0,82	0,93	0,89	0,87	51007	3908	5407	24278
0,83	0,93	0,90	0,88	51165	3750	5092	24593
0,84	0,92	0,89	0,88	50564	4351	4748	24937
0,82	0,92	0,89	0,87	50780	4135	5322	24363
0,84	0,92	0,90	0,88	50698	4217	4613	25072
0,84	0,92	0,89	0,88	50610	4305	4775	24910
0,83	0,93	0,89	0,88	51080	3835	5071	24614
0,86	0,92	0,90	0,89	50512	4403	4015	25670
0,85	0,92	0,89	0,88	50439	4476	4475	25210
0,84	0,92	0,89	0,88	50786	4129	4854	24831
0,83	0,93	0,89	0,88	51022	3893	5052	24633

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 de lluvia mediante un radioenlace satelital 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 o partes de estos, 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 de 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 desensamblado de los programas o partes de estos. 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 o partes de estos, 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 pre-procesado, 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). Tanto los datos meteorológicos, como los datos del radioenlace, han sido obtenidos a partir de una plataforma web gestionada por el propio proyecto VIMETRI-MAC. Los datos expuestos en este TFG 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 Elena Encinas Pérez, autora del presente Trabajo Fin de Grado, declara que:

El Trabajo de Fin de Grado con título “*Diseño de un sistema de observación meteorológica combinando estaciones meteorológicas y radioenlaces satelitales*”, 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 *dieciocho mil trescientos cincuenta y un euros y ochenta y tres céntimos (18.351,83 €)*, 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.

Autora del TFG:

Elena Encinas Pérez

Las Palmas de Gran Canaria, a 28 de diciembre de 2020

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 de 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 × 5 días × 4 semanas × 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 *quince 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 4 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.

Elementos Hardware	Coste	Valor residual (3 años)	Amortización (11 meses)
Ordenador Quad-Core Intel Core i5 a 1,4GHz (4 núcleos)	1.377 €	300 €	119,67 €
Ordenador Intel® Core™ i7 a 3 GHz, 16 GB de memoria RAM	1.299 €	350 €	105,44 €
Total, Hardware			225,11 €

Por tanto, el coste total del hardware empleado ha sido de *doscientos veinticinco euros con once céntimos (225,11 €)*.

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	Nº de elementos	Coste	Valor residual (3 años)	Amortización (11 meses)
MatLab® versión 2019b, 64 bits Educativa	1	250 €/año	0 €	83,33 €
Sistema Operativo macOS Mojave versión 10.14.6	1	0,00 €	0 €	0,00 €
Sistema Operativo macOS Catalina versión 10.15.4	1	0,00 €	0 €	0,00 €
Paquete Microsoft Office para Mac versión 16.39	2	7,00 €/mes	0 €	28,00 €
Bases de datos de AEMET	---	0,00 €	0 €	0,00 €
Total, Software				111,33 €

Por tanto, el coste total del software empleado ha sido de *ciento once euros con treinta y tres céntimos (111,33 €)*.

C-3. Costes de redacción del Trabajo de 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	225,11 €
Amortización del material software	111,33 €
Total	15.336,44 €

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:

$$R = 0,07 \cdot P \cdot C_h = 0,07 \cdot 15.336,44 \cdot 1 = 1.073,55 \text{ €}$$

Por lo tanto, el importe final de redacción del TFG asciende a la cantidad de *mil setenta y tres euros con cincuenta y cinco céntimos (1.073,55 €)*.

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	<i>15.000,00 €</i>
Amortización del material hardware	<i>225,11 €</i>
Amortización del material software	<i>111,33 €</i>
Redacción del TFG	<i>1.073,55 €</i>
Total	<i>16.409,99 €</i>

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:

$$V = 0,006 \cdot P \cdot C = 0,006 \cdot 16.409,99 \cdot 1,00 = 98,46 \text{ €}$$

Finalmente, los costes por derecho de visado del TFG ascienden a *noventa y ocho euros y cuarenta y seis céntimos (98,46€)*

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	336,44 €
- Material de hardware: 225,11 €	
- Material de software: 111,33 €	
C-3. Costes de redacción del Trabajo de Fin de Grado	1.073,55 €
C-4. Derechos de visado del COITT	98,46 €
C-5. Gastos de tramitación y envío	9,00 €
Presupuesto total antes de impuestos:	16.517,45 €
Aplicación de Impuestos (7% IGIC):	1.156,22 €
PRESUPUESTO TOTAL:	17.673,67 €

El presupuesto total, con los impuestos incluidos, asciende a la cantidad de *diecisiete mil seiscientos setenta y tres euros y sesenta y siete céntimos (17.673,67 €)*.

Fdo.: Elena Encinas Pérez

Las Palmas de Gran Canaria, a 28 de diciembre de 2020

