



MODELO DE GESTIÓN DE INFORMACIÓN METEOROLÓGICA (\mathcal{M}^3)

PARA CALIBRACIÓN Y VALIDACIÓN DE ALGORITMOS, DETECCIÓN
DE RIESGOS METEOROLÓGICOS Y OTRAS APLICACIONES

Ángel Manuel Guerrero Higuera

TESIS DOCTORAL

Junio 2017

Directores:

Eduardo García Ortega

Laura López Campano

Vicente Matellán Olivera

Programa de doctorado: *tecnologías avanzadas de producción*

Escuela de ingenierías industrial e informática

Universidad de León

Dedicado a mis padres

Agradecimientos

En primer lugar, quiero agradecer a mis directores, Eduardo, Laura y Vicente, por su orientación y paciencia durante la realización de esta tesis. Huelga decir que sin ellos, no habría sido posible.

También quiero agradecer su colaboración y apoyo a todos los miembros del Grupo de Física de la Atmósfera de la Universidad de León, con especial cariño a mis compañeros y amigos Andrés Merino, Roberto Weigand, Estíbaliz Gascón, Lucía Hermida, Marian Arias, Sergio Fernández, José Luis Marcos, Jorge Falagán, Rafael Posada y María Fernández. Además, quiero agradecer explícitamente al director del grupo, José Luis Sánchez, por darme la oportunidad de cumplir mi sueño de dejar la empresa privada y acceder al mundo académico.

No quiero olvidarme de mis compañeros del Grupo de Robótica de la Universidad de León, en especial Jesús Balsa, Gonzalo Esteban, Francisco Javier Rodríguez Lera, Miguel Ángel Conde, Francisco J. Rodríguez y Camino Fernández, con quienes he compartido estos últimos meses.

Mi agradecimiento también a Cristina Gutiérrez por su ayuda y consejos en la redacción y revisión de esta memoria.

Por último, quiero acordarme de mi familia y amigos. En especial de mis padres, Manuel y Ángela, que siempre me animaron a trabajar duro por aquello que me propusiera. También de mi hijo Manuel, cuya llegada ha supuesto el impulso final que necesitaba para cerrar este trabajo.

Resumen

La humanidad ha mostrado interés por la atmósfera y los fenómenos que en ella se producen a lo largo de la historia. Esto se debe a la gran influencia que tienen ciertos eventos atmosféricos en múltiples actividades de vital importancia: agricultura, transporte de personas o mercancías, etc.; sin mencionar otros como huracanes, lluvias torrenciales o fuertes nevadas, cuyos efectos pueden ser devastadores e incluso suponer la pérdida de vidas humanas. Todo esto ha convertido a la meteorología, y otras ciencias descriptivas como la climatología, en una de las áreas de conocimiento más estudiadas.

A largo del siglo XIX, debido al creciente número de personas dedicadas a la meteorología y al desarrollo de los sistemas de observación, las actividades teóricas para interpretar y tratar de explicar los fenómenos atmosféricos, así como las actividades de predicción, se desarrollaron notablemente. En la segunda mitad del siglo XX se produjo otro gran avance, la computación, que permitió abordar problemas hasta entonces intratables utilizando predicción numérica.

El desarrollo y la especialización de los sistemas de observación permite hoy día a los meteorólogos tener acceso a gran cantidad de información. Existe una gran variedad de fuentes de información: estaciones meteorológicas en superficie, radiosondeos, radares meteorológicos, satélites, boyas marítimas, mediciones tomadas desde barcos o aviones, etc. Muchas empresas, organismos de prevención de riesgos, servicios meteorológicos y grupos de investigación, recopilan información meteorológica que cada cual utiliza según sus propios intereses.

Además de datos de observación, en la actualidad se dispone de un volumen importante de información pronosticada. Gracias a los modelos de predicción numérica y al desarrollo de la computación de altas prestaciones, son muchas las organizaciones y centros de investigación que disponen de sus propias predicciones. Toda esta información constituye una densa base de datos distribuida, que se actualiza de manera continua y que es transmitida y/o almacenada en distintos formatos.

Esta tesis plantea la posibilidad de unificar toda esta información, distribuida entre diferentes organizaciones, en un Sistema de Gestión de Datos Meteorológicos

definido como un “conjunto de aplicaciones que permiten la recopilación y gestión informatizada de datos meteorológicos”. En concreto, hablaremos de Sistema de Gestión de Datos Meteorológicos abierto, si la información que contiene se difunde sin restricción alguna a todo aquel que lo requiera.

Este trabajo propone, describe y evalúa un modelo de gestión de información meteorológica, denominado \mathcal{M}^3 , diseñado para construir Sistemas de Gestión de Datos Meteorológicos. La especificación del modelo \mathcal{M}^3 incluye 6 elementos principales, $\langle \mathcal{P}, \mathcal{V}, \mathcal{D}, \mathcal{R}, \mathcal{U}, \mathcal{S} \rangle$, donde: \mathcal{P} representa un conjunto de localizaciones geográficas; \mathcal{V} , un conjunto de variables meteorológicas; \mathcal{D} , el repositorio de datos meteorológicos; \mathcal{R} , un conjunto de perfiles de acceso para acceder al repositorio; \mathcal{U} , un conjunto de usuarios; y \mathcal{S} , una serie de servicios que incluyen operaciones para introducir (\mathcal{S}_{in}), consultar (\mathcal{S}_{out}), modificar (\mathcal{S}_{mod}) y eliminar (\mathcal{S}_{del}) información en el repositorio, además de controlar el acceso (\mathcal{S}_{ctrl}) de los usuarios.

Para poder evaluar la validez del modelo, se ha desarrollado la plataforma GFA- \mathcal{M}^3 cuyo objetivo principal es la gestión unificada de toda la información meteorológica recopilada por el Grupo de Física de la Atmósfera de la Universidad de León en el marco de sus investigaciones. La plataforma incluye, entre otros, datos de las siguientes fuentes de información: datos de satélite y radar meteorológicos, obtenidos por el propio Grupo de Física de la Atmósfera; predicciones del modelo de predicción numérica Weather Research and Forecasting, ejecutado en el cluster de cálculo paralelo de la Fundación del Centro de Supercomputación de Castilla y León; datos de las estaciones meteorológicas de los sistemas de información hidrológica de las cuencas de los ríos Duero y Ebro, del Instituto Tecnológico Agrario de Castilla y León, y de observadores voluntarios.

Para demostrar la utilidad de la plataforma, se han evaluado una serie de aplicaciones y herramientas desarrollados gracias a la información contenida en la plataforma GFA- \mathcal{M}^3 , entre los que se destacan los siguientes: identificación y seguimiento de eventos meteorológicos, en concreto heladas, precipitación y nevadas; extracción de información para asimilación, en modelos de predicción numérica; desarrollo de herramientas como la que permite predecir tormentas con presencia de granizo a partir de la información del modelo Weather Research and Forecasting, la que calcula la probabilidad de granizo a partir de información de satélite, o el sistema de información meteorológica para un vehículo; calibración y validación de algoritmos como el cálculo de la precipitación estimada por el algoritmo Multi-sensor Precipitation Estimate o por el modelo de predicción numérica Weather Research and Forecasting, o el cálculo de la probabilidad de granizo; por último, también se menciona la posibilidad de utilizar esta información como herramienta de innovación docente.

Esta tesis, abre una serie de posibilidades para el futuro en relación a la plataforma GFA- \mathcal{M}^3 : ampliar el número de fuentes de datos que aportan información a la plataforma, lo que permitirá no solamente aumentar el volumen de datos,

sino que además incrementa el área de estudio; ampliar el número de formatos en los cuales puede obtenerse la información; desarrollar nuevas aplicaciones que utilicen los datos de la plataforma; o construir un portal de acceso a la información (Open Data) que simplifique el acceso a los datos.

Índice general

Agradecimientos	V
Resumen	VII
Índice de figuras	XV
Índice de tablas	XVII
Índice de listados	XIX
Índice de algoritmos	XXI
1. Introducción	1
1.1. Diseño de un SGDM	3
1.1.1. Acceso y disponibilidad de la información	3
1.1.2. Obtención de datos meteorológicos	4
1.1.3. Formatos comúnmente utilizados en meteorología	14
1.1.4. Modelos de gestión de información	20
1.2. Hipótesis, resultados esperados y métricas	23
1.2.1. Hipótesis	23
1.2.2. Resultados esperados	25
1.2.3. Métricas	26
1.3. Estructura de la memoria	26
2. Gestión de datos meteorológicos	29
2.1. Sistemas de gestión de datos existentes	29
2.1.1. SGDC	29
2.1.2. GIS	31
2.1.3. Discusión	31
2.2. Formalización del modelo	32

2.2.1.	Localizaciones geográficas	32
2.2.2.	Variables meteorológicas	33
2.2.3.	Repositorio de datos meteorológicos	33
2.2.4.	Perfiles de acceso	34
2.2.5.	Usuarios	34
2.2.6.	Servicios de acceso	35
2.3.	GFA- \mathcal{M}^3	40
2.3.1.	Repositorio de datos meteorológicos	43
2.3.2.	Input services	44
2.3.3.	Output services	45
3.	Clasificación de datos meteorológicos	47
3.1.	Planificación	47
3.2.	Análisis de los diferentes orígenes de información	48
3.2.1.	Datos de observación	48
3.2.2.	Datos de predicción	54
3.3.	Taxonomía	60
4.	GFA-\mathcal{M}^3	65
4.1.	Infraestructura software	65
4.1.1.	Repositorio de datos meteorológicos	65
4.1.2.	Input services	74
4.1.3.	Output services	89
4.2.	Infraestructura hardware	94
5.	Aplicaciones	97
5.1.	Identificación y seguimiento de riesgos meteorológicos en tiempo real	97
5.1.1.	Seguimiento de heladas	98
5.1.2.	Seguimiento de precipitaciones	98
5.1.3.	Seguimiento de nevadas	99
5.2.	Asimilación de datos en modelos NWP	100
5.2.1.	LITTLE_R	102
5.2.2.	Conversión de las salidas del GFA- \mathcal{M}^3 a LITTLE_R	104
5.3.	Calibración y validación de algoritmos	105
5.3.1.	WRF precipitation evaluation tool	106
5.3.2.	The Satellite Precipitation Evaluation Application	107
5.3.3.	Hail Precipitation Evaluation Application	107
5.4.	Desarrollo de Aplicaciones	108
5.4.1.	Caracterización de tormentas severas con presencia de granizo	109
5.4.2.	Probabilidad de granizo	113
5.4.3.	Sistema de información meteorológica para vehículos	119
5.5.	Educación académica	122

6. Conclusiones	125
6.1. Evaluación de las hipótesis	125
6.1.1. Evaluación de la hipótesis 1	125
6.1.2. Evaluación de la hipótesis 2	126
6.1.3. Evaluación de la hipótesis 3	127
6.1.4. Evaluación de la hipótesis 4	127
6.1.5. Desarrollo de aplicaciones	128
6.2. Conclusiones	128
6.3. Trabajos futuros	129
6.4. Publicaciones	130
6.4.1. Publicaciones	130
6.4.2. Participación en congresos internacionales	130
6.4.3. Otras contribuciones	132
A. Otras contribuciones	133
A.1. Publicaciones	133
A.2. Proyectos	134
Bibliografía	135
Acrónimos	145

Índice de figuras

1.1. Elementos fundamentales de un radar	8
1.2. Granizómetro impactado.	9
2.1. Diagrama de casos de uso.	41
2.2. Relación entre los servicios de entrada/salida y el repositorio.	42
2.3. Diagrama de componentes.	43
3.1. Puntos de observación de la RdOCyL.	50
3.2. Radiómetro modelo MP3000A.	53
3.3. Receptor MSG e imagen procesada por Nubes 4.	54
3.4. Ubicación y alcance del radar meteorológico del GFA.	55
3.5. Redes de granizómetros	56
3.6. Coordenadas verticales sigma normalizadas.	58
3.7. Dominios 1, 2 y 3 para predicciones tipo I y II	60
4.1. Diagrama E/R del repositorio.	66
4.2. Entidad Place	67
4.3. Entidad Source	68
4.4. Relación entre Source y Place.	68
4.5. Entidad Variable	69
4.6. Relación entre Place y Variable.	72
4.7. Esquema relacional del repositorio.	73
4.8. Arquitectura del módulo <i>csv2sql</i>	75
4.9. Arquitectura del módulo <i>nc2sql</i>	81
4.10. Arquitectura de PyWRFScheduler.	83
4.11. Jerarquía de las entidades que definen el flujo de trabajo.	84
4.12. Arquitectura de <i>http2sql</i>	87
4.13. Arquitectura de <i>hrit2sql</i>	88
4.14. Formulario de consulta del sitio web del GFA.	89
4.15. Módulos de los servicios consulta y sus interrelaciones.	91

4.16. Arquitectura del módulo sql2csv.	91
4.17. Arquitectura de sql2nc.	92
4.18. Arquitectura del módulo csv2png.	93
4.19. Arquitectura de csv2kml.	93
5.1. Evolución de la temperatura	99
5.2. Evolución de la precipitación	100
5.3. Precipitación en forma de nieve	101
5.4. Evaluación de las predicciones del modelo WRF	106
5.5. Intensidad de precipitación calculada mediante MPE.	107
5.6. Precipitación observada	108
5.7. Comparativa de la intensidad de precipitación	108
5.8. Imagen de radar y probabilidad de granizo	109
5.9. WVFD, L, H y CI	111
5.10. CAPE, SRH_{0-3} , SI y EHI	112
5.11. Máscaras de convección y granizo	115
5.12. Sist. de información meteorológica para un vehículo	122
5.13. Interfaz del sist. de información meteorológica para un vehículo	123

Índice de tablas

3.1.	Principales parámetros que definen la física del modelo WRF . . .	59
3.2.	Caracterización de los orígenes de datos analizados.	61
3.3.	Escalas cualitativa de periodicidad, ficheros recibidos por periodo y tamaño medio por fichero.	62
3.4.	Caracterización cualitativa de los orígenes de datos de observación analizadas.	62
3.5.	Taxonomía general de las fuentes de información meteorológica. . .	62
3.6.	Caracterización de las herramientas de recopilación de datos. . . .	63
3.7.	Taxonomía de los formatos de información meteorológica.	64
4.1.	Listado parcial (8 elementos) de ejemplares de la entidad Place. . .	68
4.2.	Listado de ejemplares de la entidad Source. La columna de la derecha muestra el número de ejemplares de la entidad Place asociados a cada ejemplar de Source.	69
4.3.	Listado de ejemplares de la entidad Variable.	69
4.4.	Listado parcial de medidas (10 elementos).	72
4.5.	Medidas almacenadas por cada origen de datos (Source).	73
4.6.	Argumentos de create-map. Los argumentos marcados en negrita son obligatorios.	90
4.7.	Información general del servidor <i>hawking</i>	95
4.8.	Información sobre la CPU del servidor <i>hawking</i>	95
4.9.	Información sobre la memoria y capacidad de almacenamiento del servidor <i>hawking</i>	95
4.10.	Herramientas instaladas en el servidor <i>hawking</i>	95
5.1.	Formato del encabezado de un report. Los campos marcados en negrita son obligatorios.	103
5.2.	Códigos WMO para identificar tipos de observación.	103
5.3.	Formato de los registros de datos.	104
5.4.	Flags de control de calidad (QC flags).	104

5.5. Parámetros para la máscara de convección.	114
5.6. Parámetros para la máscara de granizo.	115
5.7. Variables obtenidas por <i>SatDatos</i>	116

Índice de listados

1.1.	Representación parcial en modo texto de un fichero netCDF. . . .	15
1.2.	Extracto de un fichero KML con información de una estación meteorológica.	18
3.1.	Extracto de un fichero con datos del SAIH Duero.	50
3.2.	Extracto de un fichero con valores de temperatura del SAIH Ebro.	51
3.3.	Extracto de un fichero con valores de precipitación del SAIH Ebro.	51
3.4.	Extracto de un fichero con datos meteorológicos del ITACyL	52
4.1.	Plantilla para <i>lftp</i>	78
4.2.	Fichero de configuración wf.xml para wfmanager.py.	85
4.3.	Fichero de configuración de <i>station</i>	88
4.4.	Extracto de un fichero CSV generado por <i>sql2csv</i>	91
4.5.	Función <i>putMarker</i> en el script <i>csv2kml.py</i>	93
5.1.	Datos de observación obtenidos de la plataforma GFA- \mathcal{M}^3	104
5.2.	Datos de observación en formato LITTLE_R.	105

Índice de Algoritmos

1.	Procesamiento de ficheros del SAIH Duero (load-chd-stations.py).	76
2.	Procesamiento de ficheros del SAIH Ebro (load-che-stations.py).	79
3.	Procesamiento de ficheros del ITACyL (load-itacyl-stations.py).	80
4.	Procesamiento de ficheros netCDF (nc2sql).	81
5.	ungrip.py (PyWRFScheduler).	86
6.	real.py (PyWRFScheduler).	87
7.	csv2nc.ncl (csv2nc).	93
8.	Cálculo de la máscara de convección.	116
9.	Cálculo de la máscara de convección.	117
10.	Cálculo del ángulo zenital.	119
11.	Corrección del efecto parallax.	120
12.	Cálculo de la altitud a partir de las variables PH y PHB.	121
13.	Cálculo de la altura de la nube.	121

Capítulo 1

Introducción

El tiempo atmosférico ha sido objeto de interés de la humanidad a lo largo de la historia. Dada la gran influencia que tiene cualquier cambio en la atmósfera en muchas actividades vitales como pueden ser la agricultura, el transporte de personas o mercancías, etc.; además de los efectos devastadores que pueden provocar fenómenos extremos como huracanes, lluvias torrenciales, nevadas, etc.

Recopilar observaciones del tiempo, tratar de explicar el comportamiento de la atmósfera y predecir el viento y la lluvia, son prácticas muy antiguas. Hasta entrado el siglo XX, las personas interesadas en el tiempo atmosférico realizaban tres tareas principales: recopilar datos a partir observaciones, e intentar sacar alguna conclusión a partir de ellos; tratar de explicar los fenómenos atmosféricos basándose en leyes generales; e intentar predecir el tiempo [Lezaun (2002)]. Estas tres actividades están relacionadas entre sí y el término *meteorología* se aplica a todas ellas:

Formalmente, la meteorología se define como:

Def. 1 *Ciencia que trata de la atmósfera y de los fenómenos que en ella ocurren.*

En el siglo XIX, fruto del aumento del número de personas dedicadas a la meteorología, las actividades empíricas de recopilación e interpretación de datos, las actividades teóricas para tratar de explicar los fenómenos atmosféricos y las actividades de predicción, se fueron especializando.

Aquellos que se dedicaron a actividades teóricas dieron origen a la meteorología dinámica, e hicieron de la leyes fundamentales de la física su punto de partida.

Entre aquellos que se centraron en actividades empíricas, surgió un interés por el tiempo promedio, punto de partida de una ciencia descriptiva, la *climatología*,

que se define como:

Def. 2 *Ciencia que estudia el clima, sus variedades y sus cambios y las causas de estos.*

donde *clima* se define como:

Def. 3 *Conjunto de condiciones atmosféricas que caracterizan una región.*

Las tres áreas de la meteorología siguieron su propio camino hasta mediados del siglo XX, donde la relación entre ellas quedó cada vez más patente. El desarrollo de los sistemas de observación permitió que los meteorólogos tuvieran acceso a gran cantidad de información para sus investigaciones. Esto favoreció el avance en las tres áreas de la meteorología, que finalmente terminaron convergiendo entre los años 50 y 60. Este cambio fue motivado, entre otras causas, por el desarrollo de la computación y el auge de la predicción numérica.

Hoy día existe una gran variedad de fuentes de datos procedentes de la observación, entre las que se incluyen: estaciones meteorológicas en superficie, radiosondeos, radares meteorológicos, satélites, boyas marítimas y mediciones tomadas desde barcos o aviones. Muchas empresas, como pueden ser aseguradoras o empresas del sector de las energías renovables; organismos de prevención de riesgos, como Protección Civil; servicios meteorológicos, como la Agencia Estatal de Meteorología (AEMET); y, dentro del ámbito académico, numerosos grupos de investigación, como el Grupo de Física de la Atmósfera (GFA) de la Universidad de León; recopilan información meteorológica que cada organismo utiliza según sus propios intereses. Toda esta información constituye una densa base de datos distribuida, que se actualiza de manera continua y que es transmitida y/o almacenada en distintos formatos.

Esta tesis plantea la posibilidad de unificar toda esta información, distribuida entre diferentes organizaciones, en un Sistema de Gestión de Datos Meteorológicos (SGDM) cuya definición formal es la siguiente:

Postulado 1 *Un Sistema de Gestión de Datos Meteorológicos (SGDM) es un conjunto de aplicaciones que permiten la recopilación y gestión informatizada de datos meteorológicos.*

Un SGDM se considera *abierto* (*open data*, en inglés) si la información que contiene se difunde a todo aquel que lo requiera sin restricción alguna.

Para diseñar un SGDM, hay que abordar aspectos relacionados con el acceso y disponibilidad de la información, los orígenes y formatos de información me-

teorológica o los modelos de gestión de información existentes. En el siguiente apartado se tratan todos estos conceptos.

1.1. Diseño de un SGDM

Este apartado presenta una serie de aspectos relacionados con el diseño de un SGDM. como son el acceso y disponibilidad de la información, los orígenes y formatos de información meteorológica y los modelos de gestión de información existentes.

1.1.1. Acceso y disponibilidad de la información

Open Definition¹, un proyecto de Open Knowledge International, precisa el significado de *open*:

Def. 4 “*Open means anyone can freely access, use, modify, and share for any purpose (subject, at most, to requirements that preserve provenance and openness)*”.

“*Abierto significa que cualquiera puede acceder, usar, modificar y compartir libremente con cualquier propósito (sujeto, como mucho, a restricciones para preservar la autoría y la apertura)*”.

El concepto *Open Data* surge de la idea de que determinados tipos de datos deberían estar a disposición de cualquiera que desee utilizarlos de cualquier forma, sin restricciones de copyright, patentes u otros mecanismos de control [Auer *et al.* (2007)].

Los objetivos del movimiento Open Data son similares a los que persiguen otros movimientos *abiertos*, como son Open Source, Open Hardware, Open Content u Open Access. La filosofía detrás del movimiento Open Data existe desde hace tiempo, sin embargo, el termino en sí mismo es reciente, ganando popularidad con el auge de Internet e iniciativas como Data.gov o Data.gov.uk.

La filosofía Open Data puede aplicarse a información de cualquier tipo: científica, médica, datos de administraciones públicas privadas, etc. El problema surge cuando estos datos tienen valor comercial. El acceso o la reutilización de los datos, controlado por diversas organizaciones, tanto públicas como privadas, se realiza mediante restricciones como licencias, copyright, patentes y cargos por uso o reutilización. El concepto Open Data mantiene que estas restricciones van en

¹<http://opendefinition.org/>

contra del bien común y que estos datos deberían estar disponibles sin ninguna restricción o coste.

En consecuencia, una parte importante de la comunidad científica aboga por el acceso libre a la información, argumentando que ello favorecería un avance científico mayor y más rápido [Molloy (2011)]. Organizaciones como Open Knowledge International² tratan de favorecer y dar a conocer esta forma de pensar y trabajar.

1.1.2. Obtención de datos meteorológicos

La información meteorológica se obtiene de muchos y variados medios: estaciones meteorológicas que proporcionan el valor de ciertas variables meteorológicas como la temperatura, la humedad relativa o la precipitación acumulada en un punto concreto; satélites meteorológicos que ofrecen una visión a gran escala espacial en tiempo real del estado de la atmósfera; modelos numéricos de predicción que permiten predecir las condiciones meteorológicas en un horizonte temporal prefijado; etc.

A pesar de esta variabilidad, es posible clasificar la información meteorológica en las dos categorías principales que se enumeran a continuación.

1. *Datos de observación.* Aquí se engloban todos los datos cuyo origen es algún sistema de observación [Chen *et al.* (2002); Daley (1993); J.Tapiador *et al.* (2012); Levizzani *et al.* (2007); Marshall *et al.* (2013)]. Se incluyen en esta categoría, entre otros, los datos obtenidos por satélites meteorológicos, radares meteorológicos, estaciones meteorológicas, radiosondeos, boyas marítimas, datos procedentes de medidas realizadas por aviones o barcos, etc.

Dependiendo de la precisión de la instrumentación utilizada, se puede afirmar que la información obtenida por estos sistemas son datos empíricos. Antes de ser puestos a disposición de los usuarios, estos datos pasan por sistemas de filtrado y depuración, que permiten detectar valores correspondientes a medidas anómalas.

Esta información tiene múltiples aplicaciones prácticas como se explica más adelante, sin embargo, por si solos, no aportan ningún conocimiento sobre las condiciones meteorológicas futuras.

2. *Datos de predicción.* Aquí están los datos obtenidos por modelos numéricos de predicción [Coiffier (2011); Warner (2010)]. Un modelo numérico de predicción permite estimar las condiciones meteorológicas futuras, lo cual tiene también múltiples aplicaciones prácticas.

²<https://okfn.org/>

Las observaciones son esenciales para establecer las condiciones iniciales de los modelos de predicción. Además, el uso de metodologías de asimilación de datos en los sistemas de predicción establece una relación fundamental entre la modelización numérica y las observaciones. La validación de las predicciones mediante datos de observación permite evaluar la calidad del modelo numérico de predicción.

En los siguientes apartados se profundiza en los diferentes orígenes de datos meteorológicos tanto de observación como de información pronosticada.

Datos de observación

En meteorología, se denomina datos de observación a la información recogida por sistemas de observación, como por ejemplo satélites, radares o estaciones meteorológicas. Esta información es de gran importancia y tiene muchos usos posibles, como la validación de algoritmos o parametrizaciones, la asimilación en modelos numéricos de predicción, etc. Cuanto mayor sea la fiabilidad y la precisión de la instrumentación que se utiliza, mayor será la veracidad de los datos. A los datos de observación también se los conoce como datos de verdad terreno (*ground truth*).

En las siguientes secciones se explican algunos sistemas de observación conocidos: satélite meteorológico, radiómetro de microondas, estación meteorológica, radar meteorológico, granizómetro y detector de rayos.

Satélite meteorológico Un satélite meteorológico es un tipo de satélite utilizado específicamente para monitorizar la atmósfera terrestre [Schmetz y Menzel (2015)]. De forma agrupada, los satélites meteorológicos de China, Europa, Estados Unidos, India, Japón y Rusia proporcionan observación casi continua del estado global de la atmósfera.

Tipos de satélite en función de su órbita En función de su órbita, hay dos tipos de satélites meteorológicos:

1. *Geoestacionarios*. Orbitan alrededor de la Tierra sobre el ecuador a 35.786 km de la superficie terrestre. Gracias a su órbita permanecen estáticos respecto al movimiento de rotación de la Tierra. Entre los más conocidos están los satélites de la serie Geostationary Operational Environmental Satellite (GOES) de Estados Unidos, los satélites de la serie Meteosat europeos y los de la serie Fengyun de China.
2. *Polares*. Orbitan alrededor de la Tierra perpendicularmente al ecuador a una altura de entre 800 y 1200 km de la superficie terrestre. Ofrecen una

resolución espacial mucho mayor debido a su cercanía a la Tierra, si bien, y a diferencia de los anteriores, su resolución temporal es menor.

Observación activa/pasiva En función del método de observación que utilicen también se puede clasificar los satélites meteorológicos en dos tipos [Chuvieco (2016)]:

1. *Observación activa*. Emiten una señal y procesan el eco recibido.
2. *Observación pasiva*. Reciben la radiación propia de los objetos en diferentes canales del espectro electromagnético, en particular, el visible y el infrarrojo. Entre estos canales están los siguientes [Schmetz *et al.* (2002)].
 - Visibles: $0,6 \mu m - 1,6 \mu m$ (útiles para registrar la cobertura nubosa durante el día).
 - Infrarrojos: $3,9 \mu m - 7,3 \mu m$ (vapor de agua); $8,7 \mu m - 13,4 \mu m$ (imágenes térmicas).

Radiómetro de microondas El radiómetro de microondas (Multichannel MicroWave Radiometer (MMWR)) es un sistema de teledetección pasiva que permite obtener perfiles verticales de temperatura a diferentes niveles de altitud [Gascón (2015)]. El MMWR mide la intensidad, en forma de temperaturas de brillo, emitida por algunos componentes de la atmósfera como pueden ser el oxígeno, el vapor de agua o el agua líquida, que después se convierten en otras variables más representativas, como el Contenido de Vapor de Agua (Water Vapor Content (WVC)), el Contenido de Agua Líquida (Liquid Water Content (LWC)), la Cantidad integrada de Vapor de Agua (Integrated Water Vapor (IWV)) y la Cantidad integrada de Agua Líquida (Liquid Water Path (LWP)).

Estación meteorológica Se trata de un dispositivo diseñado para monitorizar diversas variables meteorológicas en superficie, midiendo y registrando valores a lo largo del tiempo. Se conoce como red de estaciones meteorológicas a un conjunto de estaciones meteorológicas que pertenece a la misma organización. AEMET posee la red más extensa del territorio español.

Una estación meteorológica dispone de una serie de sensores que acotan el número de variables meteorológicas que es capaz de medir. Entre estos sensores podemos encontrar los siguientes:

- *Termómetro*. Mide valores de temperatura.
- *Termómetros de subsuelo (geo-termómetro)*. Mide la temperatura a diferentes niveles de profundidad.

- *Termómetro de mínima junto al suelo.* Mide la temperatura mínima a una distancia de 15 cm sobre el suelo.
- *Termógrafo.* Registra las fluctuaciones de la temperatura.
- *Barómetro.* Mide la presión atmosférica en la superficie.
- *Pluviómetro.* Mide la cantidad de agua caída sobre el suelo por metro cuadrado en forma de lluvia, nieve o granizo.
- *Psicrómetro o higrómetro.* Mide la humedad relativa del aire y la temperatura del punto de rocío.
- *Piranómetro.* Mide la radiación solar global (directa + difusa).
- *Heliógrafo.* Mide las horas de luz solar.
- *Anemómetro.* Mide la velocidad del viento.
- *Veleta.* Indica la dirección del viento.
- *Ceilómetro.* Mide la altura de las nubes.

Es posible clasificar las estaciones meteorológicas en base al número y al tipo de sensores, y a otros factores. Entre otras, podemos encontrar la siguientes categorías:

- *Observatorio sinóptico de superficie.* Normalmente, las estaciones meteorológicas no requieren ninguna operación, más allá de un mantenimiento ocasional. Sin embargo, existen observatorios meteorológicos que cuentan con personal cualificado, que, además de registrar los datos de un conjunto de estaciones meteorológicas, recopilan datos adicionales como la cantidad, el espesor y la densidad de nieve, etc. La información recogida en estos observatorios es especialmente útil para el establecimiento de las condiciones iniciales de los modelos numéricos de predicción que se utilizan para obtener pronósticos meteorológicos con cierto alcance temporal.
- *Observatorio aeronáutico.* Normalmente situado en aeródromos, registra variables meteorológicas útiles para la navegación aérea.
- *Estación termopluiométrica.* Registra temperaturas máximas, mínimas, etc. y precipitación.
- *Estación pluviométrica.* Registra solamente precipitación.
- *Estación evaporimétrica.* Registra la evaporación efectiva.

En relación a la comunicación de los datos recopilados, se denominan *estaciones automáticas* a aquellas estaciones meteorológicas que disponen de un canal de comunicación capaz de enviar datos en tiempo real a quien corresponda.

Radar meteorológico Un radar meteorológico, es un tipo de radar utilizado en meteorología para localizar zonas donde se produce precipitación de cualquier tipo, en forma de lluvia, nieve o granizo, y calcular su trayectoria. Se trata de un sensor activo que emite una onda electromagnética para después recibir la energía reflejada o dispersada por el objeto que la recibe [Fukao *et al.* (2014); Huuskonen *et al.* (2014)]. Estos objetos se conocen como *blancos*, y a la radiación dispersada como *eco*. La radiación electromagnética utilizada está en el rango de las microondas, generalmente entre 300 MHz y 300 GHz, que supone un período de oscilación de 3 ns a 3 ps y una longitud de onda de 1 mm a 1 m.

El radar está compuesto por tres elementos fundamentales que se muestran en la figura 1.1 y que se detallan a continuación:

1. *Transmisor*. Genera pulsos cortos de energía en la zona de radiofrecuencias del espectro electromagnético. En concreto, el *magnetron* genera microondas, el *modulador* las mantiene cierto tiempo, y el *temporizador* las emite en forma de pulso a una frecuencia de repetición constante.
2. *Antena*. Las microondas pasan a través de un guía de ondas desde el transmisor hasta la antena, concentrando los impulsos en un haz estrecho. Toda la energía se concentra en el *lóbulo principal* que gracias a un reflector parabólico lo enfoca hacia el espacio. El haz se propaga a la velocidad de la luz hasta que es interceptado por un objeto con propiedades diferentes a las del aire, que dispersa parte de la energía. Una parte de esa energía dispersada es devuelta hacia la antena.
3. *Receptor*. Un interruptor automático permite que la radiación pase del transmisor a la antena cuando se emite, y de la antena al receptor cuando se recibe la radiación. Debido a que la señal recibida es muy pequeña, es necesario amplificarla antes de poder visualizarla.

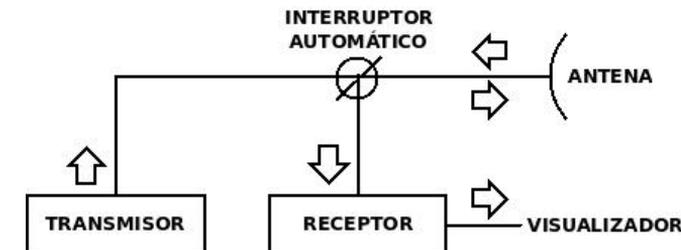


Figura 1.1: Elementos fundamentales de un radar [Gil (2009)].

Granizómetro Un granizómetro (*hailpad*) es un tipo de sensor que permite medir parámetros relacionados con la precipitación en forma de granizo, como pueden ser el tamaño, la energía, la velocidad de caída, etc [Long *et al.* (1980)]. Fue desarrollado por Chleusener y Jennings (1960). Originalmente consistía en una placa de Styrofoam® cubierta con una lámina de aluminio para evitar su deterioro debido a la exposición a la intemperie. Actualmente también se utilizan otros materiales sensibles. Las placas de material sensible se colocan sobre un soporte horizontal anclado al suelo. Las piedras de granizo impactan sobre la placa dejando registradas determinadas zonas en la superficie (figura 1.2). Analizando estas huellas [Dessens *et al.* (2015)], se puede obtener la siguiente información: Presencia o ausencia de granizo, tamaño máximo de las piedras, número de impactos por metro cuadrado, masa total de granizo precipitada (g/m^2), energía cinética total (J/m^2), etc.

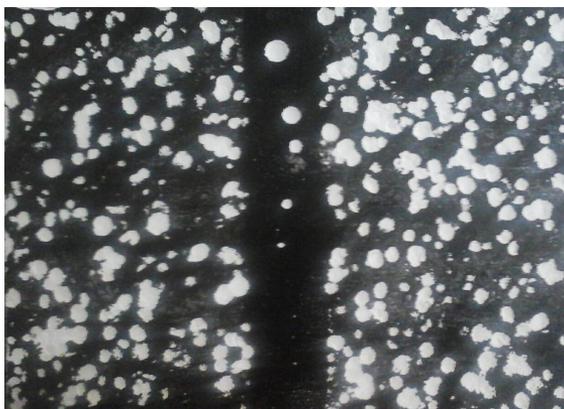


Figura 1.2: Granizómetro impactado.

Para seleccionar el momento adecuado para recoger y reemplazar los granizómetros impactados se utiliza un radar meteorológico, que permite identificar dónde y cuando se ha producido una tormenta con presencia de granizo.

Detector de rayos Un detector de rayos es un aparato que detecta los relámpagos o rayos de luz producidos por tormentas. En concreto, detectan el pulso electromagnético emitido por un relámpago; y midiendo la fuerza del mismo, el aparato puede estimar la distancia al rayo. Algunos detectores, al exponerlos a múltiples rayos detectados, pueden calcular y extrapolar la dirección del movimiento de las nubes de tormenta en relación a su posición.

Datos de predicción

La predicción meteorológica consiste en la determinación anticipada del valor, en un momento dado, de variables atmosféricas que afectan a una determinada región, como pueden ser temperatura, presión, humedad, nubosidad, precipitación, etc. La forma más habitual de realizar la predicción es haciendo uso de modelos matemáticos que emplean ecuaciones de las leyes físicas que describen el comportamiento de la atmósfera. La resolución de las mismas es complicada puesto que se trata de ecuaciones para las que no siempre existe una solución numérica exacta que permita determinar los valores futuros de las variables (dependencia temporal). Por esta razón, es necesario recurrir a técnicas de modelización numérica.

Modelos de predicción numérica Los modelos numéricos de predicción numérica (Numerical Weather Prediction (NWP)) aplican modelos físico-matemáticos de la atmósfera para predecir la circulación atmosférica. Existe un gran número de modelos desarrollados por diversas agencias y organismos meteorológicos. Estos modelos implementan diversas metodologías para pronosticar la evolución de ciertas variables meteorológicas partiendo de unas condiciones iniciales. Existen modelos de escala global, o sinóptica, diseñados para predecir la circulación atmosférica a gran escala. Entre estos, destacan el modelo Global Forecasting System (GFS) [Kalnay *et al.* (1990); Kanamitsu (1989); Kanamitsu *et al.* (1991)] de la National Oceanic and Atmospheric Administration (NOAA), y el modelo del European Center for Medium range Weather Forecasting (ECMWF) [Woods (2005)]. También hay modelos que permiten predecir el comportamiento de la atmósfera a una escala menor, mesoescala o escala regional, denominados Limited Area Models (LAM). Entre estos destacan el modelo Fifth-Generation NCAR/Penn State Mesoscale Model (MM5) [Dudhia (1993); Grell *et al.* (1995)]; su sucesor, el modelo Weather Research and Forecasting (WRF) [Skamarock *et al.* (2008, 2005)]; y el modelo High Resolution Limited Area Model (HIRLAM) [Christensen *et al.* (2007)].

Los modelos numéricos de predicción del tiempo se basan en la resolución numérica, mediante aproximaciones, de un conjunto de ecuaciones diferenciales referidas a la conservación de la masa, la energía y el momento en la atmósfera. Cada modelo hace uso de diferentes técnicas que le confieren ciertas características. A partir de unas condiciones iniciales y de contorno³, establecidas a partir de la asimilación de los datos obtenidos en un determinado instante de tiempo (t), el modelo obtiene los valores de las variables pronosticadas en un tiempo posterior ($t + \Delta t$). Para ello es preciso hacer uso de las ecuaciones que traducen las leyes

³Normalmente fijadas a partir de la salida de un modelo de mayor cobertura espacial, como GFS o como el modelo del ECMWF.

generales de la física que rigen la atmósfera terrestre.

La atmósfera es un fluido en movimiento en el que se producen procesos de intercambio de energía. En consecuencia, es necesario describir adecuadamente el movimiento de las partículas de aire sometidas a un conjunto de fuerzas, e introducir el balance energético una vez que se tienen en cuenta los intercambios que tienen lugar tanto en el medio atmosférico, como entre la atmósfera y la superficie terrestre. Para describir su comportamiento se utilizan ecuaciones generales de la Mecánica de Fluidos y de la Termodinámica. Además, el movimiento de las partículas de fluido atmosférico debe ser estudiado teniendo en cuenta que nuestro planeta es un sistema de referencia no inercial. Por lo tanto, para que se verifiquen las leyes de la mecánica clásica, es necesario introducir algunos términos denominados fuerzas ficticias.

Las ecuaciones matemáticas que describen estos procesos no son lineales y las soluciones no se pueden obtener de forma analítica. Para resolverlas es preciso recurrir al cálculo numérico, que proporciona una solución aproximada. Para ello, el primer paso consiste en discretizar el área geográfica objeto de análisis, definiendo una malla tridimensional (*grid*) que la contenga. A continuación se resuelven las ecuaciones, utilizando cálculo numérico, para obtener los valores de las variables en cada punto de la malla.

Características generales de los modelos mesoescalares Se entiende por modelo mesoescalar aquel que permite estudiar y/o pronosticar los fenómenos meteorológicos que incluyen factores mesoescalares.

Históricamente, los primeros modelos eran *hidrostáticos* porque los tamaños de las mallas horizontales que se empleaban eran comparables o mayores que la profundidad vertical de las características de interés. Esto ocurre con los modelos que simulan a escala sinóptica, válidos para establecer el comportamiento de las masas de aire, frentes polares, etc. En estos casos se mantiene la condición hidrostática y la presión está completamente determinada por las masas de aire extendidas, quedando inhibidos los movimientos del aire en la vertical en las condiciones iniciales de partida.

Sin embargo, cuando la escala de la malla del modelo se reduce, la resolución espacial aumenta y, en consecuencia, se puede incorporar la posibilidad de movimientos en la vertical. En este caso la hipótesis hidrostática deja de verificarse. Esto es importante pues permite incorporar algunos factores meteorológicos de tipo mesoescalar como la convección, ondas, etc.

Según lo expuesto, existen dos grandes categorías de modelos numéricos:

1. *Hidrostáticos*, como HIRLAM [Källén (1996); Männik y Room (2001)], utilizado por la AEMET.
2. *No hidrostáticos*, como Méso NH [Lafore *et al.* (1998)], MM5 o WRF.

Los modelos numéricos mesoescalares de área limitada permiten realizar simulaciones de la atmósfera bajo condiciones no hidrostáticas, y resultan muy útiles para el estudio de fenómenos con precipitaciones intensas y severas. Estos modelos hacen uso de distintos esquemas de parametrizaciones de cúmulos, microfísica, capa límite, etc., para dar cuenta de aquellos fenómenos cuya escala espacial es inferior a la resolución del modelo.

La incertidumbre asociada a las condiciones iniciales es un factor determinante en los resultados de la predicción [Bauer *et al.* (2015)]. Este efecto se conoce en Teoría del Caos como dependencia sensible a las condiciones iniciales [Lorenz (1993, 1963)], que establece que un pequeño cambio en las condiciones iniciales puede generar cambios radicales en el comportamiento de un sistema a largo plazo.

Para minimizar este impacto negativo se utiliza la predicción por conjuntos (*ensemble forecasting*) [Epstein (1969); Mullen y Baumhefner (1988); Stensrud *et al.* (2000)] como método de predicción numérico basado en la generación de una serie de muestras representativas del estado futuro de un sistema dinámico. Existen varias técnicas de predicción por conjuntos, como por ejemplo, la utilización de métodos Monte Carlo [Leith (1974)], donde se generan múltiples predicciones numéricas variando ligeramente las condiciones iniciales, siendo todas ellas plausibles y basadas en datos de observación.

Inicialmente la especificación de las condiciones de inicialización de un modelo NWP se basaban únicamente en el análisis e interpolación de los datos obtenidos mediante observatorios sinópticos. Estos métodos de interpolación han sido después reemplazados por técnicas de asimilación de datos [Kalnay (2003)].

La evolución del estado actual de la atmósfera (comúnmente denominado *análisis*) se trata como un problema de inversión Bayesiana utilizando datos de observación y datos de predicciones a corto plazo, junto con sus respectivas incertidumbres y restricciones, además del propio modelo numérico. Este cálculo se realiza en cuatro dimensiones para obtener un *análisis* físicamente consistente en espacio y tiempo, que pueda tratar con grandes cantidades de datos de observación distribuidos de manera heterogénea en espacio y tiempo, como es el caso de los datos de satélite. Dado que la incertidumbre del estado inicial es crucial para la predicción por conjuntos, y que la asimilación de datos utiliza observaciones y predicciones imperfectas, los métodos de predicción por conjuntos se han convertido en parte fundamental del proceso de asimilación de datos.

Asimilación de datos El objetivo de la asimilación de datos consiste en mejorar la exactitud del pronóstico combinando datos de observación con un pronóstico a corto plazo, comúnmente denominado campo de fondo o *guess*. Muchos sistemas de observación producen datos continuamente y el proceso de asimilación permite incorporar los datos recibidos entre un análisis y otro. La diferencia

entre la observación y el campo de fondo se denomina incremento de observación. La asimilación de datos analiza los incrementos de observación y crea una malla de correcciones para el pronóstico a corto plazo (o campo de fondo) anterior.

Un dato de importancia crítica para la asimilación de los datos es la covarianza del error de fondo en diferentes puntos. Esto significa que si el error del modelo en un lugar está relacionado con un error en otro lugar, la aplicación de una corrección en el primer lugar implica la necesidad de corregir también el segundo. Un concepto clave de la asimilación de datos es que el peso asignado al fondo y a las observaciones es inversamente proporcional a su varianza de error: la información con una varianza de error más pequeña (lo cual implica un error esperado menor) recibe más peso. La combinación de correcciones y pronósticos a corto plazo constituye las condiciones iniciales del ciclo de pronóstico siguiente.

Los procedimientos de asimilación de datos son bastante complejos, ya que existen diferencias de resolución espacial y temporal, y de exactitud, entre las observaciones. La calidad de las observaciones varía según la plataforma de observación, la hora del día, y las condiciones meteorológicas. Además, algunas observaciones no coinciden con las variables pronosticadas por el modelo, y en estos casos es preciso desarrollar técnicas que permitan relacionar los análisis del modelo con las observaciones.

Los incrementos de observación se ponderan según la percepción de exactitud y validez. El peso otorgado a los distintos elementos de información se basa en su covarianza de error. Para que el modelo funcione de manera razonable, es preciso aplicar las covarianzas de error necesarias y contar con buenas estimaciones de las mismas, lo cual complica la asimilación de datos. El error de observación puede ser grande si el instrumento no es muy exacto o si la resolución de la red de observación difiere de la resolución del modelo. Los datos aislados en regiones con pocos datos representan un problema de asimilación, ya que no hay nada que corrobore el incremento de observación.

La asimilación de datos en cuatro dimensiones, Four-Dimensional Data Assimilation (FDDA), permite introducir información nueva mediante reanálisis periódicos, inserción gradual o corrección, y técnicas matemáticas más avanzadas, como la asimilación de datos variacional. Existen distintos métodos variacionales de asimilación de datos:

- El método variacional tridimensional, 3DVAR, emplea covarianzas estáticas de error de fondo a fin de reducir al mínimo la distancia entre una solución del modelo y las observaciones disponibles.
- El método 4DVAR es una extensión de 3DVAR que permite un ciclo de asimilación de datos continuo durante una ventana temporal específica (que normalmente abarca varias horas).

La implementación operacional de estas técnicas de asimilación de datos en

cuatro dimensiones supone el mayor hito en lo que a predicción numérica se refiere.

1.1.3. Formatos de información comúnmente utilizados en meteorología

Como se ha visto en el apartado 1.1.2, existen fuentes de datos muy diversas y no existe ningún estándar que normalice la representación de datos meteorológicos; en consecuencia, cada una de estas fuentes utiliza el formato que más se adecúa a su finalidad o intereses.

En los siguientes apartados, se explican los formatos más utilizados para representar información meteorológica, en concreto, los que se enumeran a continuación: Network Common Data Form (netCDF), High Rate Information Transmission (HRIT), Meteorological Data Volume (MDV), representaciones en modo texto, representaciones basadas en eXtensible Markup Language (XML), representaciones basadas en JavaScript Object Notation (JSON) y formatos digitales de imagen.

netCDF

netCDF [Rew *et al.* (2015)] es un formato de archivo destinado a almacenar datos científicos (variables) en forma de vector (*array*). Un *array* es una estructura de n dimensiones que contiene elementos del mismo tipo (caracteres de 8 bit, enteros de 32 bits, etc.). Un valor escalar, es un *array* de cero dimensiones.

Entre las características que definen un fichero en formato netCDF están las siguientes:

- *Autodescrito*. Un fichero netCDF incluye información (metadatos) sobre los datos que contiene, como por ejemplo, en que unidades se miden.
- *Portable*. Se puede acceder a su contenido desde diferentes arquitecturas o plataformas. La representación física de los ficheros netCDF esta diseñada para ser independiente de la máquina en la que fueron escritos.
- *Escalable*. Se puede acceder a un subconjunto pequeño de los datos de manera eficiente.
- *Ampliable*. Se puede añadir información a un fichero netCDF ya existente sin necesidad de copiar los datos incluidos en el fichero.
- *Compartido*. Una operación de escritura y múltiples operaciones de lectura pueden realizarse al mismo tiempo.
- *Archivable*. Las versiones recientes de las librerías netCDF son compatibles con las anteriores.

netCDF fue desarrollado y es mantenido por Unidata, organización dentro de la University Corporation for Atmospheric Research (UCAR). Los autores principales son Glenn Davis, Russ Rew, Ed Hartnett, John Caron, Dennis Heimburger, Steve Emmerson, Harvey Davies y Ward Fisher; con contribuciones de muchos otros usuarios. Unidata mantiene las interfaces netCDF para C, FORTRAN 77, FORTRAN 90 y C++. El software netCDF de Unidata está disponible libremente para favorecer su uso extendido.

netCDF implementa un tipo abstracto de datos, lo que significa que todas las operaciones para consultar y manipular los datos en un fichero netCDF, deben utilizar el conjunto de funciones que permite la interfaz. La representación de los datos está oculta para las aplicaciones que utilizan la interfaz, lo que permite cambiar la forma en que se almacenan los datos sin afectar a estas aplicaciones.

Estructura de un fichero netCDF Los ficheros netCDF, habitualmente con extensión *nc*, son de tipo binario. Sin embargo, para poder analizarlos y editarlos, se pueden representar en modo texto. Es posible guardar esta representación en formato de texto en un fichero independiente, habitualmente con extensión *cdl*, generar un netCDF binario a partir de este y viceversa, utilizando *ncdump* y *ncgen*⁴.

Los datos en formato netCDF se almacenan en forma vectorial. Cada vector almacena datos de una variable concreta y tiene unas dimensiones y atributos acordes con los datos que contiene. La estructura básica del fichero netCDF, en su representación en modo texto, tiene 4 partes, como se puede ver en el listado 1.1:

- *Dimensiones*. Especificación de las diferentes dimensiones que tienen las variables contenidas en él, comúnmente: latitud, longitud, altitud sobre el nivel del mar y tiempo.
- *Variables*. Definición de las diferentes variables incluidas en el fichero, especificando dimensiones y atributos.
- *Atributos globales*. Lista de atributos globales aplicables a todas las variables del fichero.
- *Datos*. Valores de las variables definidas previamente.

```

1 netcdf wrfout_d01_2011-12-22_00:00:00 {
2   dimensions:
3     Time = UNLIMITED ; // Horizonte temporal
4     west_east = 122 ; // Longitud
5     south_north = 111 ; // Latitud
6     bottom_top = 34 ; // Altitud sobre el nivel del mar
7     // ...

```

⁴Ambos incluidos en las librerías netCDF.

```

8 variables:
9   float T(Time, bottom_top, south_north, west_east) ;
10   T:description = "perturbation potential temperature (theta-t0)" ;
11   T:units = "K" ;
12   float RAINC(Time, south_north, west_east) ;
13   RAINC:description = "ACCUMULATED TOTAL CUMULUS PRECIPITATION" ;
14   RAINC:units = "mm" ;
15   // ...
16   // global attributes:
17   :START_DATE = "2011-12-22_00:00:00" ;
18   ...
19 data:
20   T = -7.835083, -7.864655, -7.891296, -7.904358, -7.911102, -7.931854,
21       -7.968079, -8.019257, -8.089844, -8.18219, -8.287872, -8.430267,
22       ... ;
23   RAINC = 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
24           0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
25           ... ;
26   // ...
27 }

```

Listado 1.1: Representación parcial en modo texto de un fichero netCDF.

HRIT

HRIT y Low Rate Information Transmission (LRIT) son el estándar propuesto por la Coordination Group for Meteorological Satellites (CGMS) para la difusión de información digital obtenida de satélites geoestacionarios [Wolf (1999)]. La diferencia entre ambos formatos, como sugiere su nombre, es el ancho de banda necesario para transmitir contenidos. LRIT se difunde normalmente a velocidades alrededor de los 256 Kbps, mientras que HRIT se difunde normalmente a velocidades por encima de los 10 Mbps. Habitualmente, el contenido de un fichero en formato LRIT es un subconjunto de los mismos datos en formato HRIT, aplicando técnicas de muestreo y compresión.

MDV

El formato MDV [Dixon (2006)] fue desarrollado por el Research Applications Laboratory (RAL), National Center for Atmospheric Research (NCAR), en los años sesenta. Está diseñado para representar información en forma de malla (*grid*) de dos o tres dimensiones. No tiene ningún componente temporal, un conjunto de datos MDV contiene información para un único instante de tiempo. Permite almacenar varias variables, por ejemplo, un fichero MDV con información de radar puede contener datos de reflectividad y velocidad radial. Para las dimensiones horizontales (longitud y latitud), soporta diversos tipos de proyecciones geográficas. Para la dimensión vertical, permite distintos tipos de coordenadas, como por ejemplo altura (m) o en niveles de presión (hPa).

Representaciones basadas en texto

Muchas aplicaciones, dispositivos de observación, etc., del ámbito de la meteorología trabajan con representaciones personalizadas en modo texto. Un formato habitual en estos casos, pero no el único, es el formato Comma Separated Values (CSV).

CSV El formato CSV [Shafranovich (2005)] se utiliza, desde hace mucho tiempo, para intercambiar y convertir datos entre diversos programas. Sorprendentemente, a pesar de ser un formato muy común, nunca ha sido documentado formalmente.

Las siguientes reglas son utilizadas por casi todas las especificaciones CSV:

- Cada registro ocupa una única línea y están delimitados por un salto de línea (CRLF). El último registro no tiene porque terminar en un salto de línea
- Cada registro se compone de uno o varios campos separados por comas. Después del último registro no debe haber una coma.
- Todos los registros deben tener el mismo número de campos.
- Opcionalmente un campo puede estar entrecomillado para marcar el inicio y fin del mismo, y poder así introducir algún carácter especial en el campo.
- Opcionalmente, puede haber un encabezado que contenga los nombres de los campos de cada registro. El encabezado tiene que tener tantos campos como el resto de registros.

Representaciones basadas en XML

Existen varios formatos de representación de datos geográficos basados en XML. Los más populares son Keyhole Markup Language (KML) y Geography Markup Language (GML).

KML KML es un lenguaje de marcado basado en XML para representar datos geográficos en tres dimensiones [OGC (2008)]. Los ficheros KML se suelen distribuir comprimidos (con extensión KMZ) y de esta manera pueden incluir archivos de imágenes y otros recursos asociados.

Un fichero KML contiene información asociada a ubicaciones geográficas: un nombre, una descripción, sus coordenadas (latitud y longitud) y alguna otra información relevante. Como ejemplo, en el listado 1.2 se muestra un extracto de fichero KML con información de una estación meteorológica.

```

1 <kml xmlns="http://earth.google.com/kml/2.2">
2 <Document>
3   <Placemark>
4     <name>AF605</name>
5     <description>
6       <![CDATA[
7         Origen: Confederación Hidrográfica del Duero
8         Identificador de estación: AF605
9         Fecha/Hora: 2013-11-20 12:00:00
10        Longitud: -6.881656
11        Latitud: 40.970301
12        Altura: 140.00 m
13        Temperatura media: 1.06 C
14      ]]>
15    </description>
16    <Point>
17      <coordinates>-6.881656,40.970301,0</coordinates>
18    </Point>
19  </Placemark>
20  ...
21 </Document>
22 </kml>

```

Listado 1.2: Extracto de un fichero KML con información de una estación meteorológica.

Las especificaciones KML 2.2 se presentaron ante el Open Geospatial Consortium (OGC) con el fin de alcanzar un estatus de estándar abierto como formato de intercambio de información geográfica. En noviembre de 2007 el OGC creó un nuevo grupo de trabajo sobre KML 2.2 y se solicitan alegaciones hasta el 4 de enero de 2008. El 14 de abril de 2008 se convirtió definitivamente en nuevo estándar de la industria.

GML Descrito como gramática en XML Schema, GML también se utiliza para la representación de información geográfica. Se diseñó a partir de la especificación abstracta producida por el OGC y los documentos de la serie ISO 19100. Es el formato utilizado para el intercambio de información entre los organismos que gestionan el registro de la propiedad en España.

Representaciones basadas en JSON

El formato JSON también se ha utilizado para la representación de información geográfica.

GeoJSON GeoJSON es un formato estándar abierto diseñado para representar elementos geográficos simples basado en JSON. El formato es ampliamente utilizado en aplicaciones de cartografía en entornos web al permitir el intercambio de datos de manera rápida, ligera y sencilla.

La gramática GeoJSON está basada en el estándar Well Known Text (WKT) del OGC. Incluye una serie de geometrías que pueden ser de tipo punto para direcciones, ubicaciones, puntos de interés, etc.; líneas para calles, carreteras, fronteras, etc.; y polígonos para países, provincias, parcelas catastrales, etc. Además, todas ellas se pueden agrupar en colecciones.

GeoJSON es mantenido por una comunidad de desarrolladores en Internet.

Formatos digitales de imagen

Las representaciones gráficas son muy habituales en meteorología ya que permiten visualizar variables continuas. Las imágenes digitales se pueden guardar en diversos formatos, siendo los más habituales los que se enumeran a continuación.

BMP Las imágenes en formato Bitmap (BMP) se forman mediante una parrilla de píxeles. El formato BMP no sufre pérdidas de calidad y por tanto resulta adecuado para guardar imágenes que se desean manipular posteriormente. Guarda gran cantidad de información de la imagen. El mayor inconveniente es que los archivos tienen un tamaño muy grande.

GIF Graphics Interchange Format (GIF) fue diseñado para comprimir imágenes digitales. Reduce la paleta de colores a 256 colores como máximo (profundidad de color de 8 bits). No es recomendable para fotografías de cierta calidad.

JPEG Joint Photographic Experts Group (JPEG) también fue diseñado para comprimir imágenes digitales, pero a diferencia de GIF admite una paleta de hasta 16 millones de colores. La compresión JPEG puede suponer cierta pérdida de calidad en la imagen. En la mayoría de los casos esta pérdida es asumible ya que permite reducir el tamaño del archivo considerablemente y su visualización es aceptable.

TIFF Tagged Image File Format (TIFF) utiliza cualquier profundidad de color de 1 a 32 bits, lo que permite una calidad de imagen excelente. Sin embargo, como BMP produce archivos muy grandes.

PNG Portable Network Graphics (PNG) es un formato gráfico basado en un algoritmo de compresión sin pérdida para bitmaps no sujeto a patentes. Este formato fue desarrollado en buena parte para solventar las deficiencias del formato GIF y permite almacenar imágenes con una mayor profundidad de contraste.

1.1.4. Modelos de gestión de información

Existen varias alternativas a la hora de almacenar y gestionar un volumen grande de información. La elección de una u otra se hace en función de las necesidades concretas de cada caso. En los siguientes apartados se exponen algunos modelos de gestión de información ampliamente desarrollados y que guardan relación con el objetivo principal de este trabajo, establecer las bases y desarrollar un SGDM.

Sistemas de información geográfica

Un Geographic Information System (GIS) es un conjunto de herramientas que permiten la organización, almacenamiento, manipulación, análisis y modelización de grandes cantidades de datos vinculados a una referencia espacial, facilitando la incorporación de aspectos socio-culturales, económicos y ambientales que conducen a la toma de decisiones de una manera más eficaz. Se considera GIS cualquier sistema de información capaz de integrar, almacenar, editar, analizar, compartir y mostrar la información geográficamente referenciada.

La tecnología GIS puede ser utilizada para investigaciones científicas, gestión de los recursos, gestión de activos, evaluación del impacto ambiental, planificación urbana, cartografía, sociología, marketing, logística, etc.

El primer paso hacia la creación de información geo-referenciada implica el establecimiento de un modelo que permita interpretar la realidad geográfica. Se trata de representar el espacio estudiado, la variable tratada y la variación de esta a lo largo del espacio. Este modelo geográfico constituye una forma de entender un hecho geo-referenciado, pero que todavía no incorpora elementos relativos a su representación o almacenamiento.

Existen muchos modelos geográficos distintos, entre los cuales cabe destacar dos de ellos [Coclelis (1992)]: modelo geográfico de *campos* y modelo geográfico de *entidades discretas*.

Modelo geográfico de campos Un campo es un modelo de variación dentro de un marco n -dimensional, en el cual en cada punto dentro de dicho marco se tiene un valor de la variable estudiada. En su concepto matemático, un campo es una función de la forma $\varphi : R^n \rightarrow R^m$, esto es, una función que asocia cada punto de un espacio vectorial con otro en un espacio vectorial distinto.

En el caso más habitual, $m = 1$, es decir, que a cada punto del espacio vectorial origen se le asocia un único valor escalar. Se tiene así lo que se denomina un campo escalar. La mayoría de las variables que se emplean en un GIS necesitan un único valor para describirse (piénsese en variables como la elevación, la temperatura o la presión atmosférica, que solo requieren de un número para expresarse), por lo que los campos escalares son los más habituales en el ámbito geográfico.

No obstante, también encontramos los denominados campos vectoriales, en el cual el espacio vectorial de destino es multidimensional. Por ejemplo, para definir el viento en un punto geográfico no basta con un único valor, sino dos: la velocidad y la dirección en la que sopla dicho viento. Dentro de un GIS, es habitual recoger los campos vectoriales como un conjunto de varios campos escalares, cada uno de ellos en una capa distinta. Así, se tendría una capa con la dirección y otra con la velocidad, ambas magnitudes escalares. Operando de esta manera, la solución no es única, ya que el vector resultante puede definirse mediante su módulo y dirección (como en el caso anterior), pero también por sus propias coordenadas en la base del espacio vectorial destino (en el caso anterior, las componentes x , y y z del vector que indican el movimiento del viento).

El espacio vectorial de origen puede ser bidimensional, es decir, una función de la forma $f(x, y)$, representando x e y las coordenadas geográficas. Este es el caso habitual en las capas que se emplean en un GIS, donde las variables que estudiamos adquieren uno u otro valor en función de su posición dentro de un sistema coordenado de referencia.

Puede añadirse una tercera dimensión, de tal modo que los valores dependan no solo de la posición sino igualmente de la elevación. Se tendría una función de la forma $f(x, y, z)$. Para el caso, por ejemplo, de la temperatura del aire, esta depende no solo de la localización, sino también de la altura. Aunque en general es poco habitual en el marco de los GIS, puede añadirse la variable tiempo (t), teniéndose funciones de la forma $f(x, y, t)$ o $f(x, y, z, t)$

Por definición, un campo es continuo, cuando todos los puntos tienen un valor asociado. De igual modo, este valor es único, y no existe un elemento del espacio vectorial de partida que tenga asociados varios elementos del de destino, sean estos escalares o vectores. Por su propia naturaleza los campos son ideales para modelizar variables que varían de forma continua en el espacio, entre ellas la practica totalidad de variables físicas del medio, tales como temperatura del aire, presión atmosférica, elevación, etc. Los campos se asocian con las denominadas coberturas, término este más empleado en el ámbito GIS. En una cobertura existe un valor único para todos los puntos de una región dada.

Modelo geográfico de entidades discretas A diferencia de los campos, el modelo de entidades discretas no asocia a cada punto geográfico un valor, sino que concibe un entorno geográfico como un espacio vacío sobre el que se sitúan distintos elementos (entidades) que lo van rellenando. Cada una de dichas entidades posee unas características propias, constantes para toda ellas, que son las que conferirán sus propiedades particulares a los puntos que se sitúan en su interior.

Un punto puede no pertenecer a ninguna entidad, o bien a varias de ellas, según sea la disposición de estas. Para un espacio dado, las entidades pueden ser todos aquellos elementos geométricos existentes en el mismo, tales como puntos,

líneas, polígonos o, en el caso de ser dicho espacio de dimensión mayor que dos, también volúmenes.

Es fácil ver que el modelo de entidades discretas no es tan adecuado como los campos para conceptualizar variables continuas, ya que la continuidad de estas es opuesta al esquema discreto planteado. No obstante, otras variables no continuas se modelizan mejor mediante entidades discretas, ya que la forma en que se presentan coincide en cierta medida con ellas como unidades mínimas.

Uno de los ejemplos que puede asimilarse a este modelo es la presencia de vías de comunicación. Se tiene un espacio vacío (sin vías), en el cual se disponen los distintos viales en una serie de localizaciones concretas. Hay puntos que no estarán afectados por ninguna entidad, mientras que otros (los situados en las intersecciones) lo están por varias de ellas.

Las variables de tipo nominal y alfanumérico, que no son continuas, tales como el tipo de suelo en un punto o el número de parcela catastral al que pertenece dicho punto, también se adaptan bien al modelo de entidades discretas.

Otra diferencia entre los campos y las entidades discretas es que estas últimas son en general más sencillas de comprender como concepto fuera de un ámbito técnico. Los campos son conceptos matemáticos que requieren un mayor grado de abstracción, y para la mayoría de la gente no resultan tan claros. Como algunos apuntan [Goodchild y Kemp (1990)], el lenguaje habitual contiene un número mayor de expresiones y recursos para describir la realidad geográfica en base a entidades discretas que en base a campos o conceptos abstractos similares.

Sistemas de información climatológica e hidrológica

Un Sistema de Gestión de Datos Climáticos (SGDC) tiene por objetivo almacenar y explotar de manera eficiente un cierto volumen, normalmente grande, de información climatológica. Conceptualmente un SGDC y un SGDM son similares, pero su implementación es totalmente diferente. La World Meteorological Organization (WMO)⁵ ha definido una serie de directrices y buenas prácticas para la gestión de datos climáticos [Plummer *et al.* (2007)]. Si bien, estas directrices son un buen punto de partida, hay que tener en cuenta las diferencias entre un SGDM y un SGDC. Por lo general, un repositorio diseñado para actividades meteorológicas en tiempo casi real será consultado *sinópticamente*, es decir, obteniendo todos los datos referidos a ciertas ubicaciones o áreas correspondientes a un período de tiempo definido, relativamente corto. Las aplicaciones climáticas, en cambio, suelen alimentarse de datos referentes a una o varias localizaciones concretas, aunque comprendidos en períodos largos.

⁵<https://www.wmo.int>

1.2. Hipótesis, resultados esperados y métricas

Este apartado presenta una serie de hipótesis, los resultados esperados de su cumplimiento y las métricas para evaluar los mismos.

1.2.1. Hipótesis

Un SGDM que contenga datos de observación meteorológica en tiempo real de diferentes orígenes simplifica notablemente el seguimiento y monitorización de ciertos eventos meteorológicos, como pueden ser las tormentas severas, las nevadas, etc. Esto permite plantear la siguiente hipótesis:

Hipótesis 1 *Un SGDM facilita el seguimiento de eventos meteorológicos en tiempo real.*

Además de permitir el seguimiento de meteoros, tener disponibles un conjunto de datos de observación es imprescindible hoy día para mejorar la precisión de las predicciones obtenidas con un modelo NWP. Los modelos NWP mejoran su precisión gracias a los procesos de asimilación de datos [Bauer *et al.* (2015)]. Una predicción meteorológica certera salva vidas, supone un apoyo importante en la gestión de emergencias y permite reducir el impacto y prevenir las pérdidas económicas derivadas de un fenómeno meteorológico adverso.

Los datos de observación vía satélite son especialmente útiles a la hora de mejorar las condiciones iniciales de una predicción. El proyecto Global Precipitation Measurement (GPM) de la National Aeronautics and Space Administration (NASA) propone el uso de datos de observación vía satélite como datos de asimilación para mejorar la precisión de los pronósticos generados por modelos NWP [J.Tapiador *et al.* (2012)]. Sin embargo, el acceso a estos datos no es libre para cualquiera que lo requiera. Lo anteriormente expuesto permite establecer también la siguiente hipótesis:

Hipótesis 2 *Un SGDM permite obtener información para asimilación de datos en modelos NWP.*

Los investigadores necesitan datos que les permitan contrastar sus hipótesis. Por ejemplo, García-Ortega *et al.* (2012) hablan sobre la dificultad de obtener un pronóstico de precipitación convectiva certero con las predicciones obtenidas a partir de modelos NWP. Esto es así, porque la precipitación en forma de granizo no es una variable que se obtenga directamente a partir de las ecuaciones dinámicas que utilizan los modelos NWP. Para mejorar las predicciones, los autores estudian la importancia de los factores sinópticos en la formación de las tormentas severas con presencia de granizo.

Existen numerosos estudios acerca de la sensibilidad a las parametrizaciones físicas de los modelos NWP. Tapiador *et al.* (2012) realizan una comparativa de diferentes parametrizaciones del modelo WRF en eventos de precipitación. García-Ortega *et al.* (2017) llevan a cabo un análisis de sensibilidad de diferentes combinaciones de parametrizaciones físicas del modelo WRF para la mejora de la predicción del campo de precipitación convectiva en el noreste de España utilizando datos de los pluviómetros del Sistema Automático de Información Hidrológica (SAIH) de la Confederación Hidrográfica del Ebro. En ambos casos, fue necesario una gran cantidad de tiempo para procesar la información de los pluviómetros del SAIH y compararla con las salidas del modelo. Si los datos de los pluviómetros de la Confederación Hidrográfica del Ebro (CHE), hubieran estado previamente procesados y a disposición de los investigadores, este proceso de contraste habría sido mucho más rápido para ambos equipos de investigadores.

Un modelo que permita almacenar y gestionar de manera eficiente un volumen muy grande de datos meteorológicos y, además, ponerlos a disposición de la comunidad científica, sería una herramienta muy poderosa para apoyar la investigación en meteorología y climatología. El acceso a estos datos permite a los investigadores contrastar sus hipótesis y avanzar en actividades tales como la elección de parámetros para modelos numéricos de predicción meteorológica o la calibración y validación de algoritmos de detección de ciertos riesgos meteorológicos. Todo esto permite plantear la siguiente hipótesis:

Hipótesis 3 *La información contenida en un SGDM facilita la calibración y validación de algoritmos en el ámbito de la meteorología y la climatología.*

A otro nivel, pero por idénticos motivos, un SGDM puede suponer un importante apoyo a la educación académica, favoreciendo el aprendizaje autónomo y la evaluación de conocimientos sobre meteorología y climatología. Un SGDM puede permitir a un estudiante de meteorología comprobar la veracidad de sus propios pronósticos o elegir el mejor modelo numérico de predicción en una situación determinada. Además, sería de utilidad para la comunidad docente, ya que puede servir como método objetivo de evaluación de ciertas actividades, como la elaboración de pronósticos. Por tanto se que plantea la siguiente hipótesis:

Hipótesis 4 *Un SGDM favorece el aprendizaje en el ámbito de la meteorología y la climatología.*

Además de servir como apoyo a la investigación y a la docencia, la información meteorológica tiene usos comerciales si se cumplen los postulados planteados por las hipótesis 1 y 2. Algunos ejemplos pueden ser los siguientes: las empresas

aseguradoras pueden hacer uso de esta información para comprobar o valorar siniestros en agricultura; las empresas de producción de energía eólica pueden optimizar sus recursos arrancando sus aerogeneradores solamente cuando haya presencia de viento; se pueden desarrollar aplicaciones que pongan a disposición de los usuarios información meteorológica de interés, como la lluvia o la presencia de nieve en las carreteras.

La información meteorológica puede también ser muy útil en la prevención, o atenuación de las consecuencias, de ciertos riesgos meteorológicos como pueden ser las lluvias torrenciales, los huracanes o las fuertes nevadas. Adelantarse a este tipo de episodios, no solamente supone un importante ahorro de recursos, sino que incluso puede salvar vidas humanas.

1.2.2. Resultados esperados

Establecidas las hipótesis 1, 2, 3 y 4, el desarrollo de un SGDM debe aportar los siguientes resultados:

1. Desarrollo de un modelo de gestión de información meteorológica, en adelante Meteorological data Management Model (\mathcal{M}^3), que ofrezca las siguientes funcionalidades:
 - a) Repositorio de información meteorológica de diferente origen.
 - b) Mecanismo de consulta de la información contenida en el repositorio, accesible a cualquier persona física u organización autorizados.
 - c) Mecanismo de adquisición de nuevos datos meteorológicos, accesible a cualquier persona física u organización autorizados.
2. Taxonomía de los diferentes orígenes de información meteorológica, tanto de datos de observación, como de datos de predicción.
3. Taxonomía de los formatos en los que se representa la información meteorológica.
4. Implementación y validación del modelo \mathcal{M}^3 . Por cuestiones relacionadas con el acceso a la información, un SGDM normalmente contendrá datos de un área geográfica determinada. Sin embargo, todas las ideas y aplicaciones pueden ser extrapolables a cualquier otra ubicación geográfica.

Por estos motivos, el SGDM desarrollado, denominado GFA Meteorological data Management Model (GFA- \mathcal{M}^3), cubre las áreas geográficas de las cuencas de los ríos Duero y Ebro, y obtiene la información de las fuentes que se enumeran a continuación.

- Datos meteorológicos recopilados por el propio GFA.

- Datos meteorológicos recopilados por los organismos públicos y privados con los que colabora actualmente el GFA.
 - Datos meteorológicos recopilados por observadores voluntarios. Existe una gran cantidad de observadores aficionados con instrumentación propia cuyos datos pueden ser interesantes para cubrir ciertas áreas de estudio.
5. Desarrollo de herramientas de contraste de información meteorológica.
 6. Desarrollo de aplicaciones que hagan uso de la información contenida en el SGDM desarrollado.

1.2.3. Métricas

En este apartado se presentan una serie de métricas que permiten evaluar la validez de las hipótesis 1, 2, 3 y 4 establecidas previamente.

Para la validación de la hipótesis 1, se plantea la siguiente métrica:

Métrica 1 *Evaluación numérica y cualitativa de las herramientas para el seguimiento de eventos meteorológicos desarrolladas a partir de la información del GFA- \mathcal{M}^3 .*

En relación a la hipótesis 2, se plantea la siguiente métrica:

Métrica 2 *Evaluación de la plataforma GFA- \mathcal{M}^3 como fuente de datos de asimilación para modelos NWP.*

Por último, para la validación de la hipótesis 3 y 4 establecidas previamente se plantea la siguiente métrica:

Métrica 3 *Evaluación de las contribuciones del GFA.*

1.3. Estructura de la memoria

Además de la presente introducción, esta memoria se estructura en cuatro capítulos y unas conclusiones. Incluye también un resumen, una bibliografía y un glosario de acrónimos.

En primer lugar, el capítulo 2 se establecen las bases del modelo \mathcal{M}^3 en relación a otros modelos de tratamiento de información.

A continuación, en el capítulo 3 se realiza un análisis exhaustivo de las diferentes fuentes de información a las que tiene acceso el GFA. Fruto de este análisis

se obtiene una taxonomía detallada de los diferentes orígenes de información, que incluye información relativa al formato de los datos, el volumen, la periodicidad con que se generan, etc.

El capítulo 4 presenta la implementación de un SGDM, denominado GFA- \mathcal{M}^3 , siguiendo las directrices del modelo \mathcal{M}^3 planteadas en el capítulo 2.

En el capítulo 5 se evalúan algunas de las aplicaciones y resultados experimentales obtenidos a partir de la información contenida en el GFA- \mathcal{M}^3 . Con ello, se pretende dar validez a las hipótesis 1, 2, 3 y 4.

Por último, en el capítulo 6 se exponen las conclusiones alcanzadas tras el desarrollo del modelo GFA- \mathcal{M}^3 y su utilización en diferentes ámbitos; evaluando después la validez de las hipótesis 1, 2, 3 y 4 planteadas previamente en relación a las métricas 1, 2 y 3.

Capítulo 2

Modelo de gestión de datos meteorológicos (\mathcal{M}^3)

Este capítulo presenta la definición formal del modelo Meteorological data Management Model (\mathcal{M}^3) que permite implementar un SGDM. En primer lugar se exponen algunos sistemas de gestión de información alternativos. Después, se explica el modelo propuesto enumerando sus componentes, para posteriormente profundizar en cada uno de ellos. Por último, se expone el diseño de la plataforma GFA- \mathcal{M}^3 , siguiendo la especificación del modelo \mathcal{M}^3 .

2.1. Sistemas de gestión de datos existentes

En el capítulo 1 se presentaron dos familias de modelos de gestión de información: SGDC y GIS. En este apartado se abordan algunas implementaciones tanto de SGDC, como de GIS, para discutir después su idoneidad para ser utilizados como SGDM.

2.1.1. SGDC

Como se expone en el apartado 1.1.4, a diferencia de un SGDM, un SGDC está optimizado para la obtención del valor de determinadas variables meteorológicas, en una localización geográfica concreta, a lo largo de un periodo de tiempo prolongado, para de esta forma poder estudiar su evolución.

A continuación se enumeran dos plataformas de almacenamiento y explotación de información climatológica. En concreto: MeteoSIX y Meteorological, Climatological and Hydrological Database Management System (MCH).

MeteoSIX

El proyecto MeteoSIX [Regueiro *et al.* (2012)] nace en diciembre de 2009 y tiene por objetivo la difusión de datos meteorológicos y oceanográficos recogidos por MeteoGalicia¹, servicio meteorológico dependiente de la Junta de Galicia. El proyecto está coordinado por MeteoGalicia y cuenta con la participación del grupo de Arquitectura de Computadores de la Universidad de La Coruña (UDC), del Laboratorio de Sistemas de la Universidad de Santiago de Compostela (USC) y del Centro de Supercomputación de Galicia (CESGA)².

MeteoSIX dispone de una Application Programming Interface (API) que permite el acceso a los resultados de los modelos de predicción ejecutados diariamente por MeteoGalicia, además de información sobre las mareas y sobre las horas de salida y puesta del sol. El objetivo de esta API es facilitar la consulta de la información de predicción meteorológica y oceanográfica existente, presentándola de un modo estructurado y ordenado. La API puede ser útil tanto para aquellos que deseen desarrollar aplicaciones meteorológicas usando estos datos, como para usuarios finales que necesiten la información para distintos tipos de análisis. El acceso es público y gratuito. El único requisito es la solicitud de una clave de acceso a través de su portal web.

Las variables que se pueden consultar a través de la API son las siguientes: estado del cielo, temperatura, viento, precipitación, cota de nieve, humedad relativa, cobertura nubosa, presión al nivel del mar. Además, en puntos sobre el mar se pueden consultar las siguientes variables: período de ola, dirección de ola, temperatura del agua, salinidad y altura de ola.

MCH

MCH [WMO (2011)] es un Data Base Management System (DBMS) basado en aplicaciones de código abierto. Fue desarrollado en México y adaptado por el Programa Iberoamericano de Cooperación, destinado al fortalecimiento de diversos componentes de los Servicios Meteorológicos Nacionales de países iberoamericanos. En 2011, este sistema fue puesto a disposición de la WMO para su distribución gratuita a los servicios hidro-meteorológicos nacionales de todo el mundo y, casi cuatro años más tarde, este software libre es utilizado por numerosos países americanos, africanos y de Europa del Este.

MCH está dirigido principalmente a servicios meteorológicos nacionales, que no dispongan de ningún DBMS, para digitalizar, almacenar y procesar grandes cantidades de datos; o que busquen una solución sencilla, personalizable y de libre distribución, para almacenar, analizar y generar informes sobre una cantidad grande de datos meteorológicos, climatológicos e hidrológicos.

¹<http://www.meteogalicia.gal>

²<http://www.cesga.es/>

2.1.2. GIS

Entre las soluciones GIS más populares destacan: ArcGIS, QGIS, GRASS y GvGIS. Cualquiera de ellas maneja archivos vectoriales y raster con diferentes formatos. Los archivos vectoriales siguen el modelo geográfico de entidades discretas, explicado en el capítulo 1. Los archivos raster, siguen el modelo geográfico de campos, también abordado en el capítulo 1

ArcGIS, desarrollado y comercializado por ESRI, agrupa varias aplicaciones para la captura, edición, análisis, tratamiento, diseño, publicación e impresión de información geográfica. Estas aplicaciones se engloban en familias temáticas como ArcGIS Server, para la publicación y gestión web, o ArcGIS Móvil para la captura y gestión de información en campo.

QGIS es un GIS de código abierto para plataformas GNU/Linux, Unix, Mac OS, Microsoft Windows y Android. Está desarrollado en C++, usando la biblioteca Qt para su Interfaz gráfica de usuario. Además, permite la integración de plugins desarrollados tanto en C++ como Python. Una de sus mayores ventajas es la posibilidad de usar QGIS como GUI de GRASS, utilizando toda la potencia de análisis de este último en un entorno de trabajo más amigable.

GvSIG es una solución de código abierto inicialmente impulsada por el gobierno regional de la Comunidad Valenciana, dentro de un proceso de migración a software libre de todos los sistemas informáticos de la organización; precisamente gvSIG abrevia la denominación Generalitat Valenciana Sistema de Información Geográfica. Desde el año 2010 la gestión y el mantenimiento del catálogo tecnológico gvSIG es llevado a cabo por la Asociación gvSIG, una asociación que engloba tanto entidades empresariales como no-empresariales (universidades, administraciones públicas, institutos geográficos, institutos tecnológicos, etc.).

2.1.3. Discusión

De los SGDC estudiados, MCH es una solución genérica y de libre distribución. Lo primero permite adaptarla a cualquier área de estudio. Con respecto a la adquisición de datos, la herramienta de serie, solamente permite hacerlo a través ficheros de texto plano en un formato determinado. Sin embargo, al ser una herramienta de libre distribución es posible personalizarla y adaptarla a unas necesidades concretas. Por ejemplo, importar datos en otros formatos. MCH también permite visualizar resultados y obtener informes, pero la obtención de datos está orientada hacia la climatología, es decir, está optimizada para consultar valores de ciertas localizaciones geográficas concretas en periodos de tiempo muy largos. Como se afirma en el capítulo 1, esta forma de acceder a los datos no es óptima para realizar análisis meteorológicos, donde se trabaja sobre áreas geográficas muy grandes, en periodos de tiempo cortos. Por otro lado, MeteoSIX sí dispone de un acceso optimizado a ciertos datos útiles para aplicaciones meteo-

rológicas. Sin embargo, se trata de una solución ad hoc para MeteoGalicia, y sólo dispone de datos meteorológicos de la comunidad gallega.

Con respecto a los GIS, cualquier solución que sea capaz de manejar archivos raster puede ser una buena alternativa. El modelo geográfico de campos (capítulo 1) detrás de la tecnología raster, es una solución ideal para representar variables continuas en tres dimensiones. Sin embargo, los GIS tienen un gran inconveniente, que no están diseñados para gestionar la dimensión tiempo. Aunque en meteorología los rangos temporales son considerablemente menores que en climatología, la dimensión tiempo es muy importante, ya que permite monitorizar la evolución de las variables meteorológicas, por ejemplo precipitación.

En definitiva, los SGDC no están diseñados para ofrecer acceso eficiente a los datos que se necesitan para realizar un análisis meteorológico. Por otro lado, los GIS no gestionan la dimensión tiempo, lo cual impide monitorizar la evolución de las variables. Ninguno de los dos modelos es adecuado para implementar un SGDM óptimo.

2.2. Formalización del modelo \mathcal{M}^3

Un SGDM se compone de los siguientes elementos: un repositorio que permita almacenar un volumen importante de datos meteorológicos, un mecanismo de consulta que permita acceder a ellos, y un mecanismo adquisición de datos para introducir nueva información. Además, es necesario controlar quien y de qué forma puede acceder a la información.

El modelo \mathcal{M}^3 se define como una tupla de varios elementos, en los términos de la ecuación 2.1:

$$\mathcal{M}^3 = \langle \mathcal{P}, \mathcal{V}, \mathcal{D}, \mathcal{R}, \mathcal{U}, \mathcal{S} \rangle \quad (2.1)$$

donde \mathcal{P} representa un conjunto de localizaciones geográficas; \mathcal{V} representa un conjunto de variables meteorológicas; \mathcal{D} representa el repositorio de datos meteorológicos; \mathcal{R} representa un conjunto de perfiles de acceso (roles) para acceder al repositorio \mathcal{D} ; \mathcal{U} representa un conjunto de usuarios con autorización para acceder al repositorio \mathcal{D} ; y \mathcal{S} representa una serie de servicios de acceso a la información del modelo \mathcal{M}^3 .

En los siguientes apartados se profundiza en cada uno de los elementos del modelo.

2.2.1. Localizaciones geográficas (\mathcal{P})

Los elementos p_i del conjunto \mathcal{P} (ecuación 2.2) representan las localizaciones que definen y acotan el área geográfica que abarca el modelo \mathcal{M}^3 . Es decir, \mathcal{P} es una discretización del área geográfica objeto de estudio del modelo.

$$p_i \in \mathcal{P} = \{p_1, p_2, \dots, p_n\}, n \in \mathbb{N} \quad (2.2)$$

Cada elemento p_i del conjunto \mathcal{P} (ecuación 2.3) está definido por sus coordenadas geográficas, longitud (lon) y latitud (lat), y su altura sobre el nivel del mar (hgt).

$$p_i = \langle lon, lat, hgt \rangle \quad (2.3)$$

2.2.2. Variables meteorológicas (\mathcal{V})

Cada elemento v_i del conjunto \mathcal{V} (ecuación 2.4) representa una variable meteorológica como puede ser la temperatura, la humedad relativa o la intensidad y dirección del viento.

$$v_i \in \mathcal{V} = \{v_1, v_2, \dots, v_n\}, n \in \mathbb{N} \quad (2.4)$$

Cada variable (ecuación 2.5) se define por un identificador único, v_{id} , y unas unidades de medida, v_{unit} .

$$v_i = \langle v_{id}, v_{unit} \rangle \quad (2.5)$$

2.2.3. Repositorio de datos meteorológicos (\mathcal{D})

\mathcal{D} (ecuación 2.6) representa un repositorio de información que almacena un conjunto finito de datos meteorológicos.

$$\mathcal{D} = \{d_1, d_2, \dots, d_n\}, n \in \mathbb{N} \quad (2.6)$$

La información meteorológica tiene atributos comunes con independencia de cual sea su origen, lo que permite plantear el siguiente postulado:

Postulado 2 *Cualquier dato meteorológico representa el **valor** (m) de una determinada **variable** meteorológica (v), en un **instante** de tiempo conocido (t), y en una determinada **localización** geográfica (p).*

Esta aseveración se cumple en todas las fuentes de datos estudiadas. Cuando se trata de una estación meteorológica o similar es muy claro, ya que se dispone de los valores de determinadas variables meteorológicas, como pueden ser la temperatura, la humedad relativa, etc., a lo largo del tiempo, en una ubicación claramente localizada. En el caso de modelos NWP o imágenes vía satélite, esto no es tan sencillo, ya que estas fuentes de datos aportan información relativa a un área geográfica y no solamente a uno o varios puntos. Sin embargo, tanto los

modelos NWP como las imágenes vía satélite y similares, basan su procesamiento en una discretización del área geográfica objeto de estudio en una malla (*grid*) de dos o tres dimensiones. Esto permite el tratamiento de los datos como un conjunto de puntos independientes, de manera análoga a como se hace cuando se trabaja con una estación meteorológica o similar.

Por tanto, si cada elemento d_i del conjunto \mathcal{D} representa una medida, esta se puede representar según la tupla que presenta la ecuación 2.7:

$$d_i = \langle p_i, v_i, t, m \rangle \quad (2.7)$$

donde p_i es un elemento del conjunto \mathcal{P} que contiene una serie de localizaciones geográficas (ver ecuación 2.2); v_i es un elemento del conjunto \mathcal{V} que contiene una serie de variables meteorológicas (ver ecuación 2.4); t representa un instante de tiempo ($t_0 + \Delta t$); y m representa un valor numérico ($m \in \mathbb{R}$).

2.2.4. Perfiles de acceso (\mathcal{R})

\mathcal{R} (ecuación 2.8) representa un conjunto de perfiles de acceso (roles) al repositorio \mathcal{D} .

$$\mathcal{R} = \{r_1, r_2, \dots, r_n\}, n \in \mathbb{N} \quad (2.8)$$

Cada elemento r_i de \mathcal{R} (ecuación 2.9) representa un perfil de acceso al repositorio \mathcal{D} . Cada perfil de acceso r_i define un subconjunto de datos del repositorio, $\mathcal{D}' \subset \mathcal{D}$, al que concede acceso para realizar un conjunto de operaciones, $\mathcal{S}' \subset \{\mathcal{S}_{in} \cup \mathcal{S}_{out} \cup \mathcal{S}_{del} \cup \mathcal{S}_{mod}\}$, explicadas más adelante en el apartado 2.2.6.

$$r_i = \langle \mathcal{S}', \mathcal{D}' \rangle \quad (2.9)$$

2.2.5. Usuarios (\mathcal{U})

\mathcal{U} (ecuación 2.10) representa un conjunto de usuarios con autorización de acceso al repositorio \mathcal{D} .

$$\mathcal{U} = \{u_1, u_2, \dots, u_n\}, n \in \mathbb{N} \quad (2.10)$$

Cada elemento u_i de \mathcal{U} (ecuación 2.11) representa un usuario con autorización para acceder a la información contenida en el repositorio \mathcal{D} . Cada usuario se define por un identificador único v_{id} y uno o varios perfiles de acceso al repositorio $\mathcal{R}' \subset \mathcal{R}$.

$$u_i = \langle u_{id}, \mathcal{R}' \rangle \quad (2.11)$$

2.2.6. Servicios de acceso (\mathcal{S})

Los usuarios necesitan un mecanismo para introducir y consultar datos en el repositorio. Además, es necesario controlar quien accede al repositorio y qué datos consulta. Para ello el modelo \mathcal{M}^3 define una serie de servicios que se encargan de realizar estas tareas. El conjunto \mathcal{S} se define como una tupla de varios elementos. Cada elemento representa una clase de servicios de acceso, en función de la tarea que realice. \mathcal{S} se define según los términos la ecuación 2.12:

$$\mathcal{S} = \langle \mathcal{S}_{in}, \mathcal{S}_{out}, \mathcal{S}_{del}, \mathcal{S}_{mod}, \mathcal{S}_{ctrl} \rangle \quad (2.12)$$

donde \mathcal{S}_{in} representa un conjunto de servicios de adquisición de datos, que permiten introducir datos en el repositorio; \mathcal{S}_{out} representa un conjunto de servicios de consulta para obtener datos en el repositorio; \mathcal{S}_{del} representa un conjunto de servicios de borrado, que permiten borrar información; \mathcal{S}_{mod} representa un conjunto de servicios de actualización de información; y \mathcal{S}_{ctrl} representa un conjunto de servicios de control para gestionar el acceso de los usuarios a los datos del repositorio.

Servicios de adquisición de datos (\mathcal{S}_{in})

El modelo \mathcal{M}^3 incluye un mecanismo de adquisición de datos que permite introducir información meteorológica en el repositorio \mathcal{D} . Este mecanismo se compone de uno o varios servicios en función del formato en el cual se permite introducir información en el repositorio. \mathcal{S}_{in} (ecuación 2.13) representa el conjunto de servicios de adquisición de datos.

$$\mathcal{S}_{in} = \{s_{in_1}, s_{in_2}, \dots, s_{in_n}\}, n \in \mathbb{N} \quad (2.13)$$

Cada elemento s_{in_i} de \mathcal{S}_{in} representa un único servicio de adquisición de datos que permite introducir nuevos datos en el repositorio en un formato específico. Dado que una medida está representada por la tupla de la ecuación 2.7, un servicio de adquisición de datos s_{in_i} permite la inserción de un nuevo elemento $\langle p_i, v_i, t, m \rangle$ en \mathcal{D} según la ecuación 2.14:

$$s_{in_i} = \mathcal{D} \leftarrow \mathcal{D} \cup \langle p_i, v_i, t, m \rangle \quad (2.14)$$

donde $p_i \in \mathcal{P}$ es la localización del dato que se quiere introducir en el repositorio \mathcal{D} ; $v_i \in \mathcal{V}$ es la variable meteorológica del dato que se quiere introducir en el repositorio \mathcal{D} ; t es un instante de tiempo concreto ($t_0 + \Delta t$); y m es el valor numérico de la nueva medida ($m \in \mathbb{R}$).

Dado que una localización está representada por la tupla de la ecuación 2.3, un servicio de adquisición de datos s_{in_j} permite la inserción de un nuevo elemento $\langle lon, lat, hgt \rangle$ en \mathcal{P} según la ecuación 2.15:

$$s_{in_j} = \mathcal{P} \leftarrow \mathcal{P} \cup \langle lon, lat, hgt \rangle \quad (2.15)$$

Por último, dado que una variable meteorológica está representada por la tupla de la ecuación 2.5, un servicio de adquisición de datos s_{in_k} permite la inserción de un nuevo elemento $\langle v_{id}, v_{unit} \rangle$ en \mathcal{V} según la ecuación 2.16:

$$s_{in_k} = \mathcal{V} \leftarrow \mathcal{V} \cup \langle v_{id}, v_{unit} \rangle \quad (2.16)$$

Servicios de consulta (\mathcal{S}_{out})

El modelo \mathcal{M}^3 incluye un mecanismo de consulta de los datos contenidos en el repositorio \mathcal{D} . Este mecanismo se compone de uno o varios servicios en función del formato en el cual puede consultarse la información del repositorio. \mathcal{S}_{out} (ecuación 2.17) representa el conjunto de servicios de consulta.

$$\mathcal{S}_{out} = \{s_{out_1}, s_{out_2}, \dots, s_{out_n}\}, n \in \mathbb{N} \quad (2.17)$$

Cada elemento s_{out_i} de \mathcal{S}_{out} representa un único servicio de consulta que permite obtener datos del repositorio en un formato específico. Dado que cada medida está representada por la tupla de la ecuación 2.7, un servicio de consulta s_{out_i} es la proyección de m sobre la restricción $p_i = p'_i \cap v_i = v'_i \cap t = t'$ en el conjunto \mathcal{D} , como muestra la ecuación 2.18):

$$s_{out_i} = \Pi_m(\sigma_{p_i=p'_i} \cap \sigma_{v_i=v'_i} \cap \sigma_{t=t'}(\mathcal{D})) \quad (2.18)$$

donde m representa el valor de la variable v'_i , en la localización p'_i , en el instante t' ; $p'_i \in \mathcal{P}$ es la localización concreta de la cual se quiere obtener información; $v'_i \in \mathcal{V}$ es la variable meteorológica concreta de la cual se quiere conocer el valor; y t' es un instante de tiempo concreto ($t_0 + \Delta t$).

Dado que una localización está representada por la tupla de la ecuación 2.3, un servicio de consulta s_{out_j} es la proyección de lon, lat y hgt , sobre la restricción $lon = lon' \cap lat = lat' \cap hgt = hgt'$ en el conjunto \mathcal{P} , como muestra la ecuación 2.19):

$$s_{out_j} = \Pi_{lon,lat,hgt}(\sigma_{lon=lon'} \cap \sigma_{lat=lat'} \cap \sigma_{hgt=hgt'}(\mathcal{P})) \quad (2.19)$$

donde lon' , lat' y hgt' son las coordenadas (longitud, latitud y altura sobre el nivel del mar) de la localización que se quiere consultar.

Por último, dado que una variable meteorológica está representada por la tupla de la ecuación 2.5, un servicio de consulta s_{out_k} es la proyección de v_{id} y v_{unit} , sobre la restricción $v_{id} = v'_{id}$ en el conjunto \mathcal{V} , como muestra la ecuación 2.20):

$$s_{out_k} = \Pi_{v_{id}, v_{unit}} (\sigma_{v_{id}=v'_{id}}(\mathcal{V})) \quad (2.20)$$

donde v'_{id} es el identificador de la variable meteorológica que se quiere borrar.

Servicios de borrado (\mathcal{S}_{del})

Además de poder consultar la información del repositorio \mathcal{D} , el modelo \mathcal{M}^3 permite eliminar información del repositorio si es necesario. Incluye uno o varios servicios de borrado en función de cuántos métodos de borrado de información de implementen. \mathcal{S}_{del} (ecuación 2.21) representa el conjunto de servicios de borrado.

$$\mathcal{S}_{del} = \{s_{del1}, s_{del2}, \dots, s_{deln}\}, n \in \mathbb{N} \quad (2.21)$$

Cada elemento s_{del_i} de \mathcal{S}_{del} se corresponde con un único servicio de borrado que permite suprimir datos del repositorio. Dado que cada medida está representada por la tupla de la ecuación 2.7, un servicio de borrado s_{del_i} permite el borrado de un elemento $d_i \in \mathcal{D}$ según la ecuación 2.22:

$$s_{del_i} = \mathcal{D} \leftarrow \mathcal{D} - (\sigma_{p_i=p'_i} \cap \sigma_{v_i=v'_i} \cap \sigma_{t=t'}(\mathcal{D})) \quad (2.22)$$

donde $p'_i \in \mathcal{P}$ es la localización del elemento que se quiere borrar; $v'_i \in \mathcal{V}$ es la variable meteorológica del elemento que se quiere borrar; y t' es un instante de tiempo concreto ($t_0 + \Delta t$) del elemento que se quiere borrar.

Dado que una localización está representada por la tupla de la ecuación 2.3, un servicio de borrado s_{del_j} permite el borrado de un elemento $p_i \in \mathcal{P}$ según la ecuación 2.23:

$$s_{del_j} = \mathcal{P} \leftarrow \mathcal{P} - (\sigma_{lon=lon'} \cap \sigma_{lat=lat'} \cap \sigma_{hgt=hgt'}(\mathcal{P})) \quad (2.23)$$

donde lon' , lat' y hgt' son las coordenadas (longitud, latitud y altura sobre el nivel del mar) de la localización que se quiere borrar.

Por último, dado que una variable meteorológica está representada por la tupla de la ecuación 2.5, un servicio de borrado s_{del_k} permite el borrado de un elemento $v_i \in \mathcal{V}$ según la ecuación 2.24:

$$s_{del_k} = \mathcal{V} \leftarrow \mathcal{V} - (\sigma_{v_{id}=v'_{id}}(\mathcal{V})) \quad (2.24)$$

donde v'_{id} es el identificador de la variable meteorológica que se quiere borrar.

Servicios de actualización (\mathcal{S}_{mod})

El modelo \mathcal{M}^3 permite actualizar información del repositorio si es necesario. Incluye uno o varios servicios de actualización en función de cuantos métodos de actualización de información de implementen. \mathcal{S}_{mod} (ecuación 2.25) representa el conjunto de servicios de actualización de datos.

$$\mathcal{S}_{mod} = \{s_{mod1}, s_{mod2}, \dots, s_{modn}\}, n \in \mathbb{N} \quad (2.25)$$

Cada elemento s_{mod_i} de \mathcal{S}_{mod} representa un único servicio de actualización que permite modificar datos del repositorio. Dado que cada medida está representada por la tupla de la ecuación 2.7, un servicio de actualización s_{mod_i} permite la modificación de un elemento $d_i \in \mathcal{D}$ según la ecuación 2.26:

$$s_{mod_i} = \mathcal{D} \leftarrow \left(\mathcal{D} - (\sigma_{p_i=p'_i} \cap \sigma_{v_i=v'_i} \cap \sigma_{t=t'}(\mathcal{D})) \right) \cup \langle p'_i, v'_i, t', m' \rangle \quad (2.26)$$

donde $p'_i \in \mathcal{P}$ es la localización del elemento que se quiere actualizar; $v'_i \in \mathcal{V}$ es la variable meteorológica del elemento que se quiere actualizar; t' es un instante de tiempo concreto ($t_0 + \Delta t$) del elemento que se quiere actualizar; y m' es el nuevo valor del elemento que se quiere actualizar.

Dado que una localización está representada por la tupla de la ecuación 2.3, un servicio de actualización s_{mod_j} permite la modificación de un elemento $p_i \in \mathcal{P}$ según la ecuación 2.27:

$$s_{mod_j} = \mathcal{P} \leftarrow \left(\mathcal{P} - (\sigma_{lon=lon'} \cap \sigma_{lat=lat'} \cap \sigma_{hgt=hgt'}(\mathcal{D})) \right) \cup \langle lon'', lat'', hgt'' \rangle \quad (2.27)$$

donde lon' , lat' y hgt' son las coordenadas (longitud, latitud y altura sobre el nivel del mar) de la localización que se quiere actualizar y lon'' , lat'' y hgt'' representan los nuevos valores.

Por último, dado que una variable meteorológica está representada por la tupla de la ecuación 2.5, un servicio de actualización s_{mod_k} permite la modificación de un elemento $v_i \in \mathcal{V}$ según la ecuación 2.28:

$$s_{mod_k} = \mathcal{V} \leftarrow \left(\mathcal{V} - (\sigma_{v_{id}=v'_{id}}(\mathcal{V})) \right) \cup \langle v''_{id}, v''_{unit} \rangle \quad (2.28)$$

donde v'_{id} es el identificador de la variable meteorológica que se quiere actualizar y v'_{unit} las unidades en las que se mide; y v''_{id} y v''_{unit} representan los nuevos valores.

Servicios de control (\mathcal{S}_{ctrl})

El modelo \mathcal{M}^3 incluye una serie de servicios que permiten controlar quién y en qué forma se accede a los datos. \mathcal{S}_{ctrl} (ecuación 2.29) representa el conjunto de servicios de control de acceso al repositorio.

$$\mathcal{S}_{ctrl} = \{s_{ctrl1}, s_{ctrl2}, \dots, s_{ctrln}\}, n \in \mathbb{N} \quad (2.29)$$

Cada elemento s_{ctrl_i} de \mathcal{S}_{ctrl} representa un único servicio de control de acceso al repositorio. Hay que incluir servicios que permitan realizar, como mínimo, las siguientes operaciones:

1. Asignar y revocar perfiles de acceso a los usuarios.
2. Crear y borrar perfiles de acceso.
3. Crear y borrar usuarios.
4. Control de acceso

Asignación y revocación de perfiles de acceso a usuarios El modelo \mathcal{M}^3 incluye un servicio s_{ctrl_i} que permite asignar uno o varios perfiles de acceso del conjunto \mathcal{R} a un usuario del conjunto \mathcal{U} , como muestra la ecuación 2.30:

$$s_{ctrl_i} = \mathcal{R}' \longleftarrow \mathcal{R}' \cup r_i \quad (2.30)$$

donde $\mathcal{R}' \subset \mathcal{R}$ es el conjunto de perfiles de acceso de un usuario determinado $u_i \in \mathcal{U}$, ver ecuación 2.11; y $r_i \in \mathcal{R}$ es el nuevo perfil de acceso que se quiere asignar a u_i .

El modelo \mathcal{M}^3 también incluye un servicio s_{ctrl_j} que permite revocar uno o varios perfiles de acceso del conjunto \mathcal{R} a un usuario, como muestra la ecuación 2.31):

$$s_{ctrl_j} = \mathcal{R}' \longleftarrow \mathcal{R}' - r_i \quad (2.31)$$

donde \mathcal{R}' es el conjunto de perfiles de acceso de un usuario determinado $u_i \in \mathcal{U}$; y $r_i \in \mathcal{R}$ es el perfil de acceso que se quiere revocar a u_i .

Creación y borrado de perfiles de acceso El modelo \mathcal{M}^3 incluye un servicio s_{ctrl_k} para crear nuevos perfiles de acceso en el conjunto \mathcal{R} , como muestra la ecuación 2.32:

$$s_{ctrl_k} = \mathcal{R} \longleftarrow \mathcal{R} \cup \langle \mathcal{S}', \mathcal{D}' \rangle \quad (2.32)$$

donde $\mathcal{D}' \subset \mathcal{D}$ y $\mathcal{S}' \subset \{\mathcal{S}_{in} \cup \mathcal{S}_{out} \cup \mathcal{S}_{del} \cup \mathcal{S}_{mod}\}$.

También incluye un servicio s_{ctrl_l} para borrar perfiles de acceso, como muestra la ecuación 2.33:

$$s_{ctrl_l} = \mathcal{R} \leftarrow \mathcal{R} - r_i \quad (2.33)$$

donde $r_i \in \mathcal{R}$.

Adicionalmente, en \mathcal{S}_{ctrl} se pueden incluir otros servicios que permitan asignar o revocar servicios concretos de \mathcal{S}_{in} , \mathcal{S}_{out} , \mathcal{S}_{del} o \mathcal{S}_{mod} a un elemento $r_i \in \mathcal{R}$.

Creación y borrado de usuarios El modelo \mathcal{M}^3 incluye un servicio s_{ctrl_m} para crear nuevos usuarios en \mathcal{U} , como muestra la ecuación 2.34:

$$s_{ctrl_m} = \mathcal{U} \leftarrow \mathcal{U} \cup \langle u_{id}, \mathcal{R}' \rangle \quad (2.34)$$

donde u_{id} es el identificador del nuevo usuario y $\mathcal{R}' \in \mathcal{R}$ es el conjunto de perfiles de acceso concedidos al nuevo usuario.

También incluye un servicio s_{ctrl_n} que permite borrarlos, como muestra la ecuación 2.35:

$$s_{ctrl_n} = \mathcal{U} \leftarrow \mathcal{U} - u_i \quad (2.35)$$

donde $u_i \in \mathcal{U}$.

Control de acceso Por último, el modelo también incluye un servicio s_{ctrl_o} para verificar si un usuario tiene acceso a una determinada información del repositorio, como muestra la ecuación 2.35:

$$s_{ctrl_o} = \Pi_{u_{id}}(\sigma_{u_{id}=u'_{id}} \cap \sigma_{r_i \subset \mathcal{R}'(\mathcal{U})}) \quad (2.36)$$

donde u'_{id} es el usuario cuyas credenciales se quieren verificar; \mathcal{R}' es el conjunto de perfiles de acceso del usuario u'_{id} , ver ecuación 2.11; y r_i es el perfil de acceso que se quiere comprobar.

2.3. GFA- \mathcal{M}^3

En este apartado se presenta la plataforma GFA- \mathcal{M}^3 . Esta plataforma es conocida como GFA Open Data [Guerrero-Higueras *et al.* (2013c)], ya que se diseñó como SGDM abierto. GFA- \mathcal{M}^3 es una implementación del modelo \mathcal{M}^3 propuesto en el apartado 2.2. Se construyó para organizar los datos meteorológicos, procedentes de diferentes fuentes, que recopila el GFA para llevar a cabo sus investigaciones.

Además de recopilar sus propios datos, el GFA recibe información de diferentes organismos públicos y de observadores voluntarios, como se explica detalladamente en el capítulo 3. El objetivo último del SGDM construido por el GFA es facilitar el acceso a esta información a la comunidad científica o a cualquier organismo o individuo que lo precise. Estas fuentes de datos, junto a investigadores y usuarios, representan a los actores principales en la interacción con el repositorio, como muestra el diagrama de casos de uso de la figura 2.1.

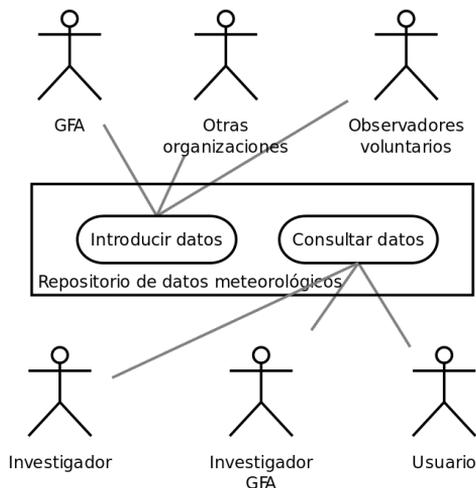


Figura 2.1: Diagrama de casos de uso.

El primer paso es definir el repositorio que almacenará toda la información, \mathcal{D} en la especificación del modelo expuesta en la ecuación 2.1. Posteriormente, hay que diseñar el mecanismo por el cual se accede a los datos (\mathcal{S}). Hay que definir un mecanismo que permita introducir información en el repositorio ($S_{in} \in \mathcal{S}$) y otro para poder consultarla después ($S_{out} \in \mathcal{S}$). Para ello, se define una primera capa de servicios que permite guardar datos en el repositorio, denominada *input services*, y una segunda capa de servicios que permite consultarlos y que se denomina *output services*. La figura 2.2 muestra la relación entre estas dos capas de servicios y el repositorio.

De cara a introducir información en el repositorio, se deben tener en cuenta las diferencias en los datos de entrada. Hay que manejar diferentes formatos: netCDF, utilizado normalmente por modelos NWP en sus salidas; Text/CSV, utilizado con frecuencia por estaciones meteorológicas, aunque la sintaxis exacta depende del fabricante; y HRIT, utilizado habitualmente para representar datos de satélite. En el momento de la redacción de este trabajo, el GFA trabaja



Figura 2.2: Relación entre los servicios de entrada/salida y el repositorio.

con estos tres formatos de entrada. Esto afecta directamente a los servicios de adquisición de datos en el repositorio (*input services*) que se han desarrollado y que se explican en los siguientes apartados. La introducción de nuevos servicios de adquisición de datos que permitan trabajar con otros formatos es una de las líneas en las que profundizar en futuros trabajos, como se explica en el capítulo 6.

También hay que considerar el formato de salida de los datos que se podrán obtener del repositorio. Los datos que los usuarios pueden obtener dependen directamente de los servicios de la capa de consulta (*output services*) de los que dispone el repositorio. En el momento de la redacción de este trabajo, estos servicios permiten la obtención de datos en formato CSV o netCDF, así como también obtener visualizaciones gráficas en formato PNG. Otra línea de trabajo futuro, como se explica en el capítulo 6, sería la inclusión de nuevos servicios de consulta que permitan trabajar con otros formatos.

La figura 2.3 muestra todos los componentes de la solución desarrollada, así como su interrelación. Coloreados en blanco, se muestran tanto el repositorio, que contiene los datos, como un *front-end* de acceso al mismo. Los elementos \mathcal{D} , \mathcal{P} , \mathcal{V} , \mathcal{R} y \mathcal{U} del modelo \mathcal{M}^3 , definido en los términos de la ecuación 2.1 expuesta en el apartado 2.2, se implementan en el propio repositorio. Coloreados en verde, se muestran los diferentes módulos de la capa de adquisición de datos (*input services*): *csv2sql*, *nc2sql*, *http2sql* y *hrit2sql*. Cada módulo transforma sus datos de entrada, cuyo formato depende del origen de los datos, en datos válidos para el modelo relacional del repositorio. Para ello hacen uso del front-end de acceso para introducir datos en el repositorio. Por último, coloreados en azul, se muestran los módulos de la capa de consulta: *sql2csv*, *sql2nc*, *csv2png* y *csv2kml*. Todos ellos permiten consultar la información del repositorio, haciendo uso del front-end de acceso. La implementación de los servicios de acceso al repositorio \mathcal{S} del modelo \mathcal{M}^3 se reparte entre los servicios de las capas de adquisición de datos y consulta, y el front-end de acceso al repositorio: los diferentes servicios de S_{in} , S_{del} y S_{mod} se implementan en los módulos de la capa de adquisición de datos; los servicios de S_{out} se implementan en los módulos de consulta; y los servicios de S_{ctrl} se implementan en el front-end de acceso al repositorio.

En los siguientes apartados se profundiza en los detalles de cada uno de los

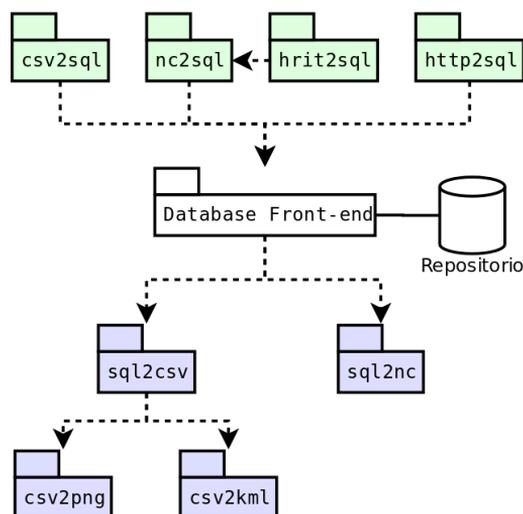


Figura 2.3: Diagrama de componentes.

componentes de la figura 2.3.

2.3.1. Repositorio de datos meteorológicos

Una vez identificadas todas las fuentes de información, hay que pensar en la forma de almacenar estos datos para su posterior utilización de una manera eficiente. Combinar datos meteorológicos de diferente origen es una tarea compleja, ya que como se ha visto en el capítulo 1, el formato de estos es totalmente diferente dependiendo de cual sea su procedencia. Para definir un modelo de datos capaz de almacenar toda esta información, es preciso identificar los puntos en común que tienen los datos que se desean guardar, con independencia de cómo estén almacenados: netCDF, CSV, etc.

Todos los datos que contienen información meteorológica tienen puntos en común. En el postulado 2 (apartado 2.2) se afirma que cualquier dato meteorológico representa el valor numérico de una determinada variable meteorológica (temperatura, humedad relativa, etc.); en un instante de tiempo conocido y en una determinada localización geográfica en los que se produce dicho valor. Esta aseveración se cumple en todos los formatos de datos estudiados.

Una vez identificados los atributos comunes de los datos meteorológicos, el siguiente paso es definir un repositorio donde almacenarlos todos. Existen varias alternativas. Por un lado están los repositorios estructurados, que siguen un esquema relacional clásico [Chen (1976); Codd (1970)], y por otro lado están los

repositorios no estructurados, conocidos como Not Only SQL (NoSQL) [Han *et al.* (2011)]. Estos últimos tienen una serie de ventajas relativas a la escalabilidad y la distribución de los datos muy interesantes para el problema que nos ocupa. Sin embargo, dada la fuerte relación entre variable meteorológica, ubicación temporal y localización de los datos, y sobretodo, la naturaleza crítica de muchas de las aplicaciones que los utilizan³, se ha optado por el modelo clásico, ya que es el único que garantiza la consistencia de las transacciones en todo momento [Cattell (2011)]. No obstante, evaluar el rendimiento utilizando un repositorio NoSQL, es una de las líneas en las que trabajar en el futuro, como se explica más adelante en el capítulo 6.

2.3.2. Input services

La plataforma GFA- \mathcal{M}^3 incluye módulos que permiten transformar datos con diferentes formatos y guardarlos en un repositorio relacional. En concreto, la capa de adquisición de datos de la plataforma GFA- \mathcal{M}^3 , se compone de los siguientes módulos: *csv2sql*, *nc2sql*, *http2sql* y *hrit2sql*; que se explican a continuación. La interrelación entre cada uno de ellos y el repositorio puede verse en la figura 2.3.

csv2sql

Este módulo se ha diseñado para introducir datos en formato texto o CSV en el repositorio. Como ya se ha explicado previamente en el capítulo 1, este formato de datos es muy común y se utiliza normalmente para estaciones meteorológicas y dispositivos similares. Como también se ha explicado, no existe un estándar, siendo el formato concreto muy diferente dependiendo del dispositivo o del fabricante.

En concreto, el módulo *csv2sql* está diseñado para trabajar con el formato de las estaciones meteorológicas de la Confederación Hidrográfica del Duero (CHD), el Instituto Tecnológico Agrario de Castilla y León (ITACyL) y la CHE. Aunque todos los ficheros recibidos tienen formato de texto, cada uno tiene su propia estructura y contenido. Partiendo de estos ficheros, *csv2sql* genera un fichero intermedio, en formato CSV, que posteriormente se utiliza para almacenar los datos en el repositorio.

nc2sql

El módulo *nc2sql* ha sido diseñado para introducir datos en formato netCDF en el repositorio. La mayor parte de los modelos NWP generan sus salidas utilizando este formato.

³Como la gestión de riesgos meteorológicos que se explica en el capítulo 5.

http2sql

Además de obtener información de organismos públicos; es interesante la información que aportan los observadores voluntarios. Existe un gran número de observadores aficionados, que dispone de estaciones meteorológicas propias, cuyas medidas pueden ser de gran interés a la comunidad científica. Para ello, se ha desarrollado una herramienta que permite a observadores voluntarios enviar los datos de sus estaciones meteorológicas a la plataforma GFA- \mathcal{M}^3 . Los datos se envían embebidos en peticiones Hypertext Transfer Protocol (HTTP), lo que significa que es necesario tener conexión a internet.

hrit2sql

El módulo hrit2sql permite introducir información en formato HRIT en el repositorio. Los datos obtenidos por receptores de datos vía satélite, habitualmente utilizan este formato. El GFA dispone de uno de estos receptores en sus instalaciones del Instituto de Medio Ambiente, Recursos Naturales y Biodiversidad (IMARENABIO).

Para guardar la información en el repositorio se transforman los datos en formato HRIT a formato netCDF, utilizando la aplicación *SatDatos*, incluida en el paquete *Nubes 4* desarrollado por *PraPro* [PraPro (2009)].

2.3.3. Output services

La plataforma GFA- \mathcal{M}^3 incluye módulos que permiten consultar la información del repositorio. En concreto, la capa de consulta de la plataforma GFA- \mathcal{M}^3 , se compone de los siguientes módulos: *sql2csv*, *sql2nc*, *csv2png* y *csv2kml*; que se detallan a continuación. La relación entre estos y el repositorio puede verse en la figura 2.3.

sql2csv y sql2nc

Los módulos *sql2csv* y *sql2nc* acceden a los datos del repositorio, y transforman la información obtenida de este en ficheros con formato CSV y netCDF respectivamente.

csv2png y csv2kml

Los módulos *csv2png* y *csv2kml* permiten obtener representaciones gráficas, de la información contenida en el repositorio, en formato PNG y KML respectivamente. Para ello hacen uso de las salidas que genera por el módulo *sql2csv*, de modo que no acceden directamente al repositorio.

Capítulo 3

Clasificación de datos meteorológicos

En este capítulo se realiza un estudio exhaustivo de las diferentes fuentes de información meteorológica a las que tiene acceso el GFA y cuyos datos utiliza en sus investigaciones. El objetivo de este análisis es obtener una clasificación de los datos meteorológicos que se incorporarán a la plataforma GFA- \mathcal{M}^3 , el SGDM desarrollado por el GFA, siguiendo la especificación del modelo \mathcal{M}^3 explicado en el capítulo 2.

También se persigue obtener una taxonomía general de las diferentes fuentes de información, que simplifique el desarrollo de otras plataformas siguiendo el modelo \mathcal{M}^3 .

3.1. Planificación

Antes de desarrollar un SGDM es necesario conocer el detalle de las distintas fuentes de información a las que se tiene acceso y las herramientas utilizadas para recopilar los datos. Hay que realizar un análisis en profundidad del volumen de información, de los distintos formatos, de la periodicidad con la que se reciben de los datos, etc. Fruto de este análisis se obtiene lo siguiente:

1. Definición del área de estudio del SGDM. El alcance geográfico de los datos meteorológicos de las distintas fuentes acotará el área de estudio cubierta por el SGDM.
2. Taxonomía de las diferentes fuentes de datos. El estudio de las diferentes fuentes de datos permite conocer con detalle el volumen, el formato y la periodicidad con que las distintas fuentes de información producen los datos.

3. Taxonomía de las herramientas utilizadas para la recopilación de los datos.
4. Taxonomía de todos los formatos de información utilizados. Conocer en profundidad los distintos formatos utilizados por cada fuente de datos permite seleccionar la mejor estrategia para construir el repositorio de información (Base de datos, sistema de ficheros, etc.).

En los siguientes apartados se realiza un análisis de diferentes fuentes de información, analizando las herramientas utilizadas para la recopilación de datos y los diferentes formatos utilizados. Con toda esa información se construye después una taxonomía que permite abordar el desarrollo de la plataforma GFA- \mathcal{M}^3 .

3.2. Análisis de los diferentes orígenes de información

En este apartado se realiza un estudio pormenorizado de las diferentes fuentes de información, tanto de observación, como de predicción, a las que tiene acceso el GFA. Primero se estudian los datos de observación, y después, los datos de predicción, analizando, cuando corresponda, las herramientas de adquisición de datos utilizadas.

3.2.1. Datos de observación

El GFA tiene a su disposición una gran cantidad de datos de observación fruto de todos los proyectos de investigación en los que ha trabajado y también de sus colaboraciones con diferentes organismos públicos y privados.

En el momento de la redacción de este trabajo, el GFA trabaja activamente con las siguientes fuentes de información meteorológica:

- *Estaciones meteorológicas*. Gracias a diferentes acuerdos de colaboración el GFA tiene acceso a los datos de las redes de estaciones meteorológicas del SAIH Duero, del SAIH Ebro y del ITACyL. Toda esta información se integra en la Red de observadores de Castilla y León (RdOCyL).
- *Radiómetro de microondas*. El GFA cuenta con un MMWR modelo Microwave Profiles 3000A (MP3000A).
- *Imágenes de satélite Meteosat Second Generation (MSG)*. El GFA dispone de un receptor de imágenes de satélite MSG en sus instalaciones en el IMARENABIO de la Universidad de León.
- *Radar meteorológico*. El GFA trabaja con un radar de la banda C modelo WR100-2/77.

- *Granizómetros*. El GFA dispone de una amplia de granizómetros en la cuenca del río Ebro.

En los siguientes apartados se profundiza en las diferentes fuentes de información, analizando el volumen de datos, el formato, la periodicidad, etc.

Estaciones meteorológicas (RdOCyL)

En las cuencas de los ríos Duero y Ebro, el estudio de los riesgos meteorológicos es extremadamente complicado debido a su diversa topografía, su gran extensión y la variedad de características climáticas que le afectan. Una de las mayores dificultades encontradas en la calibración de las herramientas necesarias para el seguimiento de estos riesgos es la incertidumbre presente en los datos de observación. En concreto, conocer con exactitud la zona en la que se registran nevadas, granizo o niebla.

Por todo lo anterior, en 2013 se creó la RdOCyL¹ en las cuencas de los ríos Duero y Ebro, cuyo objetivo es integrar en una plataforma única todos los datos de observación obtenidos de las estaciones meteorológicas de diferentes instituciones públicas y también de observadores voluntarios.

Diversos organismos públicos colaboran cediendo los datos de sus estaciones meteorológicas. Entre los organismos públicos que colaboran con la red se encuentran el SAIH Duero, el SAIH Ebro, el ITACyL y la Red de seguimiento del Cambio Global en Parques Nacionales. Además, como ya se ha mencionado, la red se alimenta con los datos aportados por observadores voluntarios.

Esta red ayuda a mejorar de manera significativa la calidad y cantidad de los datos de observación meteorológica de la mitad norte de la península ibérica. La figura 3.1 detalla las estaciones meteorológicas que aportan sus datos a la red en el momento de la redacción de este trabajo.

Red de estaciones meteorológicas del SAIH Duero El SAIH de la cuenca del río Duero es un organismo dependiente de la CHD, que gestiona las aguas de su demarcación hidrográfica, la parte española de la cuenca del río Duero.

El SAIH Duero envía datos de precipitación (mm), velocidad del viento (m/s), dirección del viento (Azimut Norte), temperatura ($^{\circ}C$), humedad relativa ($\%$), presión atmosférica media (mb), radiación solar (W/m^2) y evaporación (mm); de sus estaciones meteorológicas, que durante la redacción de este trabajo suman un total de 229.

Fruto de un acuerdo de colaboración entre ambas entidades, cada 15 minutos, el SAIH Duero envía a los servidores del GFA, vía File Transfer Protocol (FTP), un fichero de texto plano que contiene los valores de las citadas

¹En el marco del proyecto LE220A11-2 financiado por la Junta de Castilla y León.

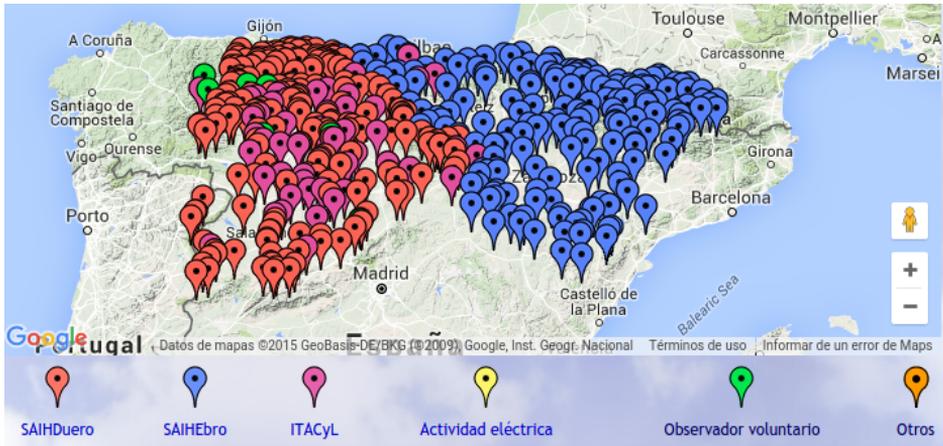


Figura 3.1: Puntos de observación de la RdOCyL.

variables en cada una de las estaciones. Los ficheros siguen la siguiente nomenclatura MMDDHHMMI.EYY. Por ejemplo, el fichero 07081015.E16 corresponde a los datos del 7 de agosto de 2016 a las 10:15 am.

El listado 3.1 muestra parte del contenido de uno de estos ficheros, con los valores de temperatura y precipitación de 4 estaciones. Las primeras líneas, marcadas con el carácter “*”, contienen una serie de cabeceras con información descriptiva. A continuación se incluyen las medidas a razón de una por línea. Cada línea se corresponde con el valor de una de las variables meteorológicas antes citadas en una estación concreta, de la que se incluye un identificador, sus coordenadas geográficas en formato Universal Transverse Mercator (UTM) y su altura sobre el nivel del mar.

```

1 *
2 *CONFEDERACION HIDROGRAFICA DEL DUERO
3 *
4 *Sistema automatico informacion hidrologica. Fichero estandar de intercambio.
5 *
6 *Incremento del tiempo: 15 min
7 *Numero de Intervalos : 1
8 *Archivo de Episodio que contiene variables del Tipo:
9 * P.- Precipitaciones (mm)
10 * V.- Velocidad Media de Viento (m/s)
11 * D.- Direccion de Viento (Azimut Norte)
12 * T.- Temperatura Media (C)
13 * H.- Humedad Relativa Media (%)
14 * A.- Presion Atmosferica Media (mb)
15 * R.- Radiacion Solar Media (w/m2)
16 * E.- Evaporacion (mm)
17 *
18 G 1 15
19 T "AF001_EATE01 " 517856 463766 1095.0 0 20.09

```

20	P	"AF001_ECPL01	"	517856	463766	1095.0	0	0.00
21	T	"AF002_EATE01	"	525544	463631	1055.0	0	18.81
22	P	"AF002_ECPL01	"	525544	463631	1055.0	0	0.00
23	T	"AF003_EATE01	"	545400	462760	1015.0	0	23.46
24	P	"AF003_ECPL01	"	545400	462760	1015.0	0	0.00
25	T	"AF004_EATE01	"	550618	461277	995.0	0	25.42
26	P	"AF004_ECPL01	"	550618	461277	995.0	0	0.00

Listado 3.1: Extracto de un fichero con datos del SAIH Duero.

Red de estaciones meteorológicas del SAIH Ebro El SAIH de la cuenca del río Ebro es un organismo dependiente de la CHE, que es el organismo encargado de la gestión, regularización y el mantenimiento de las aguas y riegos de la cuenca hidrográfica del río Ebro.

Fruto de un acuerdo entre ambas entidades, el SAIH Ebro pone a disposición del GFA los datos de temperatura ($^{\circ}\text{C}$) y de precipitación (mm) de 188 estaciones. Los datos se descargan de los servidores del SAIH cada 15 minutos, vía FTP, como ficheros de texto plano. Se genera un fichero por cada una de las variables anteriormente citadas. Estos ficheros contienen el valor de una variable concreta en cada una de las estaciones en intervalos de 15 minutos. A medida que se van obteniendo nuevas medidas, estas se añaden al fichero, que crece a medida que avanza el día. Los ficheros siguen la siguiente nomenclatura: ID_YYMMDD.txt, donde ID es un identificador que permite distinguir los ficheros con datos de temperatura (ID=ULE) y con datos de precipitación (ID=ULE_PQUIN). Por ejemplo, los ficheros ULE_20160708.txt y ULE_PQUIN_20160708.txt, contienen los datos de temperatura y precipitación respectivamente del 7 de agosto de 2016.

Los listados 3.2 y 3.3 muestran parte de los ficheros enviados por la CHE, con valores de temperatura y precipitación respectivamente. En ambos casos, cada línea incluye el identificador de la estación, el instante de tiempo en el que se ha tomado la medida (intervalos de 15 minutos), el valor de temperatura o precipitación y la calidad del valor tomado. Los ficheros no incluyen las coordenadas de las estaciones, su ubicación se obtiene a partir de su identificador. Por tanto, es necesario conocer de antemano la ubicación de las estaciones.

1	A021002TEMPE;	08/07/16	09:00;	0016,200;	BUENA
2	A021002TEMPE;	08/07/16	09:15;	0016,200;	BUENA
3	A021002TEMPE;	08/07/16	09:30;	0016,700;	BUENA
4	A022001TEMPE;	08/07/16	00:00;	0024,100;	BUENA
5	A022001TEMPE;	08/07/16	00:15;	0023,800;	BUENA
6	A022001TEMPE;	08/07/16	00:30;	0023,300;	BUENA

Listado 3.2: Extracto de un fichero con valores de temperatura del SAIH Ebro.

1	A005E85PQUIN;	08/07/16	09:00;	0000,000;	BUENA
2	A005E85PQUIN;	08/07/16	09:15;	0000,000;	BUENA
3	A005E85PQUIN;	08/07/16	09:30;	0000,000;	BUENA
4	A014Z85PQUIN;	08/07/16	00:00;	0000,000;	BUENA
5	A014Z85PQUIN;	08/07/16	00:15;	0000,000;	BUENA

6 A014Z85PQUIN; 08/07/16 00:30; 0000,000; BUENA

Listado 3.3: Extracto de un fichero con valores de precipitación del SAIH Ebro.

ITACyL El ITACyL, organismo dependiente de la Junta de Castilla y León, tiene como objetivo impulsar el sector agrario mediante el desarrollo de las nuevas tecnologías, la investigación y la transferencia de los avances científicos, favoreciendo la coordinación y colaboración con otras organizaciones públicas y privadas.

El ITACyL publica los valores de precipitación (mm), temperatura ($^{\circ}C$), humedad relativa (%), radiación (W/m^2) y velocidad (m/s) y dirección ($^{\circ}$) del viento de 53 estaciones meteorológicas. Están disponibles para cualquiera que quiera hacer uso de ellos². El GFA descarga estos datos de los servidores del ITACyL cada 30 minutos, vía FTP, como ficheros en formato CSV. El listado 3.4 muestra una parte de uno de los ficheros enviados por el ITACyL donde puede observarse el formato. La primera línea contiene un encabezado con la descripción de cada campo. El resto de líneas contienen las medidas de una determinada estación en un instante de tiempo dado. No se incluyen las coordenadas de las estaciones, su ubicación se obtiene a partir de su identificador y debe conocerse de antemano. Se genera un fichero diario con las medidas de todas las estaciones y de todas las variables anteriormente citadas. A medida que se van obteniendo nuevas medidas, se añaden al fichero, que crece a medida que avanza el día.

```
1 Código;Ubicación;Fecha (AAAA-MM-DD);Hora (HHMM);Precipitación (mm);Temperatura (C);
  Humedad relativa (%);Radiación (W/m2);Vel. viento (m/s);Dir. viento
2 AV01;Nava de Arévalo;2016-07-08;0600;0.00;16.43;86.70;82.40;1.29;234.60
3 AV01;Nava de Arévalo;2016-07-08;0630;0.00;17.98;80.30;164.20;1.91;253.40
4 AV01;Nava de Arévalo;2016-07-08;0700;0.00;19.87;71.79;254.20;2.61;260.20
5 AV101;Muñogalindo;2016-07-08;0030;0.00;15.36;91.90;0.00;2.27;282.70
6 AV101;Muñogalindo;2016-07-08;0100;0.00;15.00;93.10;0.00;1.92;281.00
7 AV101;Muñogalindo;2016-07-08;0130;0.00;14.68;92.10;0.00;2.15;258.00
```

Listado 3.4: Extracto de un fichero con datos meteorológicos del ITACyL

Radiómetro de microondas

El GFA dispone de un radiómetro de microondas (MMWR) modelo MP3000A (figura 3.2), diseñado por Radiometric Corporation, que permite obtener perfiles verticales de temperatura a diferentes niveles de altitud [Gascón (2015)]. El radiómetro MP3000A consta de 35 canales de microondas distribuidos en dos bandas de diferentes frecuencias. Una de ellas cuenta con 21 canales entre los 22 y 30 GHz. La otra, tiene 14 canales entre los 51 y 59 Ghz.

²<ftp://ftp.itacyl.es/>



Figura 3.2: Radiómetro modelo MP3000A.

Imágenes de satélite MSG

Los satélites de la serie MSG incluyen sensores con una resolución espacio-temporal y espectral superior a la de sus antecesores. Uno de estos sensores es el radiómetro Spinning Enhanced Visible and InfraRed Imager (SEVIRI) [Schmetz *et al.* (2002)]. Este sensor tiene 12 canales espectrales con una resolución temporal de 15 minutos y una calidad de datos de 10 bits. Dispone de un canal high-resolution visible (HRV) con una distancia de muestreo de 1 km, siendo esta de 3 km en el resto de canales.

El satélite operacional de la serie MSG está posicionado en órbita geoestacionaria (sobre el ecuador y longitud 0°). Desde esta posición puede ofrecer imágenes de toda Europa, oriente medio, el continente africano, la mayor parte del Océano Atlántico y la parte más oriental de América del Sur.

Para poder acceder a los datos es necesario disponer de un receptor MSG. El GFA dispone de uno de estos receptores en sus instalaciones en el IMARENABIO de la Universidad de León, que puede verse en la figura 3.3 (izquierda). Para procesar los datos recibidos se utiliza la herramienta *Nubes4*, desarrollada por *PraPro* [PraPro (2009)], que se encarga de guardar, y procesar como imagen, la información recibida en ficheros con formato HRIT (figura 3.3 derecha).

También se utiliza la herramienta *SatDatos*, que procesa parte de los ficheros HRIT generados por *Nubes4*, en concreto los que contienen la información relativa a la península Ibérica, y los convierte a formato netCDF.

Radar meteorológico

El GFA cuenta con un radar meteorológico de la banda C modelo WR100-2/77. Tiene una cobertura espacial que abarca un círculo teórico de 140 Km de

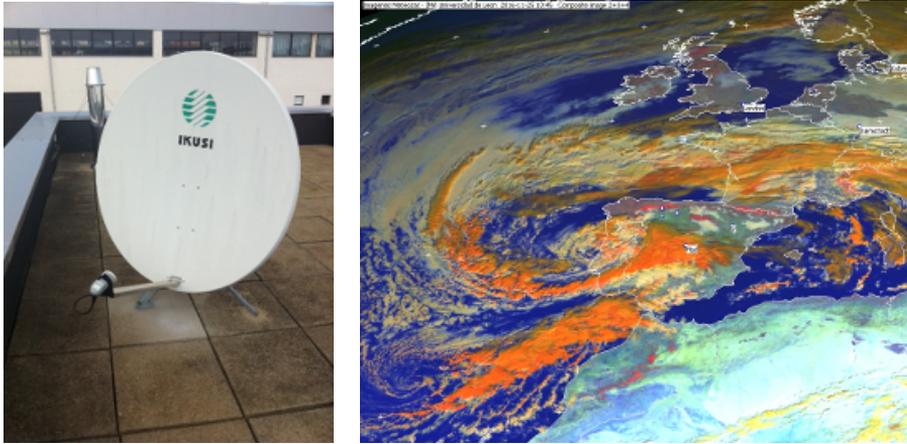


Figura 3.3: Receptor MSG (izquierda) e imagen MSG procesada por Nubes4 correspondiente al 25 de noviembre de 2016 a las 10:45 am (derecha).

radio. Las medidas del radar se obtienen en formato MDV utilizando la herramienta *TITAN* [Dixon y Wiener (1993)]. Durante la redacción de este trabajo, el radar está colocado en el aeródromo de Lugo, en la comunidad autónoma de Galicia, en el marco del proyecto Meteorisk (RTC-2014-1872-5) del Ministerio de Economía, Industria y Competitividad (MINECO). Anteriormente, en el marco de diversos proyectos de investigación del Plan Nacional de I+D+i, en colaboración con el Gobierno de Aragón y la CHE, el radar estuvo colocado en Zaragoza (figura 3.4), en la cuenca del río Ebro.

Granizómetros

El GFA dispone de una amplia red de granizómetros instalada en el suroeste de la provincia de Zaragoza y en la zona occidental de la provincia de Lérida (figura 3.5). El radar meteorológico del GFA, durante su período de operaciones en Zaragoza, daba cobertura completa a la red.

3.2.2. Datos de predicción

El GFA ejecuta el modelo MM5 en sus instalaciones en el IMARENABIO de la Universidad de León. También realiza cuatro ejecuciones diarias del modelo WRF en las instalaciones de la Fundación del Centro de Supercomputación de Castilla y León (FCSCCL).

En los siguientes apartados se abordan una serie de conceptos básicos sobre

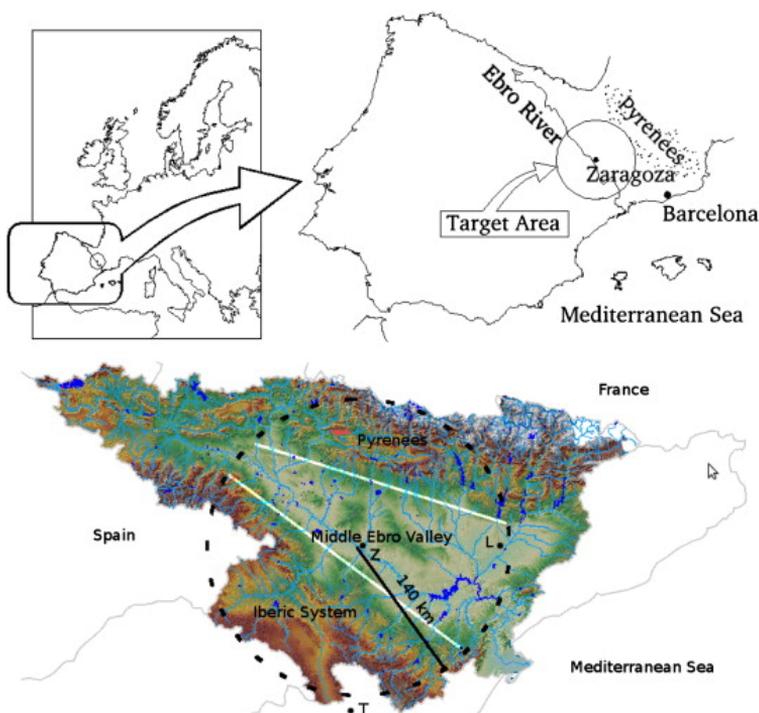


Figura 3.4: Ubicación y alcance del radar meteorológico del GFA.

los modelos MM5 y WRF y sus ejecuciones operativas para la obtención de pronósticos meteorológicos

MM5

MM5 [Dudhia (1993); Dudhia *et al.* (2015); Grell *et al.* (1995)] es un modelo de simulación de la atmósfera de área limitada, no hidrostático, sensible a las características del terreno y diseñado para predecir la circulación atmosférica a escala sinóptica, mesoescalar y regional. Se trata de un modelo a disposición de la comunidad científica y que durante varios años ha sido constantemente mejorado mediante contribuciones de usuarios de varias universidades y laboratorios del mundo.

Entre sus características están las siguientes:

- Capacidad de anidamiento múltiple.
- Dinámicas no hidrostáticas, lo que permite al modelo ser usado en una

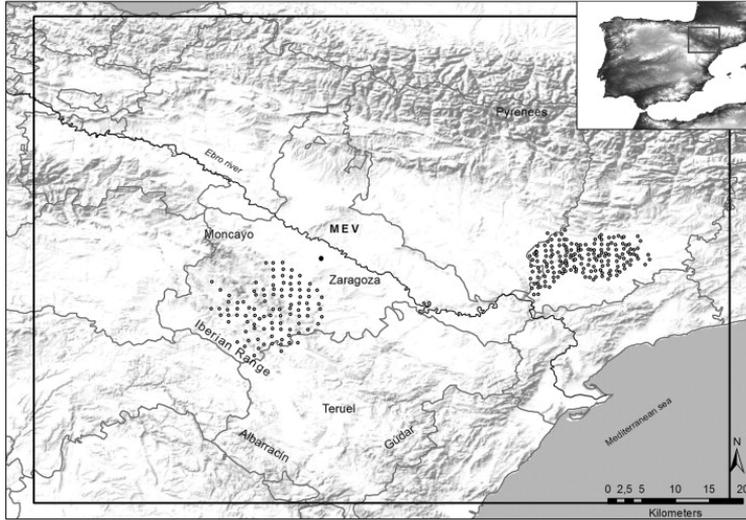


Figura 3.5: Redes de granizómetros en el suroeste de la provincia de Zaragoza y en la zona occidental de la provincia de Lérida.

escala de unos pocos kilómetros.

- Capacidad multitarea compartida (OpenMP) y máquinas de memoria distribuida (MPI).
- Una gran capacidad de asimilación de información.
- Múltiples opciones de parametrizaciones físicas.

MM5 se compone de varios módulos, que están referidos de manera colectiva bajo el nombre de sistema de simulación atmosférica MM5. Todos son necesarios y cada uno cumple una misión diferente:

- *TERRAIN* permite seleccionar el área geográfica de interés, teniendo en cuenta la topografía, usos del suelo y características físicas del terreno, en función de la época del año. El dominio seleccionado queda definido por las coordenadas geográficas de las fronteras y la resolución espacial tanto horizontal como vertical, seleccionada.

MM5 permite anidamiento múltiple con posibilidad de definir áreas de interés [Zhang *et al.* (1986)]. El anidamiento es necesario para realizar simulaciones en áreas de tamaños del orden de mesoescala. La resolución espacial de las condiciones iniciales fijadas desde un modelo de circulación general obliga a definir dominios con una resolución alta. Estos dominios no son

apropiados para predicciones en áreas del orden de mesoescala, sin embargo, si permiten establecer las condiciones de contorno necesarias para otros dominios con una resolución menor, lo cual no puede hacerse con modelos de circulación general.

Para ejecutar cualquier modelo numérico de predicción del tiempo en una región, se requiere fijar las condiciones de las fronteras laterales. En el MM5 las cuatro fronteras tienen especificados campos de viento horizontal, temperatura, presión y humedad; y pueden tener campos a escala de microfísica (que permiten desarrollar, por ejemplo, las nubes) si están disponibles. Por lo tanto, es fundamental que en las fronteras se disponga de un conjunto de valores iniciales para estos campos. Los valores de las fronteras se obtienen del análisis de tiempos futuros, o de simulaciones de otros modelos. Para pronósticos de tiempo real, las fronteras laterales dependerán de pronósticos de modelos globales, como GFS o ECMWF.

- *REGRID*. En este módulo se preparan los datos de inicialización del modelo y las condiciones de contorno, adaptándose a las características de los dominios definidos en *TERRAIN*. Los valores de los campos meteorológicos correspondientes a las condiciones iniciales, fijados por la salida de un modelo de mayor cobertura espacial (GFS, ECMWF, etc.), son interpolados a las mallas definidas, tanto en la componente horizontal como en la vertical. De este modo se establecen unívocamente las condiciones de partida, a partir de las cuales el modelo ejecutará la simulación.
- *INTERPF* realiza la interpolación vertical desde los niveles de presión hasta el sistema de coordenadas normalizadas sigma de MM5 (Figura 3.6). Se trata de la definición de un sistema de coordenadas, que sigue las irregularidades propias del terreno, adaptándose a los distintos tipos de orografía presente en los dominios de simulación. Una vez hecha esta conversión, se ejecuta la simulación.
- *MM5*. Este módulo contiene las ecuaciones físicas de pronóstico y diagnóstico que permiten realizar una predicción con un horizonte temporal prefijado. En la atmósfera existen una serie de fenómenos cuya escala espacial/temporal es inferior a la resolución fijada en las simulaciones. Por esta razón, se hace imprescindible el uso de parametrizaciones que permitan describir los fenómenos físicos que tienen lugar a estas pequeñas escalas, en función de los valores de los campos atmosféricos definidos y resueltos por el modelo.

En concreto, las parametrizaciones físicas con las que trabaja el MM5 son: parametrización de cúmulos, parametrización de la capa límite, esquema de humedad, esquema de radiación y esquema de superficie.

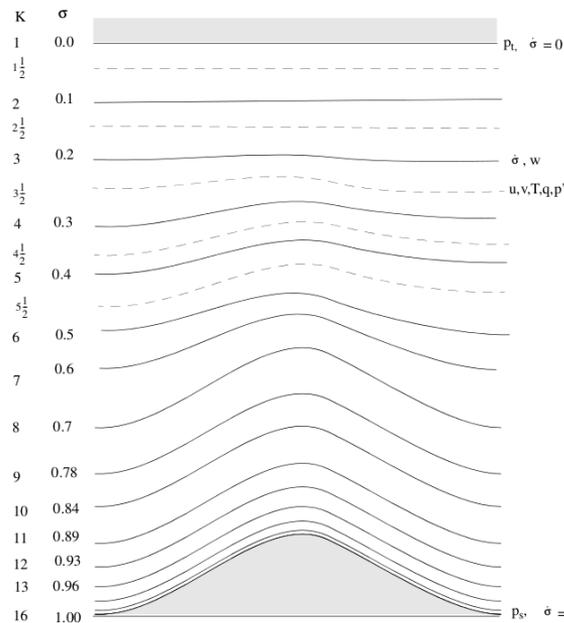


Figura 3.6: Coordenadas verticales sigma normalizadas.

Ejecución del modelo Antes de empezar a trabajar con el modelo WRF, el GFA realizaba predicciones a diario utilizando MM5. El modelo se ejecutaba en los servidores del GFA en sus instalaciones del IMARENABIO de la Universidad de León. Actualmente sólo se ejecuta de manera puntual en el marco de algunos proyectos de investigación o actividades docentes realizadas por el grupo. Para ejecutar el modelo se utiliza un script bash (*mm5operativo.sh*) que invoca a cada módulo en el orden adecuado.

Weather Research and Forecasting (WRF)

El modelo WRF es un modelo de simulación atmosférica no hidrostático de área limitada, sensible a las características del terreno y diseñado para predecir la circulación atmosférica a escala sinóptica, mesoescala y escala regional. Fue desarrollado en colaboración por la NOAA, el NCAR y otras organizaciones. El modelo WRF es la evolución de MM5.

La implementación del modelo está diseñada para trabajar en entornos de cálculo paralelo con memoria compartida, utilizando las directivas de Open Multi-Processing (OpenMP); y en entornos de cálculo paralelo con memoria distribuida, usando Message Passing Interface (MPI). Adicionalmente, tiene la capacidad de

combinar ambas tecnologías.

WRF se compone de una serie de módulos [Skamarock *et al.* (2008, 2005)], cada uno con una funcionalidad diferenciada y que deben ejecutarse para poder obtener un pronóstico:

- *GEOGRID* permite configurar el área geográfica de estudio.
- *UNGRIB* prepara los datos necesarios para inicializar el modelo (condiciones iniciales), que se obtienen normalmente de otro modelo con mayor cobertura espacial (escala sinóptica), como GFS o ECMWF .
- *METGRID* establece las condiciones de contorno adaptándose a las características de los dominios definidos por *GEOGRID*.
- *REAL* realiza la interpolación vertical partiendo de niveles de presión a un sistema de coordenadas sigma normalizadas (figura 3.6).
- *WRF*. El módulo homónimo WRF resuelve las ecuaciones físicas de pronóstico y diagnóstico que obtienen la predicción con un horizonte temporal prefijado. Para obtener un pronóstico, es imprescindible especificar una serie de parámetros que definen la física del modelo, como pueden ser la parametrización de cúmulos, la microfísica, etc. La tabla 3.1, muestra los principales parámetros que definen la física del modelo WRF.

Tabla 3.1: Principales parámetros que definen la física del modelo WRF

Parámetro	Nombre	Descripción
<i>ra_lw_physics</i>	Longwave radiation	Influencia de la radiación terrestre sobre el equilibrio radiativo.
<i>ra_sw_physics</i>	Shortwave radiation	Influencia de la radiación solar sobre el equilibrio radiativo.
<i>sf_sfclay_physics</i>	Surface layer	Capa superficial
<i>sf_surface_physics</i>	Land/water surface	Superficie de continente/océano
<i>bl_physics</i>	PBL	Capa límite planetaria
<i>cu_physics</i>	Cumulus parameterization	Parametrización de cúmulos
<i>mp_physics</i>	Microphysics	Microfísica
<i>diff_opt, km_opt</i>	Turbulence/diffusion	Turbulencia/difusión

El GFA ejecuta el modelo WRF en Caléndula, el cluster de cálculo paralelo de la FCSCCL. La FCSCCL utiliza el gestor de colas Sun Grid Engine (SGE) para gestionar el envío de trabajos a Caléndula.

Con el fin de asegurar una utilización óptima de los recursos de Caléndula, se ha desarrollado la herramienta PyWRFScheduler [Guerrero-Higueras *et al.* (2013a)], que se encarga de gestionar el envío de trabajos al gestor de colas SGE de manera secuencial, siguiendo un flujo de tareas definido previamente.

Ejecución del modelo El resultado de las ejecuciones diarias del modelo WRF están disponibles en el sitio web del GFA. Se realizan dos tipos de predicciones, sobre tres dominios con diferentes resoluciones espaciales (27 km , 9 km y 3 km respectivamente), que abarcan diferentes áreas geográficas del sudoeste de Europa (figura 3.7). Las predicciones del primer tipo se ejecutan a las 0:00 UTC y a las 12:00 UTC y tienen un horizonte temporal de 96 horas para el dominio 1 y de 48 horas para los dominios 2 y 3. Las predicciones del segundo tipo se ejecutan a las 6:00 UCT y las 18:00 UTC y tienen un horizonte temporal de 48 horas en todos los dominios.

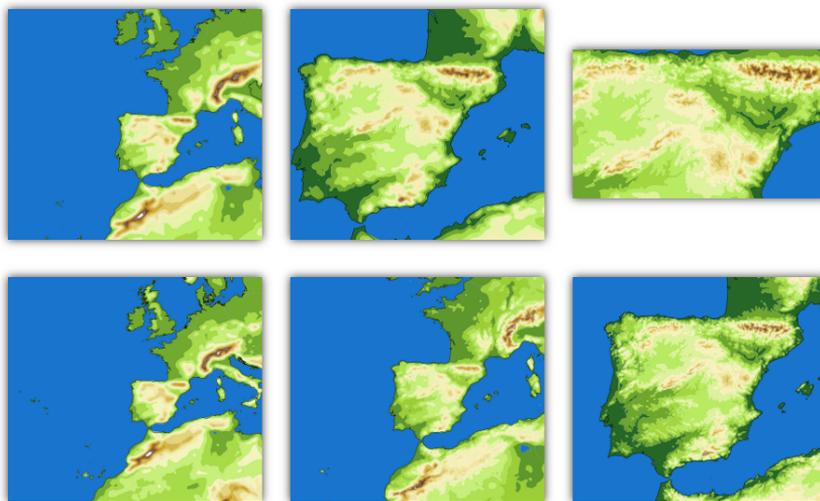


Figura 3.7: Dominios 1, 2 y 3 para predicciones tipo I (arriba). Dominios 1, 2 y 3 para predicciones tipo II (abajo).

En cada ejecución operativa se generan 50 archivos netCDF, 2 para los dominios 1 y 2, y 48 para el dominio 3 (uno por cada hora). Cada fichero ocupa aproximadamente 1 GB . Los ficheros del dominio 3, pese a contener solamente datos de una hora, son tan grandes como los ficheros de los dominios 1 y 2 debido a la mayor resolución del dominio 3.

3.3. Taxonomía

La tabla 3.2 muestra una caracterización de todas las fuentes de datos expuestas en los apartados anteriores. De cada una de ellas se detalla su categoría (datos de observación o predicción), la periodicidad con que se reciben los datos,

el número de ficheros recibidos en cada periodo y diariamente, el tamaño y el formato de cada fichero.

Tabla 3.2: Caracterización de los orígenes de datos analizados.

Origen	Observación	Predicción	Periodicidad	Ficheros/ periodo	Ficheros/ día	Tamaño/ fichero	Formato
Receptor MSG	✓		15 mins.	50	5184	50 MB	HRIT
SatDatos (Receptor MSG)	✓		15 mins.	1	95	1 MB	netCDF
SAIH Duero	✓		15 mins.	1	96	25 KB	Texto
SAIH Ebro	✓		15 mins.	*	2	500 KB	Texto
ITACyL	✓		30 mins.	*	1	70 KB	CSV
Observadores voluntarios	✓		N.D.	1	N.D.	N.D.	Texto
Radar (GFA)	✓		4 mins.	1	**	45 KB	MDV
Granizómetros (GFA)	✓		1 año***	1	N.D.	500 KB	XLS
Detector de rayos (GFA)	✓		****	*	1	N.D.	Binario
Radiómetro (GFA)	✓		2 mins.	5	3600	14 MB	CSV
MM5 (GFA)		✓	N.D.	3	N.D.	1 GB	netCDF
WRF (GFA)		✓	6 horas	50	200	1 GB	netCDF

*.- En cada periodo se añaden medidas a los ficheros diarios.

**.- Indeterminado, depende de cuanto tiempo este conectado.

***.- Se genera un único fichero con las estadísticas de toda la temporada.

****.- Indeterminada, solamente registra cuando detecta un pulso electromagnético.

N.D.- Indeterminada.

Para establecer una caracterización cualitativa de todas las fuentes de datos en lo que se refiere a volumen, tamaño y periodicidad de recepción de los datos se han definido una serie de escalas cualitativas en la tabla 3.3. Fruto de aplicar estas escalas sobre la caracterización cuantitativa de la tabla 3.2 se obtiene la caracterización cualitativa de la tabla 3.4.

La caracterización cualitativa que muestra la tabla 3.4 permite establecer la taxonomía general definida en la tabla 3.5. Los criterios de clasificación de esta taxonomía son el volumen, el tamaño, la periodicidad y el formato de los datos. Es importante destacar que esos valores cuantitativos pueden cambiar en ciertos entornos y bajo ciertas condiciones.

La tabla 3.6 presenta una caracterización de las herramientas utilizadas para recopilar los datos de las fuentes de datos que aparecen en la tabla 3.2. De cada herramienta se detalla su funcionalidad y el formato de las salidas en el caso de que sea necesario procesar los datos recibidos.

Tabla 3.3: Escalas cualitativa de periodicidad, ficheros recibidos por periodo y tamaño medio por fichero.

Periodicidad	Ficheros/periodo	Tamaño/fichero	Categoría
+2 horas	1 fichero	1–25 KB	(MB) Muy bajo/a
1–2 horas	2–4 ficheros	26–100 KB	(B) Bajo/a
16–60 mins	5–10 ficheros	101–1000 KB	(M) Medio/a
6–15 mins.	11–50 ficheros	1–1000 MB	(A) Alto/a
0–5 mins.	+50 ficheros	+1 GB	(MA) Muy alto/a

Tabla 3.4: Caracterización cualitativa de los orígenes de datos de observación analizadas.

Origen	Periodicidad	Ficheros/periodo	Tamaño/fichero
MSG Data	A	MA	A
SatDatos	A	MB	A
SAIH Duero	A	MB	MB
SAIH Ebro	A	MB	M
ITACyL	M	MB	B
Observadores voluntarios	–	MB	–
Radar	MA	MB	B
Granizómetros	MB	MB	M
Radiómetro	MA	M	A
MM5	–	B	MA
WRF	MB	MA	MA

Tabla 3.5: Taxonomía general de las fuentes de información meteorológica.

Origen	Periodicidad	Ficheros/periodo	Tamaño/fichero	Formato
Estaciones meteorológicas	A	MB	B	CSV, texto personalizado
Modelos NWP	MB	MA	MA	netCDF
Satélites meteorológicos	A	MA	A	HRIT
Radiómetro	MA	M	A	CSV
Radar	MA	MB	B	MDV
Granizómetros	MB	MB	M	XLS

En relación a los formatos utilizados por los diferentes orígenes de información y herramientas de procesamiento de los mismos, la tabla 3.7 presenta una

Tabla 3.6: Caracterización de las herramientas de recopilación de datos.

Herramienta	Utilidad	Salidas
Cliente FTP (lftp)	Descarga datos SAIH Ebro e ITACyL.	N.A.
Servidor FTP (vsftpd)	Recepción datos SAIH Duero.	N.A.
mm5operativo.sh	Ejecución MM5.	netCDF
TERRAIN	Definición de dominios.	–
REGRID	Preparación condiciones iniciales y establecimiento condiciones de contorno.	–
INTERPF	Interpolación vertical.	–
MM5	Ejecución del modelo.	–
PyWRFScheduler	Ejecución modelo WRF.	netCDF
GEOGRID	Definición de dominios.	–
UNGRIB	Preparación condiciones iniciales.	–
METGRID	Establecimiento condiciones de contorno.	–
REAL	Interpolación vertical.	–
WRF	Ejecución del modelo.	–
TITAN	Obtención de las medidas del radar.	MDV
Nubes4	Obtención de medidas MSG.	HRIT
SatDatos	Conversión de medidas MSG.	netCDF
Soft. radiómetro	Obtención de las medidas	CSV

N.A.- Los datos solamente se transfieren, no se procesan.

taxonomía mostrando las ventajas e inconvenientes de cada uno de ellos. De cada formato se detallan las siguientes propiedades: Si contiene metadatos sobre la información que contiene, si es multi-plataforma, si es escalable, si permite el acceso en paralelo, si hay un estándar claramente definido y si permite definir tipos de datos complejos.

Tabla 3.7: Taxonomía de los formatos de información meteorológica.

Formato	Criterios de clasificación					
	Autodescrito	Portable	Escalable	Acceso compartido	Estándar	Tipos complejos
<i>Formatos abiertos</i>		✓			✓	
netCDF	✓	✓	✓	✓	✓	✓
Imagen digital		✓				
XML		✓			✓	✓
JSON		✓			✓	✓
<i>Formatos propietarios</i>		✓				
HRIT					✓	
MDV					✓	
Imagen digital		✓				
<i>Formatos basados en texto</i>		✓				
CSV		✓			✓	
Personalizados		✓				

Capítulo 4

GFA- \mathcal{M}^3

Este capítulo presenta la arquitectura de la plataforma GFA- \mathcal{M}^3 , el SGDM construido por el GFA, presentada en el apartado 2.3, siguiendo la especificación del modelo \mathcal{M}^3 descrita en el apartado 2.2. Primero, se detalla la implementación de todos los componentes software, así como las diferentes herramientas utilizadas. Después, se define la infraestructura hardware necesaria para soportar la solución.

4.1. Infraestructura software

En el apartado 2.3 se identifican los principales componentes software de la plataforma GFA- \mathcal{M}^3 : Repositorio de datos meteorológicos y módulos de adquisición y consulta de datos en el mismo (*input services* y *Output services* respectivamente). En los siguientes apartados se profundiza en cada uno de ellos.

4.1.1. Repositorio de datos meteorológicos

Probablemente, la decisión más importante sea el tipo de repositorio. Como se explica en el capítulo 4 se ha optado por un modelo estructurado clásico, dejando para más adelante la evaluación de otro tipo de modelos como se detalla más adelante en el capítulo 6.

Una vez decidido el tipo de repositorio, el siguiente paso es definir el modelo de datos. Se ha utilizado una base de datos relacional alojada en un servidor *PostgreSQL* [Stinson (2001)]. Pueden utilizarse otros servidores de bases de datos, como pueden ser *MySQL*, *Oracle*, *SQL Server*, etc. Sin embargo, se ha optado por *PostgreSQL* por su probado rendimiento con bases de datos muy grandes y por su capacidad de particionado de tablas, cuya utilidad se comenta más adelante.

El servidor de bases de datos tiene que ejecutarse en un servidor físico que asegure cierto nivel de rendimiento, seguridad y disponibilidad. Los requisitos del servidor físico se definen más adelante en el apartado 4.2.

Esquema Entidad/Relación

Con respecto al contenido del repositorio, el postulado 2, expuesto en el apartado 2.2, permite identificar los atributos en común que tiene cualquier dato meteorológico: el **valor** numérico de una determinada **variable** meteorológica, la **localización** geográfica y el **instante** de tiempo en los que se produce dicho valor. Estos cuatro conceptos nos permiten identificar, entre otras, dos entidades para el modelo de datos de la plataforma GFA- \mathcal{M}^3 : *Place* y *Variable*.

Place La entidad Place representa un conjunto de localizaciones caracterizados por sus coordenadas geográficas, latitud y longitud, y por su altura sobre el nivel del mar.

Variable La entidad Variable representa un conjunto de variables meteorológicas como pueden ser la temperatura, la humedad relativa, la precipitación, la dirección del viento, etc.

La relación entre un determinado ejemplar de la entidad Place y un determinado ejemplar de la entidad Variable representa una medida, caracterizada por una marca de tiempo (*timestamp*) y un valor específico (*value*) para la variable. Por ejemplo, el 18 de octubre de 2013, a las 12:00 UTC, la temperatura en el punto identificado por las coordenadas 42°35'03" N y 5°39'00" O, ubicado en la ciudad de León (España), era de 20°C.

Cada entidad o relación en el modelo de datos tiene un determinado número de atributos que identifica cada ejemplar. La figura 4.1 muestra el diagrama E/R [Chen (1976)] del repositorio con todas las entidades identificadas y sus interrelaciones.

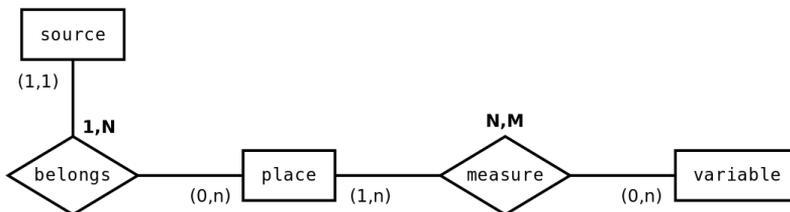


Figura 4.1: Diagrama E/R del repositorio.

En los siguientes apartados se profundiza en todas las entidades y relaciones del modelo de datos definido para el repositorio del SGDM.

Entidad Place La entidad *Place* representa una localización geográfica. Un ejemplar de la entidad *Place* puede representar la localización de una estación meteorológica, un punto de la malla (*grid*) definida por un modelo NWP o por una imagen de satélite MSG, etc.

Cada ejemplar de esta entidad se diferencia del resto por sus coordenadas geográficas, latitud y longitud; y su altura sobre nivel del mar. La figura 4.2 muestra el diagrama E/R de la entidad con sus atributos: longitud (*place_x*), latitud (*place_y*) y altura sobre nivel del mar (*place_z*). Se pueden añadir más atributos a la entidad, para completar la información de un ejemplar, como por ejemplo la ciudad más cercana, la provincia, el país, etc. En la figura 4.2, además de los mencionados longitud, latitud y altura sobre el nivel del mar; se incluyen un nombre (*name*), una descripción (*description*) y un origen (*source*). Este último se explica en el siguiente apartado. En el momento de la redacción de este trabajo, hay 151.253 ejemplares de la entidad *Place* identificados. La tabla 4.1 muestra 8 ejemplares a modo de ejemplo.

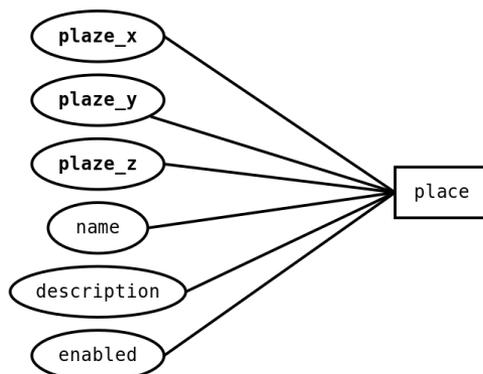


Figura 4.2: Entidad *Place*. Los atributos señalados en negrita (*place_x*, *place_y* y *place_z*) identifican unívocamente cada ejemplar.

Entidad Source Para diferenciar y agrupar el origen de los ejemplares de la entidad *Place*, se define la entidad *Source*. Un ejemplar de esta entidad representa un conjunto de ejemplares de la entidad *Place* pertenecientes a la misma organización o representando el mismo tipo de datos.

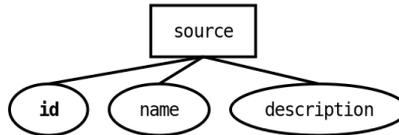
Los atributos que caracterizan un ejemplar de la entidad *Source* son simplemente un identificador (*id*), un nombre (*name*) y una descripción (*description*) como muestra el diagrama de la figura 4.3. Se pueden incluir más atributos, si se considera necesario, como por ejemplo, una persona de contacto, un responsable, etc. La tabla 4.2 muestra los orígenes de datos identificados en el momento de la

Tabla 4.1: Listado parcial (8 elementos) de ejemplares de la entidad Place.

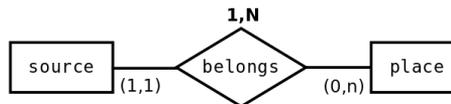
place_x	place_y	place_z	name	description	source
1.934349	42.424649	1120.00	A021	Río Segre en Puigcerdà	4
1.454244	42.361641	696.75	A022	Río Valira en Seo de Urgel	4
0.079492	42.437295	680.00	A040	Río Ara en Boltaña	4
0.186721	40.825569	580.00	A052	Río Matarraña en Beceite	4
-1.583327	41.252769	635.79	A055	Río Jiloca en Morata Jiloca	4
-1.875169	41.192056	756.00	A056	Río Mesa en Jaraba	4
-0.819056	42.863721	1187.00	A080	Río Veral en Zuriza	4
-1.953224	42.766251	481.00	A085	Río Ubagua en Riezu	4
...					

(151.253 total rows)

redacción de este trabajo.

**Figura 4.3:** Entidad Source. El atributo id, señalado en negrita, identifica unívocamente cada ejemplar.

Relación entre las entidades Source y Place Un ejemplar de la entidad Place tiene un origen único, lo que significa que debe tener un ejemplar de la entidad Source asociado. A su vez, un ejemplar de la entidad Source puede tener múltiples ejemplares de la entidad Place asociados con él. La relación entre Source y Place se muestra en la figura 4.4. La columna #Ejemplares de la tabla 4.2 muestra el número de ejemplares de la entidad Place asociados a cada ejemplar de la entidad Source.

**Figura 4.4:** Relación entre Source y Place.

Entidad Variable Las localizaciones geográficas pueden representarse utilizando ejemplares de la entidad Place, sin embargo, estas no tienen demasiado

Tabla 4.2: Listado de ejemplares de la entidad Source. La columna de la derecha muestra el número de ejemplares de la entidad Place asociados a cada ejemplar de Source.

id	name	#Ejemplares
1	Confederación Hidrográfica del Duero	229
3	Observador voluntario	11
4	Confederación Hidrográfica del Ebro	187
5	Instituto Tecnológico Agrario de Castilla y León	53
6	WRF model from GFA	150.773

interés por si mismos. El objetivo es conocer el valor de determinadas variables meteorológicas en una localización dada. Un ejemplar de la entidad Variable representa una variable meteorológica, como temperatura, precipitación, dirección del viento y velocidad del viento. Los atributos de la entidad Variable, como se muestra en la figura 4.5 son: un identificador (*id*), un nombre (*name*), una descripción (*description*) y una unidad de medida (*units*). La tabla 4.3 muestra las variables meteorológicas identificadas en el momento de la redacción de este trabajo (99).

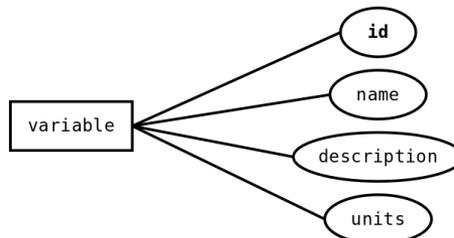


Figura 4.5: Entidad Variable. El atributo *id*, señalado en negra, identifica unívocamente cada ejemplar.

Tabla 4.3: Listado de ejemplares de la entidad Variable.

id	name	units
R	Radiación solar media	W/m^2
E	Evaporación	mm
VV10m	Velocidad media del viento	m/s
DV10m	Dirección media del viento	$^\circ$
VMAX10m	Velocidad máxima del viento	m/s
DMAX10m	Dirección velocidad max. viento	$^\circ$

Continúa en la siguiente página

Tabla 4.3: Listado de ejemplares de la entidad Variable (continuación).

id	name	units
TA	Temperatura del aire	°C
HR	Humedad relativa	%
TPR	Temperatura del punto de rocío	°C
PRES	Presión	hPa
PREC	Precipitación	mm
PRES_nmar	Presión reducida al nivel del mar	hPa
NIEVE	Capa de nieve	cm
TAMAX10m	Temperatura máxima en 10 min.	°C
TAMIN10m	Temperatura mínima en 10 min.	°C
TAMAX1h	Temperatura máxima en 1 hora	°C
TAMIN1h	Temperatura mínima en 1 hora	°C
GEO925	Altura geopotencial 925 hPa	m
GEO850	Altura geopotencial 850 hPa	m
GEO700	Altura geopotencial 700 hPa	m
HTAMAX1h	Hora y minuto de la temperatura máxima en 1 hora	HH:MM
HTAMIN1h	Hora y minuto de la temperatura mínima en 1 hora	HH:MM
VIS	Visibilidad	km
TPRE	Tiempo presente	
INSO	Insolación	minutes
RAGLOB	Radiación global	kJ/m ²
TS	Temperatura suelo	°C
TSS5cm	Temperatura subsuelo 5cm	°C
PLIQTP	Precipitación acumulada líquida	mm
PSOLTP	Precipitación acumulada sólida	mm
RVIENTO	Recorrido del viento	hm
BAT	Carga baterías	V
BATH	Carga baterías (horaria)	V
D	Dirección de Viento	Azimut norte
P	Precipitación	mm
T	Temperatura media	°C
H	Humedad relativa media	%
A	Presión atmosférica media	mb
V	Velocidad media de viento	m/s
Hi_Temp	Hi_Temp	°C
Temp_Out	Temp_Out	°C
Low_Temp	Low_Temp	°C
Out_Hum	Out_Hum	%
Dew_Pt	Dew_Pt	°C
Wind_Speed	Wind_Speed	km/h
Wind_Dir	Wind_Dir	
Wind_Run	Wind_Run	
Hi_Speed	Hi_Speed	km/h
Hi_Dir	Hi_Dir	
Wind_Chill	Wind_Chill	°C
Heat_Index	Heat_Index	°C
THW_Index	THW_Index	

Continúa en la siguiente página

Tabla 4.3: Listado de ejemplares de la entidad Variable (continuación).

id	name	units
THSW_Index	THSW_Index	
Bar	Bar	<i>hPa</i>
Rain	Rain	<i>mm</i>
Rain_Rate	Rain_Rate	<i>mm/h</i>
Solar_Rad	Solar_Rad	
Solar_Energy	Solar_Energy	
Hi_Solar_Rad	Hi_Solar_Rad	
UV_Index	UV_Index	
UV_Dose	UV_Dose	
Hi_UV	Hi_UV	
Heat_D-D	Heat_D-D	
Cool_D-D	Cool_D-D	
In_Temp	In_Temp	$^{\circ}C$
In_Hum	In_Hum	%
In_Dew	In_Dew	$^{\circ}C$
In_Heat	In_Heat	$^{\circ}C$
In EMC	In EMC	
In_Air_Density	In_Air_Density	<i>kg/m³</i>
ET	ET	
Soil_1_Moist	Soil_1_Moist	
Leaf_Wet_1	Leaf_Wet_1	
Leaf_Temp_1	Leaf_Temp_1	
Wind_Samp	Wind_Samp	
Wind_Tx	Wind_T	
ISS_Recept	ISS_Recept	
Arc_Int	Arc_Int	
temperature	temperature	$^{\circ}C$
humidity	humidity	%
dewpoint	dewpoint	$^{\circ}C$
barometer	barometer	<i>hPa</i>
windspeed	windspeed	<i>km/h</i>
direction	direction	$^{\circ}$
rainlastmin	rainlastmin	<i>mm</i>
dailyrain	dailyrain	<i>mm</i>
monthlyrain	monthlyrain	<i>mm</i>
yearlyrain	yearlyrain	<i>mm</i>
heatindex	heatindex	$^{\circ}C$
gustspeed	gustspeed	<i>km/h</i>
T2	Temperatura en superficie	<i>K</i>
t	Temperatura	$^{\circ}C$
t2m	Temperatura	$^{\circ}C$
rain	Precipitación	<i>mm</i>
rh	Humedad relativa	%
radiation	Radiación	<i>W/m²</i>
wind-speed	Velocidad del viento	<i>m/s</i>
wind-direction	Dirección del viento	$^{\circ}C$
pquin	Precipitación	<i>mm</i>

(99 total rows)

Relación entre las entidades Place y Variable Como se ha explicado en puntos anteriores, cualquier dato meteorológico se puede modelar utilizando ejemplares de las entidades Place y Variable. La relación entre un ejemplar de la entidad Place y un ejemplar de la entidad Variable en un instante de tiempo determinado representa una medida. La relación entre estas dos entidades, con sus propios atributos, se detalla en la figura 4.6. El atributo *datetime* representa el instante de tiempo. El atributo *value* es el valor numérico que toma la variable meteorológica representada por un ejemplar de Variable, en la localización representada por un ejemplar de Place, en el instante de tiempo marcado por el atributo *datetime*.

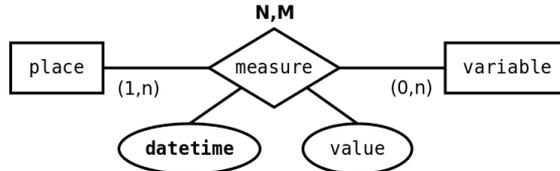


Figura 4.6: Relación entre Place y Variable.

La tabla 4.4 muestra 10 medidas a modo de ejemplo. La tabla 4.5 muestra el número de medidas, almacenadas en el repositorio en el momento de la redacción de este trabajo, por cada origen de datos. La agrupación de los datos por años permite observar la evolución del repositorio a lo largo del tiempo.

Tabla 4.4: Listado parcial de medidas (10 elementos).

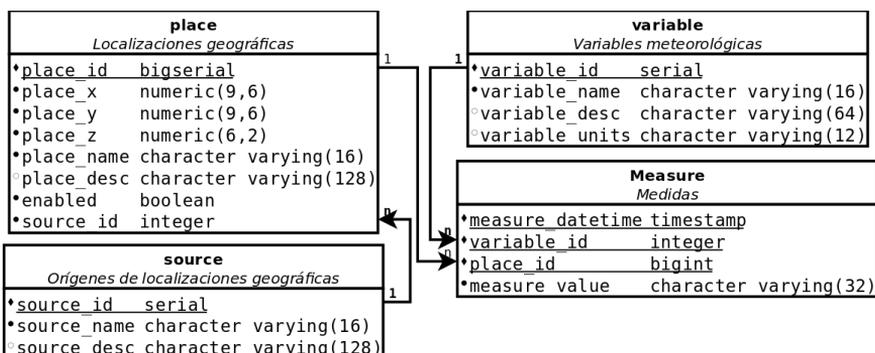
datetime	variable_name	place_x	place_y	value
2015-08-29 06:45:00	P	-4.048375	40.922394	0
2014-08-31 09:45:00	t	-1.193959	42.645959	19.1
2015-10-12 00:30:00	P	-5.399170	42.899208	0.15
2016-01-27 03:30:00	A	-4.031314	40.915792	833
2015-07-08 22:15:00	T	-5.186454	42.633060	13.93
2014-11-27 16:15:00	t	-1.970864	42.143596	6.7
2015-05-20 21:15:00	P	-5.661480	40.310346	0
2015-10-26 18:30:00	P	-4.196146	41.763668	0
2016-03-11 11:45:00	T	-4.758911	42.935513	2.4
2015-07-22 03:45:00	T	-5.373080	42.517595	19.16
...				
(99.891.247 total rows)				

Tabla 4.5: Medidas almacenadas por cada origen de datos (Source).

Source	2012	2013	2014	2015	2016	Total
CHD	1.424.204	13.359.003	12.985.047	13.049.549	4.336.555	45.154.358
Volunt.	4290	6.960.412	8.310.848	6.257.034	1.365.384	22.901.838
CHE		3.080.741	6.608.765	10.762.654	5.897.243	26.349.403
ITACyL		250.590	1.941.298	2.027.712	1.115.277	5.334.877
WRF					150.771	150.771
Total	1.428.494	23.650.746	29.845.958	32.096.949	12.865.230	99.891.247

Esquema relacional

La figura 4.7 muestra el esquema relacional de la base de datos PostgreSQL de la plataforma GFA- \mathcal{M}^3 . Es el resultado de implementar el modelo E/R expuesto en el apartado anterior.

**Figura 4.7:** Esquema relacional del repositorio.

Introducción de datos en el repositorio

Se han implementado procedimientos almacenados¹ para crear nuevos ejemplares de la entidad Place y nuevas medidas, estas últimas definidas por la relación entre ejemplares de Place y Variable. Se ha optado por su implementación como procedimientos almacenados para, de este modo, dejar todo el control de la integridad referencial en manos propio modelo de datos. De esta manera, los módulos de adquisición de datos (*input services*) son totalmente independientes del

¹Un procedimiento almacenado es un programa almacenado físicamente en una base de datos. Su implementación varía de un gestor de bases de datos a otro.

modelo de datos.

insert_place El procedimiento *insert_place* permite introducir nuevos ejemplares de la entidad Place. Para invocar a *insert_place* hay que aportar una serie de argumentos: Las coordenadas geográficas y la altitud sobre el nivel del mar de la nueva localización; un nombre; una descripción; y el origen al que pertenece, indicando en este caso el identificador del ejemplar de la entidad Source que referencia al origen. Si las coordenadas introducidas no se corresponden con las de una localización existente, se crea un nuevo ejemplar de la entidad Place, en otro caso se actualizan los datos de la localización. El procedimiento devuelve el identificador (*place_id*, ver figura 4.7) de la nueva localización.

insert_measure *insert_measure* permite introducir medidas en el repositorio. Para invocar a *insert_measure* hay que aportar una serie de argumentos: la variable meteorológica (referencia a la entidad Variable); las coordenadas de la localización (referencia a la entidad Place); el *timestamp* de la nueva medida; y el valor numérico de la medida. Si la variable a la que se hace referencia no existe, se crea un nuevo ejemplar de la entidad Variable.

insert_volunteer_measure Hay un tercer procedimiento almacenado, llamado *insert_volunteer_measure*, que también permite introducir medidas en el repositorio. Este procedimiento se invoca cuando es un observador voluntario el que envía datos al repositorio. A estos últimos, no se les permite enviar medidas de variables no definidas en el modelo. Es utilizado por el servicio *http2sql* que se explica más adelante en el apartado 4.1.2. *insert_volunteer_measure* es muy similar a *insert_measure*, con la salvedad de que no permite insertar nuevas variables.

4.1.2. Input services

En el apartado 2.3 se definen los módulos de adquisición de datos en el repositorio de la plataforma GFA- \mathcal{M}^3 : *csv2sql*, *nc2sql*, *http2sql* y *hrit2sql*; acorde con las especificaciones del modelo \mathcal{M}^3 definidas en el apartado 2.2.

csv2sql

Este módulo permite introducir datos en formato CSV en el repositorio. En concreto, está preparado para procesar los datos de las estaciones meteorológicas del SAIH Duero, del SAIH Ebro y del ITACyL. Para ello, partiendo de los datos de entrada, se genera un fichero intermedio en formato CSV, adaptado a las necesidades del modelo de datos definido en el apartado 4.1.1. La figura 4.8 muestra

la estructura del módulo `csv2sql`. En los siguientes apartados se profundiza en cada una de las partes. El formato concreto de los ficheros con las medidas, su volumen y su periodicidad, para cada origen de datos, está detallado en el capítulo 3. Dada la naturaleza periódica de los datos recibidos, se puede programar la ejecución del módulo `csv2sql`, por ejemplo, utilizando el demonio *Cron* de Unix.

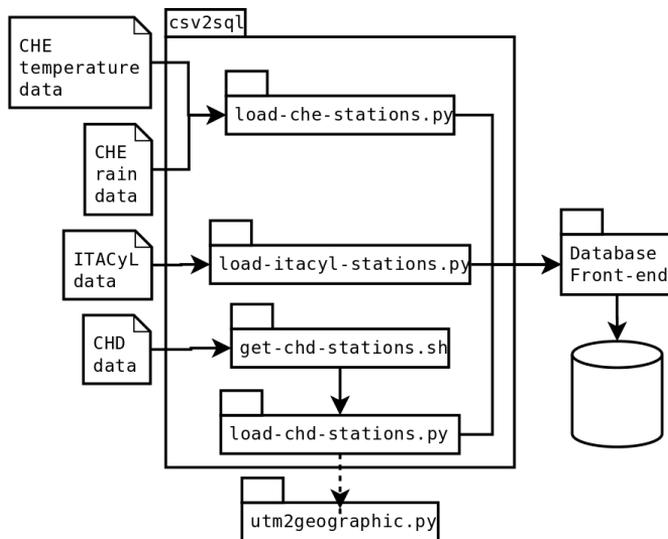


Figura 4.8: Arquitectura del módulo `csv2sql`.

Tratamiento de los datos del SAIH Duero Los datos de las estaciones meteorológicas del SAIH Duero son recibidos vía FTP en el servidor de la plataforma GFA- \mathcal{M}^3 . Se recibe un fichero cada quince minutos con las medidas del último cuarto de hora. Cada fichero contiene los datos del último cuarto de hora de todas sus estaciones. El formato y la periodicidad de los ficheros se detallan en el capítulo 3.

El script Bash `get-chd-stations.sh` (figura 4.8) se encarga de recoger los ficheros recibidos de la CHD, y dejarlos en la ruta apropiada para que el script Python `load-chd-stations.py` (figura 4.8) los procese. El algoritmo 1 muestra los pasos que sigue `load-chd-stations.py` para almacenar las medidas de un fichero en el repositorio. Se recorre el fichero línea a línea (línea 1), descartando las cabeceras. Por cada línea se realiza lo siguiente: se divide la línea en elementos, utilizando los espacios en blanco como separador (línea 2); de cada línea se toman las coordenadas y el nombre de la estación, y la variable meteorológica y el valor de la medida (líneas 3–7); el origen de los datos se marca directamente como

SAIH Duero (línea 8); a partir de las coordenadas y el nombre de la estación, se obtiene el identificador de la estación utilizando el procedimiento `insert_place` (línea 9); la fecha/hora de los datos se toma a partir del nombre del fichero (línea 10); por último, se guardan las medidas en el repositorio utilizando el procedimiento `insert_measure` (línea 11). Con la llamada a `insert_place`, si la estación no estuviera registrada en el repositorio, se crearía una nueva localización, como se explicó en el apartado anterior.

Algoritmo 1 Procesamiento de ficheros del SAIH Duero (`load-chd-stations.py`).

Input: *inputfile*

```

1: for all line in inputfile do
2:   items ← split(line, '_')                                ▷ Split line into a list of items
3:   lon, lat ← utm2geographic(items[3], items[4], 30, 'N')    ▷ Convert UTM coordinates
4:   hgt ← items[5]
5:   name ← desc ← items[1]
6:   variable ← items[0]
7:   value ← items[7]
8:   source ← 1                                               ▷ SAIH Duero
9:   place ← insert_place(lon, lat, hgt, name, desc, source)
10:  datetime ← dateformat(basename(inputfile), 'YY-MM-DD HH:MI:SS')
11:  insert_measure(datetime, variable, place, value)
12: end for

```

Las coordenadas de las estaciones del SAIH Duero están en formato UTM. `load-chd-stations.py` utiliza la función `utm2geographic` (línea 3 del algoritmo 1), del módulo homónimo `utm2geographic`, para de traducir coordenadas en formato UTM a coordenadas geográficas.

utm2geographic El módulo `utm2geographic` contiene funciones para convertir coordenadas UTM a coordenadas geográficas y viceversa. La conversión de dos coordenadas UTM X e Y en coordenadas geográficas se realiza siguiendo los pasos que se enumeran a continuación:

1. Dados los datos básicos de la geometría del elipsoide de Hayford, semieje mayor $a = 6378388,0$ y semieje menor $b = 6356911,946130$, se calculan la segunda excentricidad y el radio polar de curvatura:

$$\text{Segunda excentricidad, } e' = \frac{\sqrt{a^2 - b^2}}{b}$$

$$\text{Radio polar de curvatura, } c = \frac{a^2}{b}$$

2. Tratamiento previo de las coordenadas UTM X e Y :

- Eliminación del retranqueo del eje X : $X = X - 500000$.

– Para el eje Y la eliminación del retranqueo sólo se realiza en el caso de que se opere con coordenadas UTM correspondientes al hemisferio sur, en cuyo caso: $Y = Y - 10000000$.

3. Cálculo del meridiano central del huso: $\lambda_0 = \text{Huso} \cdot 6 - 183$. Es necesario conocer el uso UTM, o zona UTM al que pertenecen las coordenadas a X e Y . Para la comunidad de Castilla y León, el huso UTM es 30.
4. Cálculo de parámetros según las ecuaciones de Coticchia-Surace:

$$\begin{aligned} \varphi' &= \frac{Y}{6366197,724 \cdot 0,9996} \\ v &= \frac{c}{(1 + e'^2 \cdot \cos^2 \varphi')^{\frac{1}{2}}} \cdot 0,9996 \\ a &= \frac{X}{v} \\ A_1 &= \sin(2 \cdot \varphi') \\ A_2 &= A_1 \cdot \cos^2 \varphi' \\ J_2 &= \varphi' + \frac{A_1}{2} \\ J_4 &= \frac{3 \cdot J_2 + A_2}{4} \\ J_6 &= \frac{5 \cdot J_4 + A_2 \cdot \cos^2 \varphi'}{3} \\ \alpha &= \frac{3}{4} \cdot e'^2 \\ \beta &= \frac{5}{3} \cdot \alpha^2 \\ \gamma &= \frac{35}{27} \cdot \alpha^3 \\ B_\Phi &= 0,9996 \cdot c \cdot (\varphi' - \alpha \cdot J_2 + \beta \cdot J_4 - \gamma \cdot J_6) \\ b &= \frac{Y - B_\Phi}{v} \\ \zeta &= \frac{e'^2 \cdot a^2}{2} \cdot \cos^2 \varphi' \\ \xi &= a \cdot \left[1 - \frac{\zeta}{3} \right] \\ \eta &= b \cdot (1 - \zeta) + \varphi' \\ \sin h \xi &= \frac{e^\xi - e^{-\xi}}{2} \end{aligned}$$

$$\Delta\lambda = \arctan \frac{\sin h \xi}{\cos \eta}$$

$$\tau = \arctan(\cos \Delta\lambda \cdot \tan \eta)$$

5. Cálculo final de las coordenadas geográficas, longitud (λ) y latitud (φ):

- La longitud, λ , se calcula con la expresión $\lambda = \Delta\lambda + \lambda_0$. Previamente hay que pasar $\Delta\lambda$ a grados sexagesimales, λ_0 ya está representado en grados sexagesimales.
- La latitud φ se calcula con la siguiente expresión, cuyo resultado hay que pasar después a grados sexagesimales:

$$\varphi = \varphi' + \left[1 + e'^2 \cdot \cos^2 \varphi' - \frac{3}{2} \cdot e'^2 \cdot \sin \varphi' \cos \varphi' \cdot (\tau - \varphi') \right] \cdot (\tau - \varphi')$$

Tratamiento de los datos del SAIH Ebro Los datos de las estaciones meteorológicas del SAIH Ebro se descargan de sus servidores vía FTP y se procesan en el servidor de la plataforma GFA- \mathcal{M}^3 . Hay un fichero diario por cada variable meteorológica, en concreto temperatura en superficie y precipitación. Los ficheros crecen a medida que se incorporan nuevas medidas cada treinta minutos. El script Python *load-che-stations.py* (figura 4.8) se encarga de descargar los ficheros, procesarlos y guardar los datos en el repositorio. Para descargar los ficheros, *load-che-stations.py* hace uso de la herramienta *lftp*. *lftp* es un cliente FTP en modo consola que permite la automatización de descargas mediante ficheros de configuración. Partiendo de un a plantilla, se genera un fichero de configuración, que se pasa como argumento a *lftp*, por cada fichero a descargar. El listado 4.1 muestra el contenido de la plantilla a partir de la cual se generan ficheros de configuración para *lftp*. Las credenciales de acceso del GFA: nombre de usuario, contraseña y dirección IP del servidor de la CHE que aloja los datos (línea 1), fueron asignadas en el marco de un acuerdo de colaboración entre ambas entidades. Hay que sustituir las cadenas alfanuméricas DIRECTORY y FILE de la plantilla, líneas 2–3, por la ruta local en la cual se quiere depositar el fichero, y por la ruta remota del fichero que se quiere descargar respectivamente. Dado que hay un fichero por cada variable meteorológica, se generan dos ficheros de configuración y se invoca a *lftp* dos veces, una con cada fichero.

```

1 open -u USER,PASSWORD IP
2 lcd DIRECTORY
3 get FILE
4 quit

```

Listado 4.1: Plantilla para *lftp*.

Una vez descargados los datos, se procesan y se guardan en el repositorio. El algoritmo 2 detalla los pasos que sigue *load-che-stations.py* para almacenar

las medidas de un fichero una vez descargado. Se recorre el fichero línea a línea (línea 1), descartando las cabeceras. Por cada línea se siguen los siguientes pasos. Primero, se divide la línea en elementos, utilizando el carácter “;” como separador (línea 2). En concreto de cada línea se toma el nombre de la estación a la que pertenecen los datos, y la marca de tiempo y el valor de la medida (líneas 3–5). Para guardar los datos se necesita el identificador de la estación, de modo que hay que obtenerlo en el repositorio, a partir del nombre de la estación (línea 6). Por último, se guardan los datos en el repositorio utilizando el procedimiento *insert_measure* explicado en el apartado anterior (línea 7). Dado que hay un fichero por variable meteorológica, como se detalla en el capítulo 3, el valor de *variable* (línea 7) viene dado por el propio fichero que se esté procesando en cada momento.

Algoritmo 2 Procesamiento de ficheros del SAIH Ebro (load-che-stations.py).

Input: *inputfile*

```

1: for all line in inputfile do
2:   items ← split(line, ‘;’)                                ▷ Split line into a list of items
3:   station ← items[0]
4:   datetime ← items[1]
5:   value ← items[2]
6:   place ← SELECT place_id FROM place WHERE place_name = station
7:   insert_measure(datetime, variable, place, value)
8: end for

```

De igual modo que a la hora de descargar los datos, el algoritmo anterior se ejecuta una vez por cada variable meteorológica.

Tratamiento de los datos del ITACyL Los ficheros que contienen los datos de las estaciones meteorológicas del ITACyL se descargan comprimidos de sus servidores vía HTTP y se procesan en el servidor de la plataforma GFA- \mathcal{M}^3 . Hay un único fichero diario que va creciendo a medida que se incorporan nuevas medidas cada quince minutos. El script Python *load-itacyl-stations.py* (figura 4.8) se encarga de descargar el fichero con las medidas del día y guardar los datos en el repositorio. Primero, se descarga de los servidores del ITACyL el fichero, comprimido en formato ZIP, utilizando la herramienta *wget*. Después, utilizando *unzip*, se descomprime.

Una vez descargados los datos, se guardan en el repositorio. El algoritmo 3 muestra los pasos que sigue *load-itacyl-stations.py* para almacenar las medidas de un fichero una vez descargado y descomprimido. Se recorre el fichero línea a línea (línea 1), descartando las cabeceras. En cada línea del fichero de datos vienen informadas varias variables, en concreto: Precipitación (*mm*), temperatura ($^{\circ}C$), humedad relativa (%), radiación (W/m^2), velocidad del viento viento (*m/s*) y dirección del viento ($^{\circ}$). Por cada línea se siguen los siguientes pasos: se divide

la línea en elementos, utilizando el carácter “;” como separador (línea 2); de cada línea se toma el identificador de la estación a la que pertenecen los datos, y la marca de tiempo y los valores de las variables anteriores (líneas 2–10); para guardar los datos se necesita el identificador de la estación, de modo que hay que obtenerlo en el repositorio, a partir del nombre de la estación (línea 11); por último, se guardan los datos en el repositorio utilizando el procedimiento *insert_measure* explicado en el apartado anterior (líneas 12–17).

Algoritmo 3 Procesamiento de ficheros del ITACyL (load-itacyl-stations.py).

Input: *inputfile*

```

1: for all line in inputfile do
2:   items ← split(line, ';')
3:   station ← items[0]
4:   datetime ← items[2] + item[3]
5:   rain ← items[4]
6:   t2m ← items[5]
7:   rh ← items[6]
8:   radiation ← items[7]
9:   wind_speed ← items[8]
10:  wind_direction ← items[9]
11:  place ← SELECT place_id FROM place WHERE place_name = station
12:  insert_measure(datetime, 'rain', place, rain)
13:  insert_measure(datetime, 't2m', place, t2m)
14:  insert_measure(datetime, 'rh', place, rh)
15:  insert_measure(datetime, 'radiation', place, radiation)
16:  insert_measure(datetime, 'wind-speed', place, wind_speed)
17:  insert_measure(datetime, 'wind-direction', place, wind_direction)
18: end for

```

nc2sql

El módulo *nc2sql* permite introducir datos procedentes de un fichero en formato netCDF en el repositorio. El formato de un fichero netCDF se explica en el apartado 1.1.3. Se utiliza para introducir en el repositorio las medidas de la ejecución operativa del modelo WRF que realiza el GFA en *Caléndula*, el cluster de cálculo paralelo de la FCSCl, aunque se puede usar con cualquier fichero netCDF con independencia de su origen. La figura 4.9 muestra la estructura del módulo *nc2sql*.

El algoritmo 4 muestra los pasos que sigue *nc2sql* para leer una variable meteorológica de un fichero netCDF y guardar las medidas en el repositorio. Es necesario aportar dos argumentos: El identificador de la variable (*variable*) y la ruta del fichero netCDF (*inputFile*). Primero se obtienen los valores de *variable* junto con las coordenadas (longitud, latitud y altura sobre el nivel del mar) (líneas 1–5). La marca de tiempo se obtiene de los atributos globales del fichero

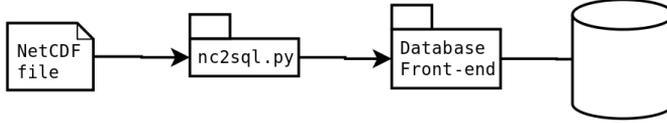


Figura 4.9: Arquitectura del módulo *nc2sql*.

(línea 6). Se calculan las dimensiones de la rejilla y se recorre punto a punto en las tres dimensiones (líneas 7–10). De cada punto de la rejilla se obtienen las coordenadas y el valor de *variable* en ese punto (líneas 11–14), y se definen un nombre, una descripción y un origen, para la localización (líneas 15–17). Por último, se guardan los datos en el repositorio utilizando los procedimientos *insert_place* y *insert_measure* (líneas 18-19).

Algoritmo 4 Procesamiento de ficheros netCDF (*nc2sql*).

```

Input: inputfile, variable                                ▷ netCDF input file path and variable name
1: ncfile ← Dataset(inputFile, 'r')
2: values ← ncfile[variable]                               ▷ Get variable values
3: LON ← ncfile['XLONG']                                   ▷ Get longitude coordinates
4: LAT ← ncfile['XLAT']                                   ▷ Get latitude coordinates
5: HGT ← ncfile['HGT']                                   ▷ Get height coordinates
6: datetime ← ncfile['START_DATE']
7: nlevs, nlats, nlons ← shape(values)                    ▷ Get values dimensions
8: for i = 1 to nlevs do
9:   for j = 1 to nlats do
10:    for k = 1 to nlons do
11:      lon ← LON[i, j, k]
12:      lat ← LAT[i, j, k]
13:      hgt ← HGT[i, j, k]
14:      value ← values[i, j, k]
15:      name ← 'WRF grid point'
16:      description ← 'lon,lat,hgt'
17:      source ← 6                                         ▷ WRF model from GFA
18:      place ← insert_place(lon, lat, hgt, name, description, source)
19:      insert_measure(datetime, variable, place, value)
20:    end for
21:  end for
22: end for
  
```

Se repite el proceso anterior por cada variable y cada fichero netCDF cuyos datos se quieran almacenar en la plataforma GFA- \mathcal{M}^3 . Los ficheros netCDF cuyos datos se introducen en el repositorio de la plataforma GFA- \mathcal{M}^3 , son los generados por la ejecución operativa del modelo WRF que realiza el GFA, explicada en el apartado 3.2.2. Estas ejecuciones se gestionan utilizando la herramienta

PyWRFScheduler [Guerrero-Higueras *et al.* (2013a)].

PyWRFScheduler La herramienta PyWRFScheduler descompone la ejecución del modelo en una serie de pasos o tareas. Cada tarea se implementa en un script Python diferente:

- *preprocess.py*. Este script se encarga de una serie de tareas previas a la ejecución del modelo.
- *geogrid.py*, se encarga de la ejecución del módulo *GEOGRID*. No es necesario ejecutar *GEOGRID* en cada ejecución del modelo, solamente cuando haya algún cambio en el área geográfica de estudio.
- *ungrib.py*, se encarga de la ejecución del módulo *UNGRIB*.
- *metgrid.py*, se encarga de la ejecución del módulo *METGRID*.
- *real.py*, se encarga de la ejecución del módulo *REAL*.
- *wrf.py*, se encarga de la ejecución del módulo *WRF*.
- *postproc.py*, obtiene una serie imágenes PNG a partir de las salidas netCDF que genera el modelo WRF.

Además de los scripts anteriores, es necesario otro más, *wfmanager.py*, que se encarga de enviar los trabajos a Caléndula en el orden adecuado y de monitorizar su correcta finalización. A diferencia de los otros scripts, que se ejecutan dentro del entorno de cálculo paralelo, *wfmanager.py* debe ejecutarse desde un nodo con capacidad de enviar trabajos al gestor de colas SGE utilizando el comando *qsub*. La figura 4.10 muestra las interrelaciones de los scripts que componen PyWRFScheduler y el gestor de colas SGE.

wfmanager.py Para monitorizar el inicio y la finalización de cada tarea, *wfmanager.py* utiliza el fichero de log que genera SGE con el resultado de la ejecución de cada trabajo. El orden en que *wfmanager.py* envía los trabajos al gestor de colas viene definido en un fichero de configuración en formato XML. Este fichero de configuración sigue una serie de reglas y contiene una serie de entidades. En concreto, un elemento de la entidad *Work-flow* que define el orden de las tareas y *n* elementos de la entidad *Task*, una por cada tarea que compone el flujo de trabajo. La figura 4.11 muestra la jerarquía entre las entidades que definen el flujo de trabajo.

Un elemento de la entidad *Work-flow* contiene una serie de elementos de la entidad *Task*. Tiene dos atributos, *date* que define la hora de inicio del pronóstico y *forecast* que define el horizonte temporal del pronóstico, en horas, a partir de la fecha de inicio.

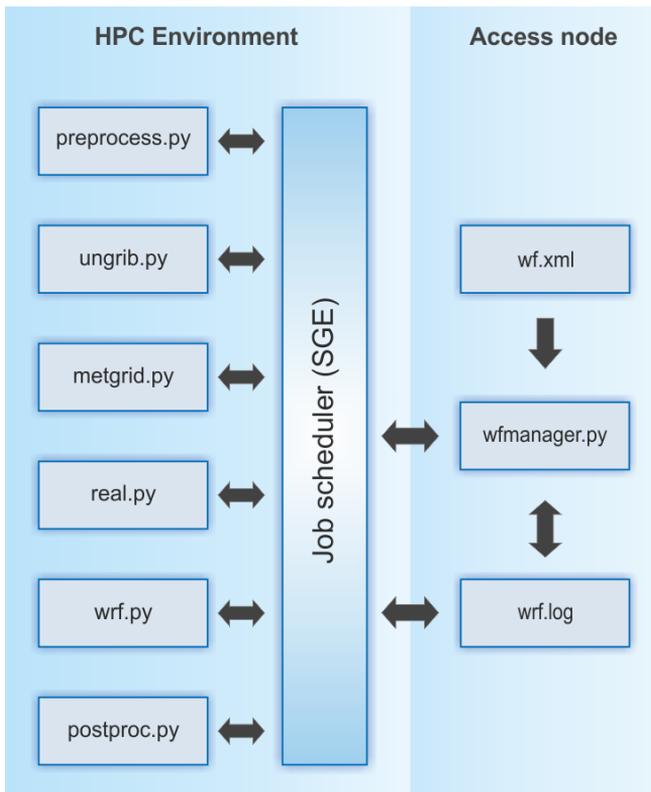


Figura 4.10: Arquitectura de PyWRFScheduler.

Cada elemento de la entidad Task define una tarea. Contiene un elemento de cada una de las siguientes entidades:

- *id*. Identifica una tarea. Tiene dos atributos: *continueTag*, que es una cadena de texto que `wfmanger.py` busca en el log de SGE para saber si la tarea ha terminado correctamente; y *abortTag*, que será el valor para saber si la tarea no ha terminado correctamente.
- *script*. Especifica el script Python que debe ejecutar la tarea. Tiene dos atributos: *nodes*, que especifica el número de nodos que se utilizarán en el entorno de cálculo paralelo para ejecutar el script; y *queue*, que identifica la cola de ejecución del entorno de cálculo paralelo.
- *paramlist*. Contiene una serie de elementos de la entidad *param* que identifican los parámetros necesarios para ejecutar la tarea.

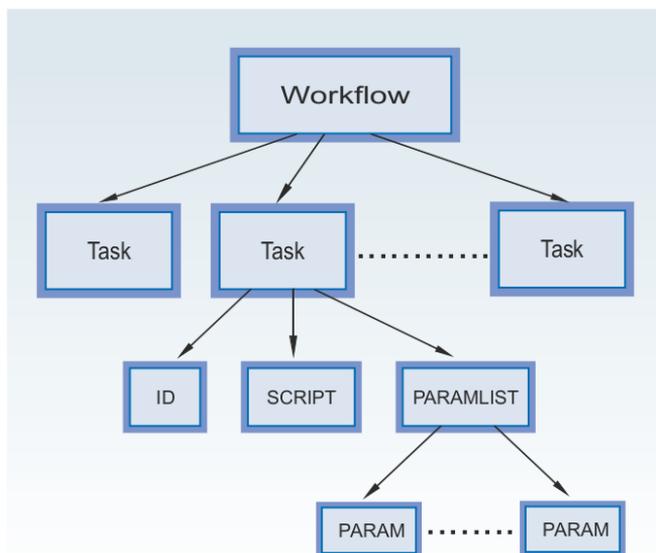


Figura 4.11: Jerarquía de las entidades que definen el flujo de trabajo.

wfmanager.py analiza el fichero de configuración XML y define el flujo de trabajo utilizando el API Document Object Model (DOM) de Python. Por cada elemento Task que aparece en el fichero, wfmanager.py envía un trabajo al gestor de colas SGE utilizando el comando qsub con una serie argumentos:

- Directorio de trabajo (-wd). Directorio base donde trabaja el script.
- Entorno paralelo (-pe). Número de nodos a reservar en la cola donde se envía el script.
- Nombre (-N). Identificador del trabajo.
- Salida estándar (-o). Ruta del fichero donde redirigir la salida estándar.
- (-j). Especifica si se mezcla o no la salida estándar de error con la salida estándar.
- Intérprete (-S). Intérprete que ejecuta el script.
- Cola (-q). Cola donde se envía el script.

wfmanager.py utiliza el valor de los atributos de la tarea y los elementos que contiene, para asignar valores a los argumentos de qsub.

El listado 4.2 muestra un fichero de configuración completo siguiendo la especificación anterior.

```

1 <?xml version="1.0" encoding="UTF-8" ?>
2 <workflow date="2011102600">
3   <task run="true">
4     <id continueTag="Preprocess finished" abortTag="Preprocess aborted">preprocess</id>
5     <script numnodos="1" queue="gfa">preprocess.py</script>
6     <paramlist>
7       <param name="initDate">2011101400</param>
8       <param name="numFiles">17</param>
9       <param name="templatePath">/home/gfa_1/gfa_1_2/templates</param>
10      <param name="geogPath">/home/gfa_1/gfa_1_2/input</param>
11      <param name="outputPath">/home/gfa_1/gfa_1_2</param>
12    </paramlist>
13  </task>
14  <task>
15    <id continueTag="ungrrib finished" abortTag="ungrrib aborted">ungrrib</id>
16    <script numnodos="1" queue="gfa">ungrrib.py</script>
17    <paramlist>
18      <param name="ungrribPath">/home/gfa_1/gfa_1_2</param>
19      <param name="inputPath">/home/gfa_1/gfa_1_2/scripts/test</param>
20      <param name="logPath">/home/gfa_1/gfa_1_2/scripts/test</param>
21    </paramlist>
22  </task>
23  <task>
24    <id continueTag="metgrid finished" abortTag="metgrid aborted">metgrid</id>
25    <script numnodos="1" queue="gfa">metgrid.py</script>
26    <paramlist>
27      <param name="metgridPath">/home/gfa_1/gfa_1_2</param>
28      <param name="logPath">/home/gfa_1/gfa_1_2/scripts/test</param>
29    </paramlist>
30  </task>
31  <task>
32    <id continueTag="real finished" abortTag="real aborted">real</id>
33    <script numnodos="16" queue="gfa">real.py</script>
34    <paramlist>
35      <param name="realPath">/home/gfa_1/gfa_1_2</param>
36      <param name="mpiPath">/home/gfa_1/gfa_1_2</param>
37    </paramlist>
38  </task>
39  <task>
40    <id continueTag="wrf finished" abortTag="wrf aborted">wrf</id>
41    <script numnodos="128" queue="gfa">wrf.py</script>
42    <paramlist>
43      <param name="wrfPath">/home/gfa_1/gfa_1_2</param>
44      <param name="mpiPath">/home/gfa_1/gfa_1_2</param>
45    </paramlist>
46  </task>
47 </workflow>

```

Listado 4.2: Fichero de configuración wf.xml para wfmanager.py.

preprocess.py Este script realiza una serie de tareas previas a la ejecución del modelo. Primero, se descarga los datos necesarios para inicializar el modelo (condiciones iniciales). En las ejecuciones operativas del GFA se utilizan las salidas del modelo GFS para inicializar el modelo. Después, se modifican algu-

nos valores en los ficheros de configuración propios del modelo, en concreto la fecha/hora de inicio y final de las predicciones.

ungrib.py ungrib.py se encarga de la ejecución el módulo UNGRIB, el algoritmo 5 muestra los pasos que sigue su implementación. Primero, hay que enlazar los datos de inicialización del modelo para que UNGRIB pueda utilizarla. Para ello, el modelo WRF incluye un script que se encarga de hacerlo, *link_grib.csh*. Después de enlazar los datos de entrada con *link_grib.csh* (línea 2), se ejecuta *ungrib.exe*, incluido en el software del modelo (línea 3). Si ungrib.exe se ejecuta correctamente, ungrib.py escribe el valor del atributo `continueTag` en el fichero de log, en otro caso escribe el valor del atributo `abortTag` (líneas 4–8).

Algoritmo 5 ungrip.py (PyWRFScheduler).

```

Input: ungribPath, inputPath           ▷ ungrip.exe and link_grib.csh path, input data path
1: chdir(ungribPath)                       ▷ Move to ungrip.exe path
2: run 'link_grib.csh inputPath/gfs*'       ▷ Link input data
3: run 'ungrib.exe'
4: if Success running of ungrib.exe then
5:   WriteLog('ungrib finished')
6: else
7:   WriteLog('ungrib aborted')
8: end if

```

metgrid.py Este script se encarga de la ejecución del módulo METGRID. Para hacerlo, hay que ejecutar el programa *metgrid.exe* incluido en el software del modelo. Si metgrid.exe se ejecuta correctamente, metgrid.py escribe el valor del atributo `continueTag` en el fichero de log, en otro caso escribe el valor del atributo `abortTag`.

real.py real.py se encarga de la ejecución del módulo REAL siguiendo los pasos del algoritmo 6. Primero, hay que ejecutar el programa *real.exe*, incluido en el software del modelo. real.exe está diseñado para ejecutarse en varios nodos con el comando *mpirun* (línea 1). Por hacerlo, hay que indicar el número de nodos como argumento de mpirun (*numNodos*). Si real.exe se ejecuta correctamente, real.py escribe el valor del atributo `continueTag` en el fichero de log, en otro caso escribe el valor del atributo `abortTag` (líneas 11–19).

wrf.py Este script se encarga de la ejecución del módulo WRF. Para hacerlo, hay que ejecutar el programa *wrf.exe*, incluido en el software del modelo. wrf.exe está diseñado para ejecutarse en varios nodos con el comando mpirun. Si

Algoritmo 6 *real.py* (PyWRFScheduler).

```

Input: realPath, numNodes                                ▷ real.exe path, number of nodes for mpirun
1: chdir(realPath)                                          ▷ Move to real.exe path
2: run 'mpirun -n numNodes real.exe'
3: if Success running of real.exe then
4:   WriteLog('real finished')
5: else
6:   WriteLog('real aborted')
7: end if

```

wrf.exe se ejecuta correctamente, *wrf.py* escribe el valor del atributo *continueTag* en el fichero de log, en otro caso escribe el valor del atributo *abortTag*.

http2sql

El módulo *http2sql* permite introducir datos meteorológicos vía HTTP en el repositorio. Es utilizado principalmente por los observadores voluntarios. La figura 4.12 muestra la arquitectura del módulo. Para obtener las credenciales que permiten su utilización, los observadores deben ponerse en contacto con el GFA a través de su sitio web.

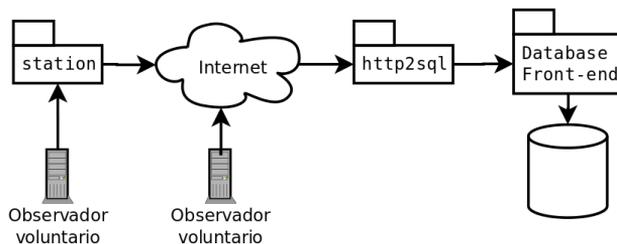


Figura 4.12: Arquitectura de *http2sql*.

http2sql se ha implementado como servicio web². Por el momento, el servicio pone a disposición de los usuarios autorizados dos métodos: *insertPlace* e *insertMeasure*. La funcionalidad de estos dos métodos se corresponde con la los de los procedimientos almacenados *insert_place* e *insert_measure*, explicados en el apartado 4.1.1. Los argumentos en ambos métodos se corresponden con los de los procedimientos almacenados, con una diferencia, *insertPlace* e *insertMeasure* requieren dos argumentos adicionales, un nombre de usuario y una contraseña, que previamente hay que solicitar al GFA para poder utilizar el servicio.

²En <http://hawking.unileon.es/rdocyl/wsd1.php> puede consultarse la especificación Web Services Description Language (WSDL) del servicio.

station Para que un observador voluntario no tenga que implementar una solución software que permita invocar al servicio web definido en el punto anterior, el GFA ha desarrollado la aplicación *Station*, que los usuarios pueden descargar e instalar en su computadora para compartir sus datos de observación. *Station* está implementado en Python y por tanto, puede ser ejecutado en cualquier sistema operativo.

La aplicación *Station* necesita que el usuario edite un fichero de configuración (*station.cfg*) y complete algunos argumentos: *id*: Identificador de la estación asignado por el GFA; *type*: Tipo de estación meteorológica; *file*: Ruta del fichero con los datos de la estación; y *time_step*: Intervalo de tiempo (en segundos) entre cada envío. El listado 4.3 muestra parte de este fichero de configuración.

```
1 id=obv_000_IMA
2 type=weatherlink
3 file=C:\WaeatherLink\Estacion\download.txt
4 time_step=900 # 15 minutos
```

Listado 4.3: Fichero de configuración de *station*.

El segundo argumento (*type*) es especialmente importante, ya que en función del tipo de estación meteorológica de la que disponga el observador, el formato en el que se guardan los datos será distinto. *Station* está preparado para trabajar con estaciones de WeatherLink y the WeatherDisplay, los dos fabricantes de estaciones meteorológicas más populares del mercado. Acepta datos de otras estaciones, pero si el fabricante no sigue ningún estándar a la hora de guardar los datos (CSV, etc.), puede haber alguna incompatibilidad.

hrit2sql

El módulo *hrit2sql* permite introducir datos en formato HRIT en el repositorio. La figura 4.13 muestra la arquitectura del módulo *hrit2sql*. Los ficheros HRIT obtenidos a partir de un receptor de datos MSG se convierten a formato netCDF. Para ello, se utiliza la herramienta *SatDatos*, contenida dentro del paquete *Nubes4* desarrollado por *PraPro* [PraPro (2009)]. Después, el fichero netCDF es procesado por el módulo *nc2sql* explicado en el apartado anterior.

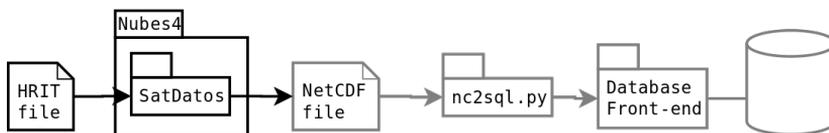


Figura 4.13: Arquitectura de *hrit2sql*.

4.1.3. Output services

Los datos meteorológicos que los usuarios pueden obtener, dependen directamente de los servicios de consulta de los que dispone el repositorio. Por el momento, los usuarios pueden obtener información en formato CSV, netCDF, o bien obtener directamente visualizaciones gráficas en formato PNG o KML. En el apartado 6.3 se plantea la posibilidad de incluir nuevos servicios de consulta para obtener información en otros formatos.

Cualquier usuario registrado en el sitio web del GFA puede hacer uso de estos servicios de consulta para obtener datos. La figura 4.14 muestra el formulario web a partir del cual se pueden hacer las consultas.

Obtener un mapa personalizado

Completa la información del siguiente formulario para obtener un mapa personalizado con los datos de la red de observación meteorológica de Castilla y León.

Fecha: 23 ▼ oct ▼ 2013 ▼ Fecha de los datos que se desea obtener

Hora: 12 ▼ 00 ▼ Hora de los datos que se desea obtener

Variable meteorológica: Temperatura ▼

Región:

- Castilla y León
- Cuencas de los ríos Duero y Ebro
- Personalizada: Lon. máxima Lon. mínima Lat. máxima Lat. mínima

Resolución:

- 200 x 200 pixels
- 400 x 400 pixels
- Personalizada: x pixels

Formato de la salida:

- Imagen (PNG)
- Google Maps/Earth (KML)
- Text/csv
- NetCDF

Obtener mapa

Figura 4.14: Formulario de consulta del sitio web del GFA.

Es posible obtener datos directamente a utilizando el servicio web *create-map*. Para ello es necesario aportar una serie de argumentos, que se corresponden con los campos que hay que rellenar en el formulario de consulta. La tabla 4.6 muestra una lista con todos los argumentos necesarios y sus posibles valores. Los que aparecen marcados en negrita son obligatorios. A continuación se muestran algunas invocaciones válidas del servicio *create-map*:

- Obtención de una gráfica en formato KML con datos de temperatura en superficie correspondientes al 3 de octubre de 2013 a las 16:00 UTC:

```
1 http://hawking.unileon.es/station/create-map-2.php?type=temperature&date=201310031600&out=kml
```

- Obtención de una gráfica en formato PNG con datos de precipitación correspondientes al 3 de octubre de 2013 a las 16:00 UTC:

```
1 http://hawking.unileon.es/station/create-map-2.php?type=rain&date=201310031600&out=png
```

Tabla 4.6: Argumentos de create-map. Los argumentos marcados en negrita son obligatorios.

Argumento	Descripción
type	Variable meteorológica.
date	Fecha y hora de los datos.
out	Tipo de salida: <ul style="list-style-type: none"> – <i>png</i> para PNG – <i>kml</i> para KML – <i>csv</i> para CSV – <i>nc</i> para netCDF
region	Región. Valores posibles: <ul style="list-style-type: none"> – <i>cyl</i> para Castilla y León. – <i>nw</i> para las riberas del Duero y del Ebro. – <i>custom</i> para personalizar la región.
maxlat	Latitud máxima para la región personalizada.
maxlon	Longitud máxima para la región personalizada.
minlat	Latitud mínima para la región personalizada.
minlon	Longitud mínima para la región personalizada.
resolution	Resolución de la rejilla (<i>grid</i>). Valores posibles: <ul style="list-style-type: none"> – <i>200x200</i> – <i>400x400</i> – <i>NxM</i>

Los servicios de consulta que atienden estas peticiones son: *sql2csv*, *sql2nc*, *csv2png* y *csv2kml*. Todos ellos fueron introducidos en el apartado 2.3.3. La figura 4.15 muestra sus interrelaciones. En las siguientes secciones se profundiza en cada uno de ellos.

sql2csv

Este módulo permite la extracción de datos del repositorio en formato CSV. Este formato es útil para los investigadores, ya que pueden tratar los datos con

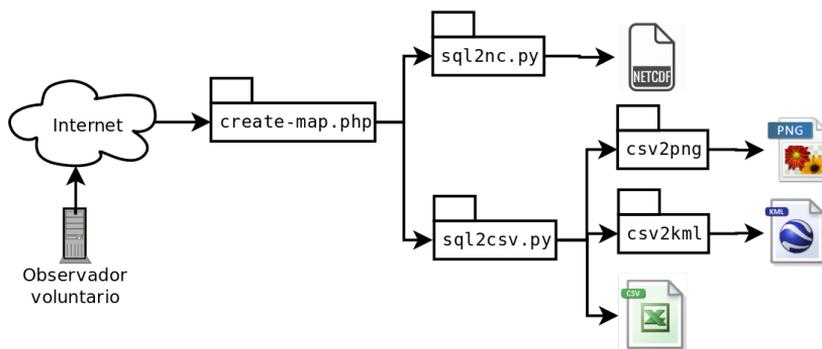


Figura 4.15: Módulos de los servicios consulta y sus interrelaciones.

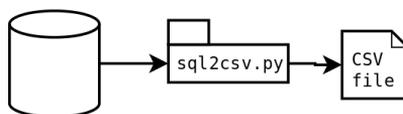


Figura 4.16: Arquitectura del módulo sql2csv.

diversas herramientas y realizar sus propios cálculos, comparativas, etc. La figura 4.16 muestra la arquitectura del módulo.

sql2csv.py es un script Python que se invoca cuando el usuario escoge la salida en formato CSV en el formulario de consulta. Recibe dos argumentos: Una cadena de texto con la consulta Structured Query Language (SQL) que se quiere realizar y la ruta del fichero CSV que se quiere crear. Primero se ejecuta la consulta y se obtienen los resultados en forma de matriz. Las filas representan las medidas y las columnas representan los campos (coordenadas, marca de tiempo, variables, valor, etc.). La matriz se procesa fila a fila para construir el fichero CSV. Una fila de la matriz, se corresponde con una fila en el fichero CSV. El listado 4.4 muestra un extracto de un fichero CSV generado por *sql2csv*. Incluye los siguientes campos: Identificador del ejemplar de Place, longitud, latitud, altura sobre el nivel del mar, origen de los datos (identificador de Ssource), fecha y hora de la medida, valor numérico de la medida, variable meteorológica y unidades en que se mide.

```

1 AF581;-6.799314;41.034013;150.00;1;2015-04-26 00:00:00;0.15;Precipitación;mm
2 PL602;-6.794821;40.265631;960.00;1;2015-04-26 00:00:00;0.6;Precipitación;mm
3 PL292;-6.754223;42.122004;1020.00;1;2015-04-26 00:00:00;0.3;Precipitación;mm
4 PL293;-6.734307;42.136882;1370.00;1;2015-04-26 00:00:00;0;Precipitación;mm
5 AF601;-6.674370;40.364087;780.00;1;2015-04-26 00:00:00;0;Precipitación;mm
6 AF604;-6.659824;40.650755;590.00;1;2015-04-26 00:00:00;0.45;Precipitación;mm
7 AF291;-6.630441;42.053328;900.00;1;2015-04-26 00:00:00;0.45;Precipitación;mm
8 PL601;-6.585792;40.291579;950.00;1;2015-04-26 00:00:00;1.4;Precipitación;mm

```

9 ...

Listado 4.4: Extracto de un fichero CSV generado por *sql2csv*.

sql2nc

sql2nc permite la extracción de datos del repositorio en formato netCDF. La figura 4.17 muestra la arquitectura del módulo. Como se observa en la figura, *sql2nc* hace uso de *sql2csv* para generar un fichero CSV. Después, el script Python *csv2nc.py* utiliza el fichero CSV generado por *sql2csv* para generar el fichero netCDF final. Para hacerlo, *csv2nc.py* se sirve del script NCAR Command Language (NCL) *csv2nc.ncl*, que inserta valores de una variable en un fichero netCDF, creándolo si este no existe. Para almacenar valores de varias variables *csv2nc.py* invoca sucesivamente a *csv2nc.ncl*.

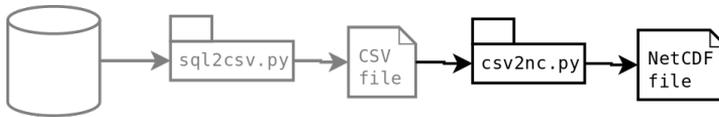


Figura 4.17: Arquitectura de *sql2nc*.

El algoritmo 7 muestra los pasos que sigue *csv2nc.ncl* para guardar los valores de una variable en un fichero netCDF. Es necesario aportar algunos argumentos: la ruta del fichero CSV con los datos de entrada (*inputFile*), la ruta del fichero netCDF donde se guardarán los datos (*outputFile*) y el nombre de la variable que se quiere añadir (*variable*). Primero, tras leer el contenido del fichero CSV (línea 1), se extraen las medidas del mismo (líneas 2–4), en concreto, las coordenadas y con el valor de la variable asociado. A continuación, se construye una malla (*grid*) utilizando las coordenadas y los valores (línea 5). Se rellenan los posibles valores incompletos (línea 6). Por último, se crea un fichero netCDF nuevo (línea 7) y se añade al mismo la nueva variable asignándole la malla definida en los pasos anteriores (línea 8).

csv2png

Este módulo permite la obtención de visualizaciones gráficas en formato PNG a partir de las salidas del módulo *sql2csv*. La figura 4.18 muestra la arquitectura del módulo. *csv2png.py* ejecuta un script NCL al que se pasa como argumento la salida que previamente produce *sql2csv*.

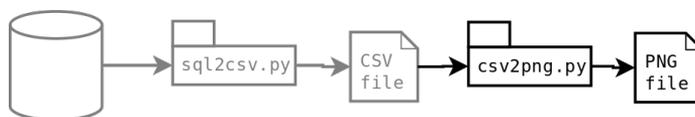
Hay un script NCL diferente para cada una de las variables que se pueden consultar, en concreto, en el momento de la redacción de este trabajo: temperatura, temperatura máxima, temperatura mínima, precipitación, precipitación

Algoritmo 7 csv2nc.ncl (csv2nc).**Input:** *inputFile*, *outputFile*, *variable*

```

1: lines ← read(inputFile)           ▷ Read file contents
2: lon ← get_field(lines, 1)         ▷ Get longitude
3: lat ← get_field(lines, 1)         ▷ Get latitude
4: values ← get_field(lines, 1)      ▷ Get values
5: grid ← build_grid(lon, lat, values)
6: fill_missing_values(grid)
7: output ← open(outputFile, 'w')
8: write(output, grid)

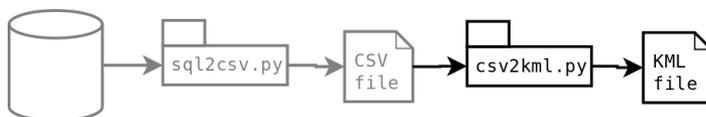
```

**Figura 4.18:** Arquitectura del módulo csv2png.

acumulada y presencia de nieve. Se ejecuta uno u otro en función del valor del argumento *type* recibido en la invocación al servicio *create-map.php*.

csv2kml

Este módulo permite la obtención de ficheros en formato KML a partir de las salidas del módulo *sql2csv*. Hay diversas herramientas que permiten la visualización de estos ficheros como *Google Earth*, *ArcGIS*, etc. La figura 4.19 muestra la arquitectura de *csv2kml*.

**Figura 4.19:** Arquitectura de csv2kml.

csv2kml.py genera un fichero en formato KML utilizando los datos del fichero CSV producido por *sql2csv*. Se añade un marcador, como el que muestra el listado 4.5, por cada medida.

```

1 <Placemark>
2   <name>AF012</name>
3   <styleUrl>#chd</styleUrl>
4   <description>
5     <![CDATA[
6     <table>

```

```

7     <tbody>
8     <tr><td>&nbsp;</td><td>Origen</td><td>Confederación Hidrográfica del Duero</td>
9     </tr><td>&nbsp;</td><td>Identificador de estación</td><td>AF012</td></tr>
10    <tr><td>&nbsp;</td><td>Fecha/Hora</td><td>2015-02-14 22:00:00</td></tr>
11    <tr><td>&nbsp;</td><td>Longitud</td><td>-6.819436</td></tr>
12    <tr><td>&nbsp;</td><td>Latitud</td><td>41.032176</td></tr>
13    <tr><td>&nbsp;</td><td>Altura</td><td>133.00 m</td></tr>
14    <tr><td></td><td>Temperatura media</td><td>5.33 C</td></tr>
16    </tbody>
17    </table>
18    ]])
19    </description>
20    <Point>
21    <coordinates>-6.819436,41.032176,0</coordinates>
22    </Point>
    </Placemark>

```

Listado 4.5: Función *putMarker* en el script *csv2kml.py*.

4.2. Infraestructura hardware

Para poder implementar la solución propuesta en el apartado anterior es necesario cubrir una serie de requisitos hardware. El primero de ellos es disponer de un servidor capaz de contener el repositorio asegurando cierto nivel de seguridad. Por ello se ha optado por una plataforma GNU/Linux por sus conocidas prestaciones, tanto a nivel de rendimiento como de seguridad.

Por otro lado, el repositorio debe ser accesible a todo el mundo, por tanto debe contar con una salida permanente a internet y una dirección IP fija. Esto último permitirá a los usuarios finales acceder al servidor. El acceso desde fuera es imprescindible tanto para poder introducir información en el repositorio como para obtenerla. Por tanto, es necesario abrir una serie de puertos de comunicaciones que permitan el tráfico de ciertos servicios como HTTP, Secure SHell (SSH), FTP, etc. Es necesario tomar las medidas necesarias para asegurar las comunicaciones (firewall, Intrusion Detection System (IDS), etc.).

La plataforma desarrollada por el GFA está funcionando de manera operativa en un servidor GNU/Linux, denominado *hawking*, dentro de las instalaciones del grupo en el IMARENABIO. Las tablas 4.7, 4.8 y 4.9 resumen las prestaciones del servidor.

En relación a las herramientas instaladas la tabla 4.10 muestra una lista con las principales. En el apartado anterior, donde se habla de la infraestructura software, se expone más detalladamente las necesidades y motivaciones de algunas de ellas.

Tabla 4.7: Información general del servidor *hawking*.

OS Type	Linux
Release	Ubuntu 12.04 (precise)
Hostname	hawking
Kernel	3.2.0-26-generic #41-Ubuntu SMP Thu Jun 14 17:49:24 UTC 2012
GCC version	4.6 (x86_64-linux-gnu)

Tabla 4.8: Información sobre la CPU del servidor *hawking*.

Fabricante	GenuineIntel
CPUs	4
Modelo	Intel(R) Core(TM)2 Quad CPU Q6600 @ 2.40GHz
Frecuencia	2394.154 MHz
Cache L2	4096 KB
Memoria L2	7920 MiB

Tabla 4.9: Información sobre la memoria y capacidad de almacenamiento del servidor *hawking*.

Memoria	7920 MB
Almacenamiento	458 GB

Tabla 4.10: Herramientas instaladas en el servidor *hawking*.

Herramienta	Versión
PostgreSQL	psql (PostgreSQL) 9.1.13
Servidor web Apache	Apache/2.2.22 (Ubuntu)
PHP	PHP 5.3.10-1ubuntu3.23 with Suhosin-Patch (cli)

Capítulo 5

Aplicaciones

Este capítulo presenta algunas de las investigaciones, aplicaciones y proyectos en los que ha participado el GFA, en todos los casos se ha utilizado la información del SGDM descrito en el capítulo 4. Con ello se pretende dar validez a las hipótesis 1, 2, 3 y 4, planteadas en el capítulo 1. Cada caso se presenta en el ámbito de su marco de aplicación entre los siguientes:

1. Identificación y seguimiento de riesgos meteorológicos en tiempo real.
2. Asimilación de datos en modelos NWP.
3. Calibración y validación de algoritmos.
4. Desarrollo de aplicaciones.

En las siguientes secciones se profundiza en cada marco de aplicación.

5.1. Identificación y seguimiento de riesgos meteorológicos en tiempo real

La identificación, y su posterior monitorización, de un determinado evento meteorológico que pueda suponer riesgo para la sociedad es vital de cara a minimizar su impacto. En esa línea, un SGDM con información en tiempo real puede ser de gran utilidad. En concreto, el GFA- \mathcal{M}^3 , que dispone de información en tiempo real de diversas fuentes, como se expuso en el capítulo 3, permite el seguimiento de este tipo de eventos nocivos dentro del área geográfica que abarca.

Una de las principales ventajas del GFA- \mathcal{M}^3 es la posibilidad de obtener, interpolados, datos de diversas fuentes. Los resultados obtenidos fruto de esta

interpolación, son más precisos, y cubren un área geográfica mayor, que obtenidos trabajando con estas mismas fuentes de manera independiente.

A continuación se exponen algunos ejemplos de seguimiento de riesgos meteorológicos en un caso real, utilizando la plataforma GFA- \mathcal{M}^3 . En concreto, los siguientes:

- Heladas.
- Precipitación.
- Nevadas.

Entre los días 30 de enero y 6 de febrero de 2015 se sucedieron en todo el norte de la Península Ibérica numerosos episodios de precipitación severa. Algunos de estos episodios de precipitación, en especial aquellos que tuvieron lugar en las inmediaciones de la Cordillera Cantábrica en su sección más occidental, unidos a las bajas temperaturas derivaron en precipitación en forma de nieve, llegando incluso a alcanzar los 20 centímetros en lugares donde tales fenómenos no ocurren habitualmente, cómo por ejemplo, la ciudad de León. Los datos de la plataforma GFA- \mathcal{M}^3 fueron de gran utilidad para el seguimiento de los eventos meteorológicos antes mencionados.

5.1.1. Seguimiento de heladas

Las heladas son un fenómeno meteorológico muy común que tiene muchas implicaciones para la sociedad, entre otras, pavimentos deslizantes, congelación de conducciones de agua, etc. Una helada precedida de precipitación en forma de nieve favorece que esta se mantenga y no desaparezca.

Se produce una helada cuando la temperatura se mantiene a 0 °C o por debajo. Conocer los valores de temperatura en tiempo real permite localizar las zonas donde se está produciendo una helada y tomar las medidas que se precisen en cada caso.

La figura 5.1 muestra la evolución de la temperatura, en las cuencas de los ríos Duero y Ebro, cada dos horas a lo largo del 31 de enero de 2015. En ella, se observa como las temperaturas se mantuvieron muy bajas incluso en las horas diurnas, lo que permitió que toda la nieve caída hasta el momento permaneciera en la superficie. Las gráficas de temperatura se han extraído de la plataforma GFA- \mathcal{M}^3 , interpolando datos de las estaciones meteorológicas del SAIH Duero, del SAIH Ebro, del acITACyL y de observadores voluntarios.

5.1.2. Seguimiento de precipitaciones

La precipitación también puede ser un riesgo para la sociedad. Las precipitaciones severas pueden provocar riadas, arruinar cosechas, desbordar ríos, etc.

5.1. IDENTIFICACIÓN Y SEGUIMIENTO DE RIESGOS METEOROLÓGICOS EN TIEMPO REAL

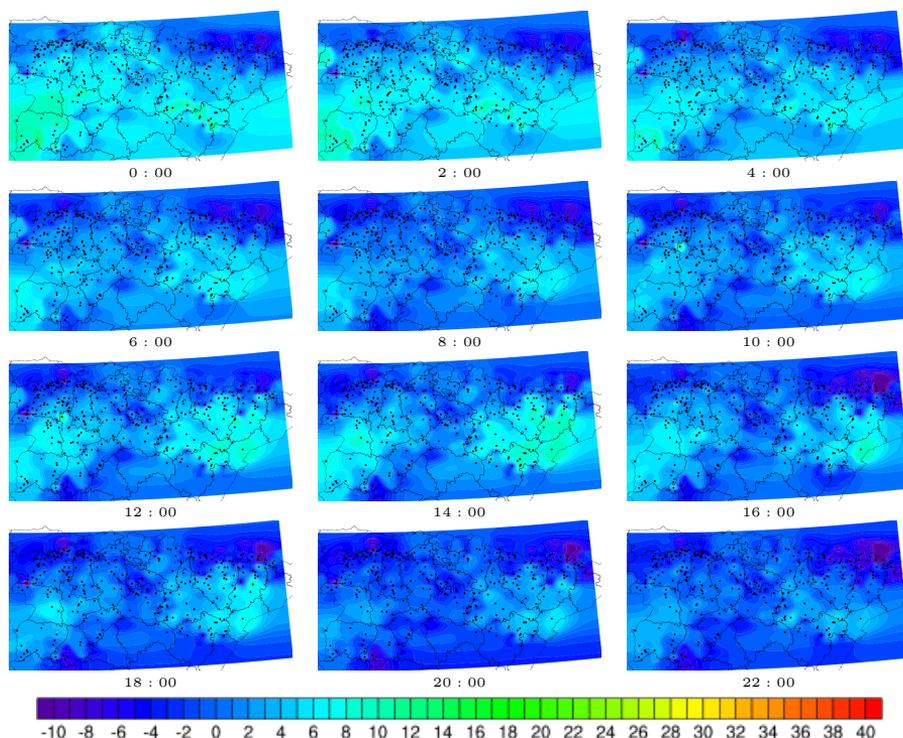


Figura 5.1: Evolución de la temperatura ($^{\circ}\text{C}$), el 31 de enero de 2015, en las cuencas de los ríos Duero y Ebro.

Saber cuanta precipitación está cayendo en una determinada localización favorece que se puedan tomar las medidas necesarias para paliar sus efectos.

La figura 5.2 muestra la evolución de la precipitación cada hora a lo largo del 31 de enero de 2015, en las cuencas de los ríos Duero y Ebro. En ella se observa como las precipitaciones, aunque no muy fuertes, fueron constantes en la mitad norte de la Península Ibérica. Las gráficas de precipitación se han extraído de la plataforma GFA- \mathcal{M}^3 , interpolando datos de las estaciones meteorológicas del SAIH Duero, del SAIH Ebro, del acITACyL y de observadores voluntarios.

5.1.3. Seguimiento de nevadas

La precipitación severa en forma de nieve supone un gran riesgo, sobre todo en lo que a comunicaciones por carretera se refiere. Adelantarse a ellas, además de evitar poner en peligro la vida de personas, supone un ahorro económico muy

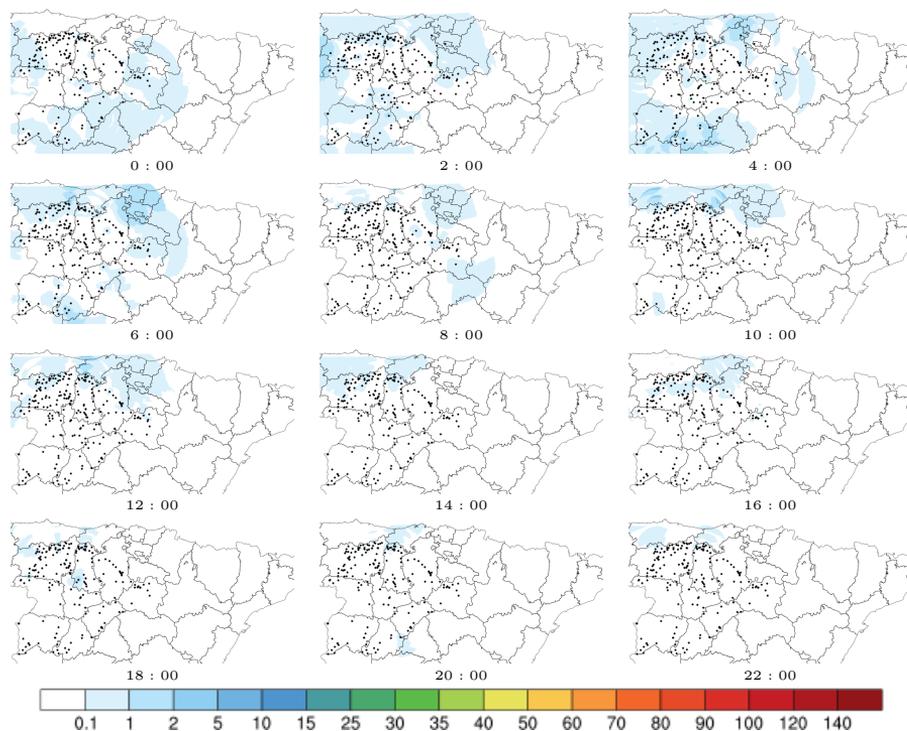


Figura 5.2: Evolución de la precipitación (mm), el 31 de enero de 2015, en las cuencas de los ríos Duero y Ebro.

importante.

Haciendo un seguimiento conjunto de la precipitación y de la temperatura, podemos detectar las localizaciones donde se está produciendo una nevada. La figura 5.3 muestra, sombreadas en azul, las ubicaciones donde se produjeron nevadas durante el 31 de enero de 2015. Para detectar estas zonas basta con identificar los puntos donde hay presencia de precipitación (figura 5.2) y donde las temperaturas están por debajo de 0°C (figura 5.1). De nuevo, los datos para generar las gráficas se han extraído de la plataforma GFA- \mathcal{M}^3 interpolando valores de precipitación y temperatura.

5.2. Asimilación de datos en modelos NWP

Otra de las aplicaciones de un SGDM es la obtención de la información necesaria para inicializar un sistema de asimilación de datos que mejore los pronósticos

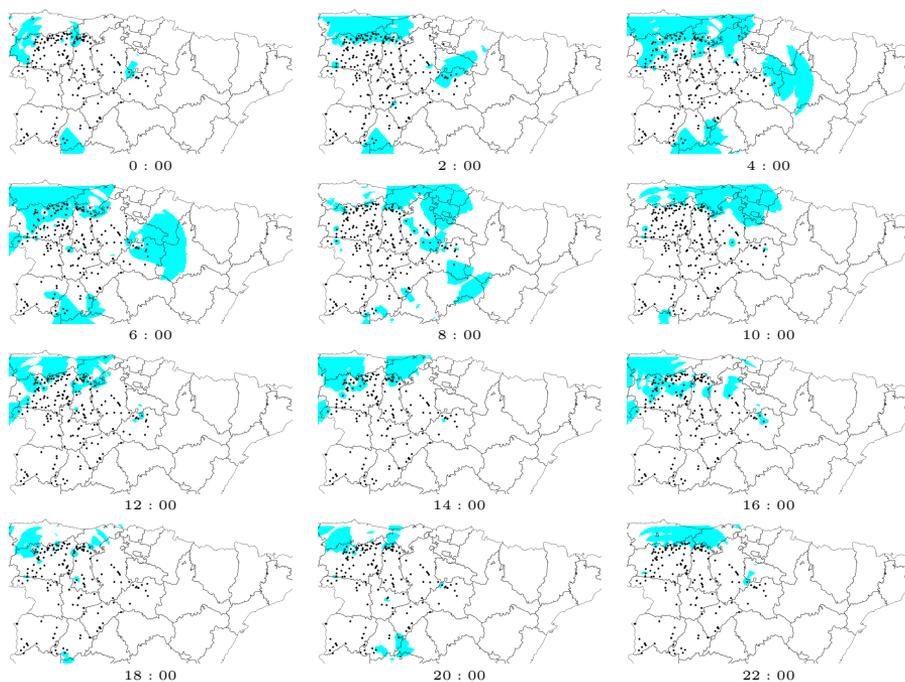


Figura 5.3: Precipitación en forma de nieve, el 31 de enero de 2015, en las cuencas de los ríos Duero y Ebro.

de un modelo NWP. En concreto la plataforma GFA- \mathcal{M}^3 , permite obtener la información necesaria para inicializar Weather Research and Forecasting model data assimilation system (WRFDA). WRFDA ha adoptado LITTLE_R como formato intermedio para introducir las observaciones en su sistema, de modo que primero es necesario convertir las observaciones de la plataforma GFA- \mathcal{M}^3 , que pueden ser extraídas en netCDF o CSV. LITTLE_R es un formato basado en texto heredado de MM5. Es responsabilidad del usuario desarrollar una interfaz para convertir los datos de observación a LITTLE_R.

Una vez se dispone de los datos de observación en el formato adecuado, la aplicación OBSPROC, incluida en el paquete WRFDA, se encarga de preprocesarlos. Este preprocesado es necesario para:

- Eliminar observaciones fuera de los dominios temporales y espaciales.
- Re-ordenar y mezclar datos duplicados.
- Transformar los datos basados en niveles de presión o altitud considerando la hipótesis hidrostática.

- Revisar la consistencia vertical y las condiciones superadiabáticas en los datos de observación con múltiples niveles.
- Asignar un grado de error a las observaciones.
- Convertir los datos de observación al formato requerido por WRFDA.

5.2.1. LITTLE_R

LITTLE_R es un formato de fichero para almacenar datos de observación basado en caracteres American Standard Code for Information Interchange (ASCII). No existe un formato estándar para datos de observación, LITTLE_R como formato intermedio para que otras aplicaciones, como WRFDA, fueran capaces de asimilar datos de observación con diferentes formatos.

Un fichero LITTLE_R se compone de informes (*reports*). Cada report se compone de un encabezado con meta-información sobre los datos, los datos de observación propiamente dichos, y una última línea con información sobre el número de observaciones que contiene el report.

El encabezado es una única línea que contiene información sobre el tipo de observación y su localización. La tabla 5.1 muestra el formato de todos los campos que componen el encabezado. Todos los campos son de ancho fijo y no se utiliza ningún carácter para separarlos. Solamente los campos que están marcados en negrita son obligatorios, el resto son opcionales. Los campos opcionales que no se completan deben rellenarse con caracteres en blanco para campos de texto o el valor '-888888.00000' para campos numéricos. El campo *platform* representa el origen de los datos de observación. La WMO ha definido una serie de códigos alfanuméricos que identifican los distintos tipos de datos de observación [WMO (2004)]. La tabla 5.2 muestra algunos de ellos.

Después del encabezado hay una o más líneas con los datos propiamente dichos. La tabla 5.3 muestra el formato de estos registros de datos. Todos los campos son de ancho fijo y no se utiliza ningún carácter para separarlos. No hay ningún campo obligatorio, todos son opcionales y dependen del tipo de observación. En aquellos campos para los que no haya valor se introduce el valor '-888888.00000'. Después de los datos hay que introducir una línea más que indica el final de los datos. El formato de esta línea es el mismo que para los registros de datos, pero debe contener el valor '-777777.00000' en los dos primeros campos y '-888888.00000' en el resto.

En algunos campos tanto del encabezado como de los registros de datos se hace referencia a unos flags de control de calidad (QC flags). Estos flags pueden tomar una serie de valores definidos en Bailey *et al.* (1994) y resumidos en la tabla 5.4.

Por último, para finalizar el report, se incluye otra línea con tres valores enteros (de 7 posiciones) con el valor del número total de datos válidos, con el

Tabla 5.1: Formato del encabezado de un report. Los campos marcados en negrita son obligatorios.

Variable	Format	Description
latitude	F20.5	station latitude (north positive)
longitude	F20.5	station longitude (east positive)
id	A40	ID of station
name	A40	Name of station
platform	A40	Description of the measurement device
source	A40	GTS, NCAR/ADP, BOGUS, etc.
elevation	F20.5	station elevation (m)
num_vld fld	I10	Number of valid fields in the report
num_error	I10	Number of errors encountered during the decoding
num_warning	I10	Number of warnings encountered during decoding
seq_num	I10	Sequence number of this observation
num_dups	I10	Number of duplicates found for this observation
is_sound	L10	T/F Above-surface or surface
bogus	L10	T/F bogus report or normal one
discard	L10	T/F Duplicate and discarded (or merged) report.
sut	I10	Seconds since 0000 UTC 1 January 1970
julian	I10	Day of the year
date_char	A20	YYYYMMDDHHmmss
slp, qc	F13.5, I7	Sea-level pressure (Pa) and a QC flag
ref_pres, qc	F13.5, I7	Reference pressure level (Pa) and a QC flag
ground_t, qc	F13.5, I7	Ground Temperature (T) and QC flag
sst, qc	F13.5, I7	Sea-Surface Temperature (K) and QC flag
psfc, qc	F13.5, I7	Surface pressure (Pa) and QC flag
precip, qc	F13.5, I7	Precipitation Accumulation and QC flag
t_max, qc	F13.5, I7	Daily maximum T (K) and QC flag
t_min, qc	F13.5, I7	Daily minimum T (K) and QC flag
t_min_night, qc	F13.5, I7	Overnight minimum T (K) and QC flag
p_tend03, qc	F13.5, I7	3-hour pressure change (Pa) and QC flag
p_tend24, qc	F13.5, I7	24-hour pressure change (Pa) and QC flag
cloud_cvr, qc	F13.5, I7	Total cloud cover (oktas) and QC flag
ceiling, qc	F13.5, I7	Height (m) of cloud base and QC flag

Tabla 5.2: Códigos WMO para identificar tipos de observación.

Código	Tipo de observación	Descripción
FM-12	SYNOP	Surface observation from a fixed land station
FM-14	SYNOP MOVIL	Surface observation from a mobile land station
FM-13	SHIP	Surface observation from a sea station
FM-86	SATEM	Report of satellite remote upper-air soundings of pressure, temperature, humidity, and thickness

Tabla 5.3: Formato de los registros de datos.

Variable	Format	Description
pressure, qc	F13.5, I7	Pressure (Pa) of observation, and QC flag
height, qc	F13.5, I7	Height (m MSL) of observation, and QC flag
temperature, qc	F13.5, I7	Temperature (K) and QC flag
dew_point, qc	F13.5, I7	Dewpoint (K) and QC flag
speed, qc	F13.5, I7	Wind speed (m/s) and QC flag
direction, qc	F13.5, I7	Wind direction (degrees) and QC flag
u, qc	F13.5, I7	u component of wind (m/s), and QC flag
v, qc	F13.5, I7	v component of wind (m/s), and QC flag
rh, qc	F13.5, I7	Relative Humidity (%) and QC flag
thickness, qc	F13.5, I7	Thickness (m), and QC flag

Tabla 5.4: Flags de control de calidad (QC flags).

Class	Quality
Flag 0	No QC done
Flag 1	Good data
Flag 2	“Probably” good data
Flag 3	“Probably” bad data
Flag 4	Bad data
Flag 5	Changed

número de errores y con el número de warnings respectivamente.

5.2.2. Conversión de las salidas del GFA- \mathcal{M}^3 a LITTLE_R

El listado 5.1 muestra un extracto de los datos obtenidos de la plataforma GFA- \mathcal{M}^3 a través del servicio sql2csv explicado en el apartado 4.1.3 del capítulo 4. En concreto, hay datos de presión atmosférica, temperatura, temperatura del punto de rocío, velocidad del viento, dirección del viento y humedad relativa, correspondientes al 31 de enero de 2015, procedentes de un observador voluntario en la provincia de Palencia.

```

1 OBV_005_PAL,-4.526667,42.007500,2015-01-31 00:00:00,993.6,Bar,hPa
2 OBV_005_PAL,-4.526667,42.007500,2015-01-31 00:00:00,8.2,Temp_Out,C
3 OBV_005_PAL,-4.526667,42.007500,2015-01-31 00:00:00,5.2,Dew_Pt,C
4 OBV_005_PAL,-4.526667,42.007500,2015-01-31 00:00:00,19.3,Wind_Speed,Km/h
5 OBV_005_PAL,-4.526667,42.007500,2015-01-31 00:00:00,W,Wind_Dir,
6 OBV_005_PAL,-4.526667,42.007500,2015-01-31 00:00:00,81,Out_Hum,%

```

Listado 5.1: Datos de observación obtenidos de la plataforma GFA- \mathcal{M}^3 .

Convertir los datos obtenidos del GFA- \mathcal{M}^3 en formato CSV a LITTLE_R es una tarea relativamente sencilla que puede realizarse mediante un script de shell o un script Python. El listado 5.2 muestra un fichero LITTLE_R con un único report que contiene la información de los datos anteriores (listado 4.4). En el listado se resaltan los espacios en blanco y, en color magenta, los valores obtenidos de las observaciones de la plataforma GFA- \mathcal{M}^3 para una mejor visualización. El resto de campos se completa con espacios en blanco o valores nulos (-888888.00000) según la especificación del apartado anterior. A los flags de control de calidad se les asigna el valor 0. La primera línea se corresponde con el encabezado. La segunda línea contiene un único registro de datos, y la tercera, marca el final de la sección de datos. Por último, la cuarta línea contiene los contadores de datos correctos, erróneos y warnings. Cabe destacar también el cambio de unidades ya que LITTLE_R exige utilizar las unidades del sistema internacional. En concreto para el ejemplo anterior: *Pa* para la presión atmosférica, grados *K* para las temperaturas, *m/s* para velocidad del viento, y grados para la dirección del viento.

```

1  _____-4.52667_____42.00750_____
   _____FM-12_SYNOP_____
   _____0.00000_____1____
-888888_____888888_____888888_____F_____F_____F_____888888_____
  _____888888_____20150131160000-888888.00000_____0-888888.00000_____
0-888888.00000_____0-888888.00000_____0-888888.00000_____0-888888.00000_____
  _____0-888888.00000_____0-888888.00000_____0-888888.00000_____0-888888.00000_____
  _____0-888888.00000_____0-888888.00000_____0-888888.00000_____0_____
2  _____99360.00000_____0-888888.00000_____281.35000_____278.15000_____0_____
  _____5.36111_____0_____180_____0-888888.00000_____0-888888.00000_____0_____
  _____81_____0-888888.00000_____0_____
3  -777777.00000_____0-777777.00000_____0-888888.00000_____0-888888.00000_____
  _____0-888888.00000_____0-888888.00000_____0-888888.00000_____0-888888.00000_____
  _____0-888888.00000_____0-888888.00000_____0_____
4  _____1_____0_____0_____

```

Listado 5.2: Datos de observación en formato LITTLE_R.

5.3. Calibración y validación de algoritmos

La información contenida en la plataforma GFA- \mathcal{M}^3 ha permitido validar y calibrar la precisión de algunos de los algoritmos utilizados por el GFA. Este apartado presenta tres herramientas utilizadas para realizar análisis de datos meteorológicos en tiempo real:

- *WRF precipitation evaluation tool*, García-Ortega *et al.* (2012).
- *Satellite Precipitation Evaluation Application*, Merino *et al.* (2014b).
- *Hail Precipitation Evaluation Application*, Guerrero-Higuera *et al.* (2013d); Merino *et al.* (2014c).

Los algoritmos utilizados en cada una de las herramientas enumeradas han sido calibrados y validados a partir de los datos de la plataforma GFA- \mathcal{M}^3 , como se explica en los siguientes apartados.

5.3.1. WRF precipitation evaluation tool

Esta herramienta permite evaluar la precisión de las predicciones obtenidas con el modelo WRF para la mitad norte de la Península Ibérica. Para ello, se comparan los resultados de éstas con los datos de observación de las estaciones meteorológicas de la plataforma GFA- \mathcal{M}^3 . La figura 5.4 muestra una comparativa de precipitación y temperatura en superficie, obtenidas en un pronóstico del modelo WRF, con datos de observación en la misma zona y periodo de tiempo.

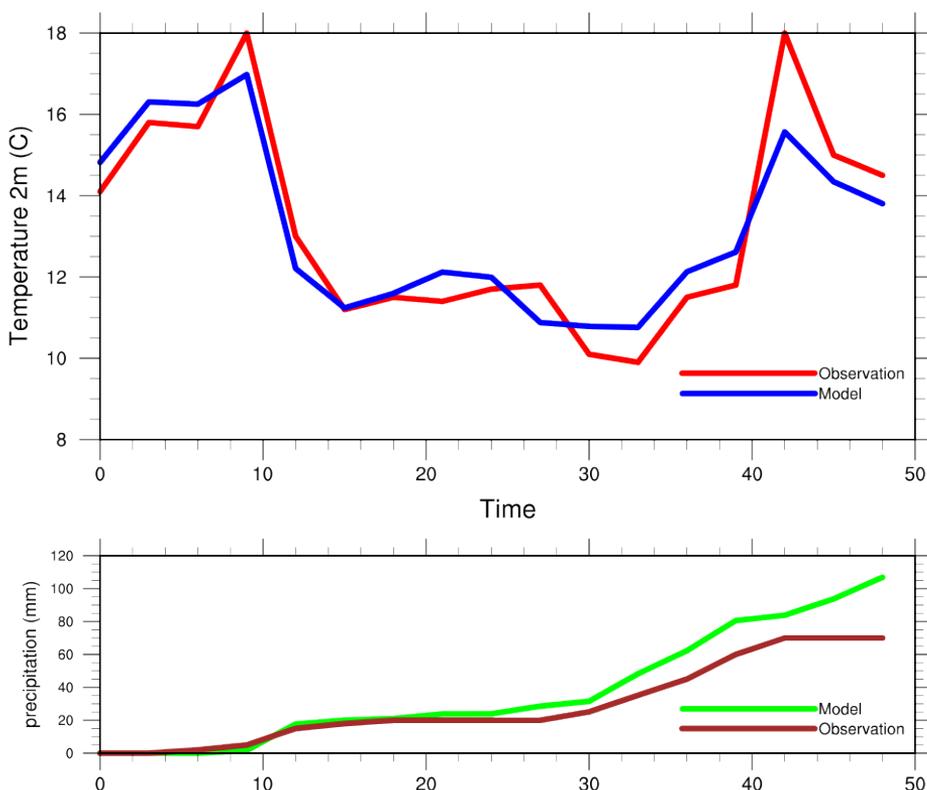


Figura 5.4: Evaluación de las predicciones del modelo WRF. Comparativa de temperatura en superficie (arriba) y precipitación total (abajo).

5.3.2. The Satellite Precipitation Evaluation Application

Esta herramienta utiliza el algoritmo Multi-sensor Precipitation Estimate (MPE) [Heinemann *et al.* (2002)] para obtener una estimación de la intensidad de la precipitación a partir de los datos de satélite MSG contenidos en la plataforma GFA- \mathcal{M}^3 . La figura 5.5 muestra la estimación de la intensidad de la precipitación, calculada mediante el algoritmo MPE, en Castilla y León, entre las 14:00 y las 23:00 horas UTC, el 1 de mayo de 2014. Las áreas marcadas en azul representan zonas con una intensidad de precipitación superior a 1 mm/h.

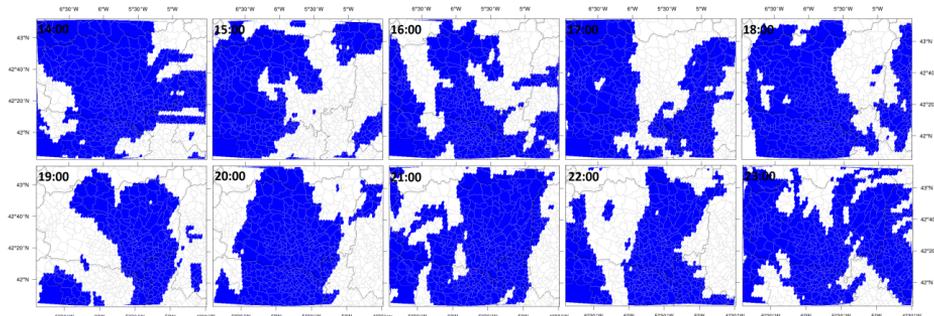


Figura 5.5: Intensidad de precipitación calculada mediante MPE.

La figura 5.6 muestra la precipitación observada por las estaciones meteorológicas de la plataforma GFA- \mathcal{M}^3 en la misma zona y en el mismo periodo de tiempo que la estimación de la figura 5.5. Las áreas marcadas en azul representan zonas con una intensidad de precipitación superior a 1 mm/h.

Comparando los resultados de la figura 5.6 con la estimación calculada mediante el algoritmo MPE (figura 5.5), se pueden identificar las zonas donde la estimación se corresponde con los datos de observación. La figura 5.7 muestra marcadas en rojo las zonas donde la estimación es correcta, las zonas marcadas en azul se corresponden con falsas alarmas y las zonas marcadas en verde identifican puntos donde no hay datos suficientes para realizar la comparativa.

5.3.3. Hail Precipitation Evaluation Application

Esta herramienta permite identificar tormentas con presencia de granizo a partir de información de satélite, Merino *et al.* (2014c). Para validar el algoritmo se han utilizado datos de radar.

La figura 5.8 muestra, a la izquierda, una imagen obtenida por el radar meteorológico del GFA. Las zonas marcadas en rojo representan localizaciones con una alta probabilidad de tormentas con presencia de granizo. A la derecha, muestra

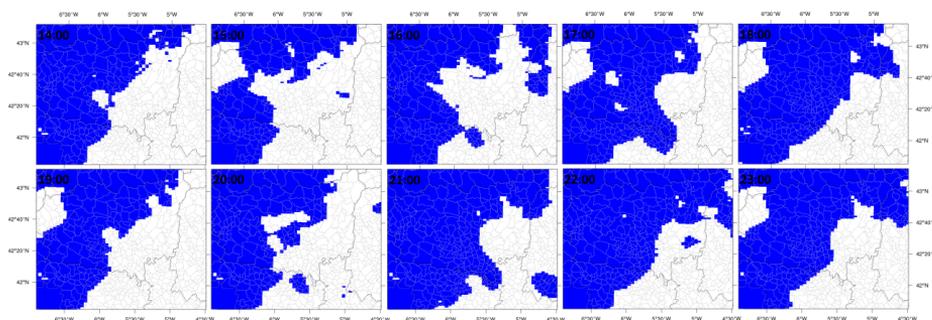


Figura 5.6: Precipitación observada por las estaciones meteorológicas de la plataforma GFA- \mathcal{M}^3 .

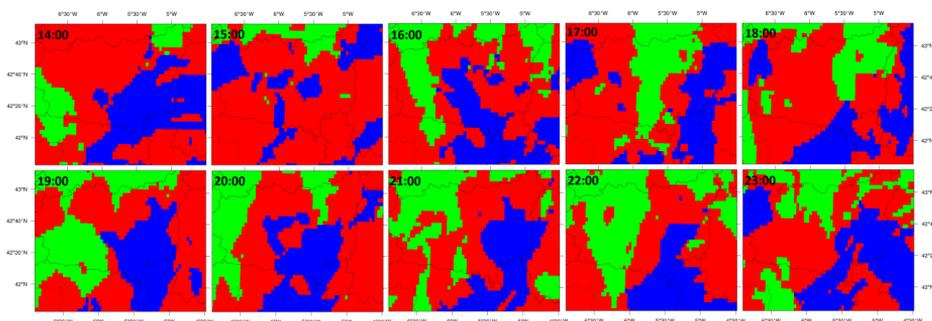


Figura 5.7: Comparativa de la intensidad de precipitación calculada mediante MPE con los datos de observación de la plataforma GFA- \mathcal{M}^3 .

la probabilidad de granizo calculada según el algoritmo propuesto en Merino *et al.* (2014c) e implementado según Guerrero-Higueras *et al.* (2013d).

Ambas imágenes corresponden a un episodio ocurrido el 21 de julio de 2010 en el nordeste de la Península Ibérica en el cual se registraron numerosas tormentas con presencia de granizo.

5.4. Desarrollo de Aplicaciones

Este apartado presenta una serie de aplicaciones desarrolladas a partir de los datos de la plataforma GFA- \mathcal{M}^3 , en concreto las siguientes:

- Cálculo de parámetros para la caracterización de tormentas severas con

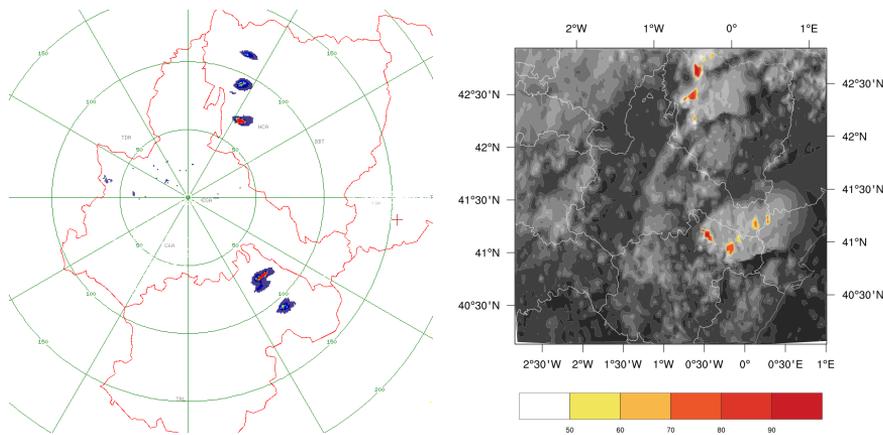


Figura 5.8: Imagen del radar meteorológico del GFA (izquierda). Probabilidad de granizo, calculada según el algoritmo propuesto en Merino *et al.* (2014c) (derecha).

presencia de granizo.

- *PyHDT*, aplicación para el cálculo de la probabilidad de granizo.
- Sistema de información meteorológica para un vehículo.

Las siguientes secciones profundizan en cada una de las herramientas enumeradas.

5.4.1. Caracterización de tormentas severas con presencia de granizo

La información contenida en la plataforma GFA- \mathcal{M}^3 puede utilizarse directamente, o bien servir como base para cálculos más complejos. Un ejemplo de esto es el método propuesto en Merino *et al.* (2013) para caracterizar tormentas severas con precipitación en forma de granizo en el valle del río Ebro, en el noreste de la Península Ibérica. Para ello, primero se analizan las características del entorno sinóptico y la distribución espacial de las tormentas en los días con precipitación en forma de granizo. A continuación, se establecen las configuraciones mesoescalares de tal manera que estas puedan relacionarse con la precipitación observada. Por último, se estudian varios parámetros que determinan la aparición de granizo dentro de los entornos convectivos. Gracias a este estudio, es posible predecir, con cierto índice de fiabilidad, la aparición de eventos con precipitación convectiva en el área de estudio. La presencia de granizo dentro de estos eventos se realiza a partir de ciertos parámetros.

Para realizar este estudio ha sido necesario disponer de un gran volumen de datos, tanto de verdad terreno, como pronosticados; registrados durante mucho tiempo. En concreto, se han seleccionado 100 días, entre de mayo y septiembre, y entre los años 2001 y el 2010, donde el radar meteorológico del GFA registró episodios de precipitación convectiva. La precipitación convectiva se identifica utilizando la herramienta TITAN [Dixon y Wiener (1993)], que permite el seguimiento de células convectivas de más de 30 dBZ.

Para calcular los campos que permiten definir las configuraciones mesoescalares se utilizan algunas de las variables meteorológicas que predice el modelo WRF. Como se explica en el apartado 3.2.2, estas ejecuciones cubren el valle del río Ebro, en el noreste de la Península Ibérica, y por tanto sus salidas pueden utilizarse para calcular las configuraciones mesoescalares que propone Merino *et al.* (2013). Los campos que permiten definir las configuraciones mesoescalares son los siguientes [García-Ortega *et al.* (2012)]:

- Divergencia del flujo de vapor de agua (WVFD). La WVFD viene dada por la expresión que muestra la ecuación 5.1:

$$WVFD = \frac{1}{g} \int_{p_0}^p \nabla \cdot (q \vec{v}) dp \quad (5.1)$$

donde \vec{v} es la velocidad del viento, $p_0 = 1000 \text{ hPa}$ y $p = 700 \text{ hPa}$. \vec{v} , en todos los niveles de presión, incluidos 1000 hPa y 700 hPa , es una de las variables que se predice con el modelo WRF.

- Índice de inestabilidad convectiva (CI). El CI se define como la diferencia entre la temperatura potencial equivalente a 700 hPa y a 900 hPa , como muestra la ecuación 5.2.

$$CI = the_{e700} - the_{e900} \quad (5.2)$$

Para calcular la temperatura potencial equivalente a 700 hPa y a 900 hPa se utilizan las siguientes variables obtenidas por el modelo WRF: Temperatura (K), proporción de mezcla de vapor de agua (kg/kg) y presión (Pa) [Bolton (1980)].

- Humedad específica a 850 hPa (Q_{850}). La humedad específica Q , en todos los niveles de presión, incluido 850 hPa , es una de las variables que se predice con el modelo WRF.
- Viento a 850 hPa (\vec{v}_{850}). El viento \vec{v} , en todos los niveles de presión, incluido 850 hPa , es una de las variables que se predice con el modelo WRF.

La figura 5.9 muestra una representación gráfica de los 4 factores utilizada en Merino *et al.* (2013).

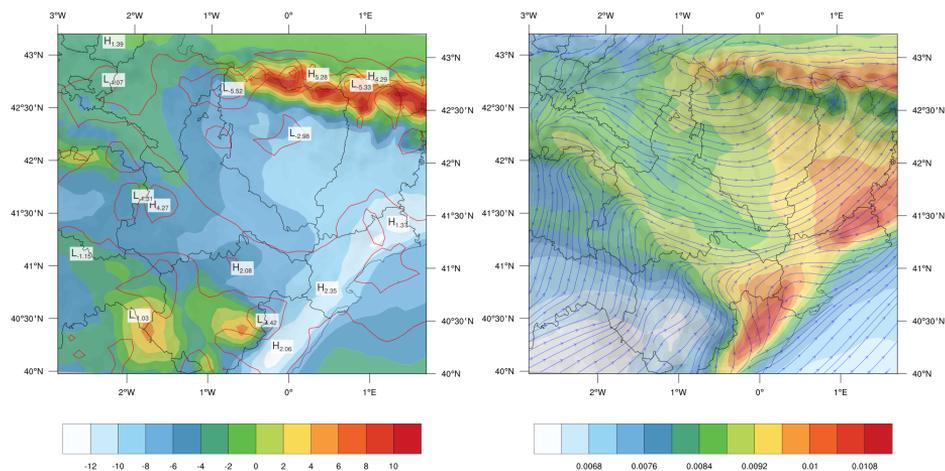


Figura 5.9: Izquierda: WVFD (líneas de contorno, $gm^{-2}s^{-1}$), convergencias (L), divergencias (H) y CI (escala de colores, $^{\circ}C$); Derecha: Q_{850} (líneas de corriente) y \vec{v}_{850} (escala de colores, $kg\ kg^{-1}$).

Con respecto a los parámetros que determinan la presencia de granizo en episodios de precipitación convectiva, se calculan también a partir de se utilizan algunas de las variables meteorológicas que predice el modelo WRF. Por tanto, también puede utilizarse la información de la plataforma GFA- \mathcal{M}^3 . Los parámetros que determinan la presencia de granizo son los siguientes:

- Convective Available Potential Energy (CAPE), Moncrieff (1981), es un indicador utilizado habitualmente para la predicción de supercélulas que producen aguaceros severos. Hay estudios en Europa que indican que la presencia de CAPE por encima de $400\ J\ hg^{-1}$ es un buen indicador de presencia de granizo [Kaltenböck *et al.* (2009)].
- Storm-relative helicity between 0 and 3 km (SRH_{0-3}), Droegemeier *et al.* (1993). En las áreas que rodean las zonas montañosas, las modificaciones de los flujos de viento en capas bajas favorecen la cizalladura direccional (cambios fuertes en la dirección del viento), jugando un papel importante en el comienzo y en el desarrollo de tormentas severas de ámbito local en Europa [Hannesen *et al.* (2000); Kaltenböck (2004)].
- Showalter Index (SI), Hart y Korotky (1991); Showalter (1953), puede ser útil para predecir tormentas, aplicando un umbral de 3. Para tormentas

severas en Europa, el valor medio es de 0, excepto cuando hay presencia de granizo, que es de -1 [Huntrieser *et al.* (1997); Kaltenböck *et al.* (2009)].

- Energy–Helicity Index (EHI), Davies (1993); Hart y Korotky (1991). Este índice relaciona CAPE y SRH_{0-3} y se utiliza habitualmente para predecir supercélulas. Cuando el valor de este índice es superior a 1, indica una alta probabilidad de supercélulas [Rasmussen y Blanchard (1998)].

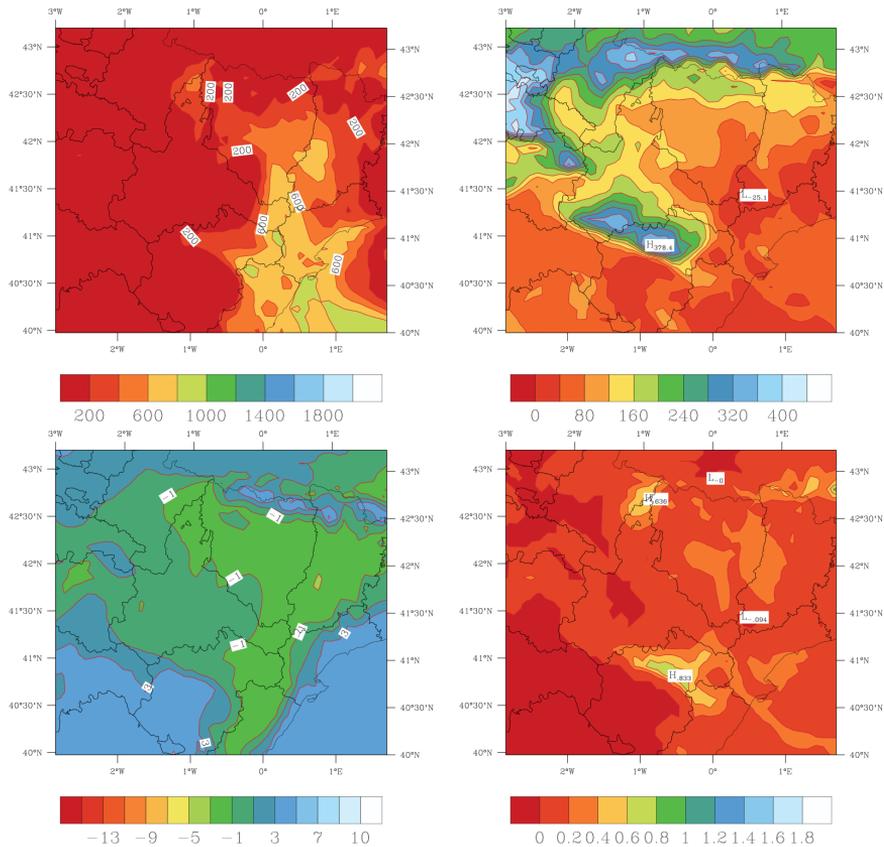


Figura 5.10: Arriba: CAPE ($\frac{J}{kg}$) y SRH_{0-3} ($\frac{m^2}{s^2}$); Abajo: SI ($^{\circ}C$) y EHI.

La figura 5.10 muestra una representación gráfica de los parámetros que determinan la presencia de granizo en episodios de precipitación convectiva Merino *et al.* (2013).

Las variables necesarias para calcular los factores mesoescalares (WVFD, CI, Q_{850} y F_{850}) y los parámetros que determinan la presencia de granizo (CAPE,

SRH_{0-3} , SI y EHI), están disponibles en las salidas del modelo WRF contenidas en la plataforma GFA- \mathcal{M}^3 , y pueden consultarse a través de los servicios de consulta del mismo. Además de las salidas del modelo WRF, y los datos del radar meteorológico del GFA, se ha utilizado la información de los pluviómetros de la CHE, para validar la veracidad de los datos de precipitación obtenidos por el modelo. Estos datos también están disponibles en la plataforma GFA- \mathcal{M}^3 .

5.4.2. Probabilidad de granizo

Las tormentas severas con precipitación en forma de granizo es uno de los riesgos meteorológicos más comunes en Europa, produciendo grandes pérdidas económicas. La identificación de este tipo de episodios es un demanda habitual de la población en general, y de los organismos de gestión de riesgos en particular. El GFA ha desarrollado un modelo que permite la identificación en tiempo real (*nowcasting*) de precipitaciones en forma de granizo, utilizando para ello la información recogida por el satélite MSG [Merino (2013); Merino *et al.* (2014c)].

El cálculo de la probabilidad de granizo en tiempo real utilizando el algoritmo propuesto por el GFA requiere varios pasos que se enumeran a continuación. Cada paso se detalla en los siguientes apartados.

1. Definición del área geográfica de estudio.
2. Cálculo de la probabilidad de granizo en cada *pixel*.
3. Corrección de la probabilidad en base al ángulo zenital.
4. Corrección de las coordenadas por el efecto *Parallax*.

El GFA ha desarrollado una herramienta, denominada *PyHDT* [Guerrero-Higueras *et al.* (2013d)], que implementa el cálculo de la probabilidad de granizo. *PyHDT* se compone de dos scripts:

- *hail.py* realiza el cálculo de la probabilidad y corrige el resultado en base al ángulo zenital.
- *hail.ncl* genera una gráfica que permite visualizar la probabilidad de granizo. También corrige la distorsión en las coordenadas producidas por el efecto *Parallax*.

Definición del área geográfica de estudio

Es necesario discretizar el área para poder estudiarla. Para ello se define una malla bidimensional que abarca el área de estudio (*grid*). Cada punto del grid se denomina *pixel*. Cuanto mayor sea el número de pixels en el grid, mayor resolución se obtendrá en los resultados. Por contra, el tiempo de cálculo también será mayor.

Obtención de la probabilidad de granizo

La probabilidad de granizo P obtenida por el modelo propuesto por el GFA viene dada por la ecuación 5.3:

$$P = \frac{\exp Z_1}{1 + \exp Z_1} \quad (5.3)$$

donde $Z_1 = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$; X_1, X_2, \dots, X_k son variables explicativas (relaciones entre variables incluidas en el modelo); k es el número total de variables explicativas; α es una constante; y $\beta_1, \beta_2, \dots, \beta_k$ representan una serie de pesos discriminantes asociados a cada variable.

El cálculo se realiza en dos pasos como se detalla a continuación.

1. Las nubes objeto de interés para el estudio se separan del resto de formaciones nubosas. En concreto, se detectan formaciones convectivas bien desarrolladas, habitualmente conocidas como cumulonimbos (*cumulonimbus clouds*). Se denomina *máscara de convección* al resultado de este primer paso.
2. Se calcula la probabilidad de precipitación en forma de granizo dentro de estas nubes. Se le denomina *máscara de granizo* al resultado de este segundo paso.

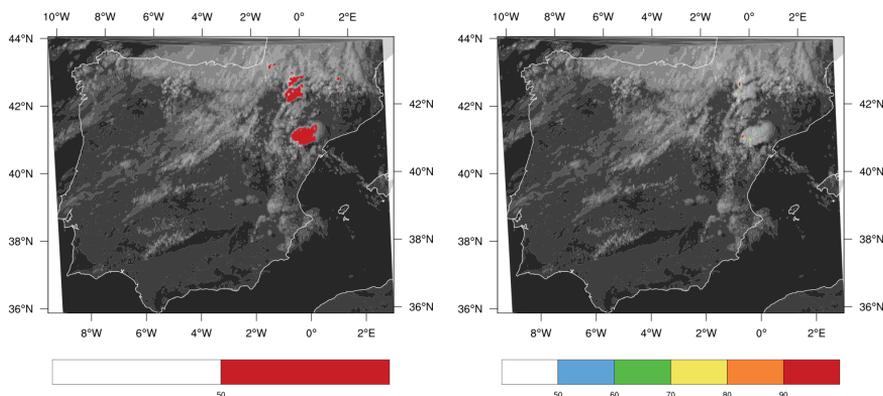
En ambos pasos se aplica la ecuación 5.3, pero con las variables explicativas de la tabla 5.5 para obtener la máscara de convección, y con las variables de la tabla 5.6) para la máscara de granizo. La figura 5.11 muestra una representación de las máscaras de convección y granizo respectivamente.

Tabla 5.5: Parámetros para la máscara de convección.

Metric explanatory variables	β
Incercept	1492.636
Brightness temperature 8.7 μm	1.188
Brightness temperature 6.2 μm	-5.186
Albedo 1.6 μm	2.226
Albedo 0.8 μm	-1.659
Albedo 3.9 μm	-0.884
Brightness temperature 7.3 μm	-7.627
(Albedo 1.6 μm) * (Brightness temperature 8.7 μm)	-0.009810
(Brightness temperature 6.2 μm) * (brightness temperature 7.3 μm)	0.026309
(Albedo 0.8 μm) * (Albedo 3.9 μm)	0.007047

Tabla 5.6: Parámetros para la máscara de granizo.

Metric explanatory variables	β
Incercept	115.039
Brightness temperature 6.2 μm	-0.624
Albedo 1.6 μm	-2.18
Albedo 0.8 μm	0.118
(Albedo 1.6 μm) * (Brightness temperature 6.2 μm)	0.010955

**Figura 5.11:** Máscara de convección (izquierda) y máscara de granizo (derecha).

Para obtener los valores de las variables necesarias, se utiliza la información obtenida de la aplicación SatDatos, mencionada en el apartado 3.2.1). SatDatos transforma la información recibida en formato HRIT a formato netCDF. SatDatos devuelve un fichero netCDF que incluye una variable denominada *britemp*. La variable *britemp* contiene a su vez información sobre todas las variables necesarias para el cálculo de la probabilidad de granizo. Las variables contenidas dentro de *britemp* pueden consultarse en la tabla 5.7.

Para calcular la máscara de convección se multiplica el resultado de la ecuación 5.3 por 100 para obtener valores de prioridad entre 0 y 100, como muestra la 5.4:

$$P_0 = 100 \cdot \left(\frac{\exp X}{1 + \exp X} \right) \quad (5.4)$$

donde X se calcula aplicando las variables y los coeficientes β de la tabla 5.5, según la ecuación 5.5:

Tabla 5.7: Variables obtenidas por *SatDatos*.

Metric explanatory variables	
Albedo 0.6 μm	C_1
Albedo 0.8 μm	C_2
Albedo 1.6 μm	C_3
Albedo 3.9 μm	C_4
Brightness temperature 6.2 μm	C_5
Brightness temperature 7.3 μm	C_6
Brightness temperature 8.7 μm	C_7
Brightness temperature 9.7 μm	C_8
Brightness temperature 10.8 μm	C_9
Brightness temperature 12.0 μm	C_{10}
Brightness temperature 13.4 μm	C_{11}
High resolution visible	C_{12}

$$\begin{aligned}
X &= 1492,636 \\
&+ (1,188 \cdot C_7) \\
&+ (-5,186 \cdot C_5) \\
&+ (2,266 \cdot C_3) \\
&+ (-1,659 \cdot C_2) \\
&+ (-0,884 \cdot C_4) \\
&+ (-7,627 \cdot C_6) \\
&+ (-0,00980977 \cdot C_3 \cdot C_7) \\
&+ (0,02630949 \cdot C_5 \cdot C_6) \\
&+ (0,00704733 \cdot C_2 \cdot C_4)
\end{aligned} \tag{5.5}$$

El algoritmo 8 muestra la implementación del cálculo de la máscara de convección.

Algoritmo 8 Cálculo de la máscara de convección.

Input: $C_2, C_3, C_4, C_5, C_6, C_7$

Output: P_0

1: $X \leftarrow 1492,636 + (1,188 \cdot C_7) + (-5,186 \cdot C_5) + (2,266 \cdot C_3) + (-1,659 \cdot C_2) + (-0,884 \cdot C_4) + (-7,627 \cdot C_6) + (-0,00980977 \cdot C_3 \cdot C_7) + (0,02630949 \cdot C_5 \cdot C_6) + (0,00704733 \cdot C_2 \cdot C_4)$

2: $P_0 \leftarrow 100 \cdot \left(\frac{\exp X}{1 + \exp X} \right)$

3: **return** P_0

Para calcular la *máscara de granizo*, también se aplica la ecuación 5.3, multiplicando el resultado por 100 para obtener valores de probabilidad. Además, también se multiplica por el resultado de la *máscara de convección* P_0 , para trabajar solamente con los puntos de interés, como muestra la ecuación 5.6.

$$P_1 = P_0 \cdot 100 \cdot \left(\frac{\exp Y}{1 + \exp Y} \right) \quad (5.6)$$

Antes de hacer esto, es necesario discretizar los valores de la máscara de convección P_0 , para señalar los puntos de la malla (grid) candidatos para producir granizo. Para hacer esto, se asigna el valor 1 a aquellos puntos con una probabilidad mayor que 50 y el valor 0 a los puntos con una probabilidad inferior a 50.

Y se calcula aplicando las variables y los coeficientes β de la tabla 5.6, como muestra la ecuación 5.7.

$$\begin{aligned} Y = & 115,039 \\ & + (-0,624 \cdot C_5) \\ & + (-2,180 \cdot C_3) \\ & + (0,118 \cdot C_2) \\ & + (0,01095546 \cdot C_3 \cdot C_5) \end{aligned} \quad (5.7)$$

El algoritmo 9 muestra la implementación del cálculo de la máscara de granizo.

Algoritmo 9 Cálculo de la máscara de convección.

Input: P_0, C_2, C_3, C_5

Output: P_1

```

1:  $Y \leftarrow 115,039 + (-0,624 \cdot C_5) + (-2,180 \cdot C_3) + (0,118 \cdot C_2) + (0,01095546 \cdot C_3 \cdot C_5)$ 
2:  $mask \leftarrow P_0$ 
3: for  $i = 0$  to  $size(mask)$  do
4:   if  $mask[i] > 50$  then
5:      $mask[i] = 1$ 
6:   else
7:      $mask[i] = 0$ 
8:   end if
9: end for
10:  $P_1 \leftarrow mask \cdot 100 \cdot \left( \frac{\exp Y}{1 + \exp Y} \right)$ 
11: return  $P_1$ 

```

Corrección de la probabilidad en base al ángulo zenital

Es necesario realizar una serie de correcciones en los resultados obtenidos en el paso anterior. Hay que ajustar la probabilidad de cada pixel, teniendo en cuenta

el ángulo zenital, dado que el algoritmo esta optimizado para el cálculo de la probabilidad en condiciones diurnas.

Para corregir los valores de probabilidad para un instante determinado, es necesario conocer el valor del ángulo zenital (*zenith angle*) θ , para cada punto y cada instante. Para calcular el valor de θ hay que seguir los siguientes pasos:

1. Calcular δ de acuerdo con la ecuación 5.8:

$$\delta = 23,45 \cdot \sin \left(360 \cdot \left(\frac{284 + J}{365} \right) \right) \quad (5.8)$$

donde J se corresponde con el día del año.

2. Calcular la hora solar H en los términos la ecuación 5.9:

$$H = hour + \left(\frac{4 \cdot lon \cdot 100}{60 \cdot 100} \right) \quad (5.9)$$

donde *hour* es la hora UTC en el instante para el cual estamos calculando el ángulo zenital; y *lon* es el valor de la longitud, en grados, del punto geográfico en el cual estamos calculando el ángulo.

3. Calcular valor de ω se calcula según la ecuación 5.10:

$$\omega = 15 \cdot (H - 12) \quad (5.10)$$

4. Por último, el valor del ángulo zenital θ , se calcula en los términos de la ecuación 5.11:

$$\theta = \arccos(\sin \delta \cdot \sin \phi + \cos \delta \cdot \cos \phi \cdot \cos \omega) \quad (5.11)$$

donde ϕ se corresponde con el valor de la latitud, en radianes, del punto geográfico en el cual estamos calculando el ángulo zenital.

`hail.py` calcula el ángulo zenital para todos los puntos del *grid* siguiendo los pasos del algoritmo 10.

Corrección de las coordenadas

Es necesario modificar las coordenadas de cada pixel para corregir el efecto Parallax. Esta distorsión en las coordenadas se produce por la posición del satélite en relación con la curvatura de la Tierra. La corrección se realiza siguiendo el método propuesto en Vicente *et al.* (2002) siguiendo los pasos del algoritmo 11.

Para realizar la corrección es necesario conocer la altura de la nube en cada punto (*cloudHeight* en el algoritmo 11). Para calcularla, se compara el valor de

Algoritmo 10 Cálculo del ángulo zenital.**Input:** $lon, lat, datetime$ **Output:** θ

```

1:  $J \leftarrow \text{dat\_of\_year}(datetime)$ 
2:  $H \leftarrow \text{hour} + \left( \frac{4 \cdot lon \cdot 100}{60 \cdot 100} \right)$ 
3:  $\phi \leftarrow \text{degrees2radians}(lat)$ 
4:  $\delta \leftarrow \text{degrees2radians} \left( 23,45 \cdot \sin \left( 360 \cdot \left( \frac{284+J}{365} \right) \right) \right)$ 
5:  $\omega \leftarrow \text{degrees2radians}(15 \cdot (H - 12))$ 
6:  $\theta \leftarrow \text{arc cos}(\sin \delta \cdot \sin \phi + \cos \delta \cdot \cos \phi \cdot \cos \omega)$ 
7: return  $\theta$ 

```

la variable *brightness temperature 12 μm* (C_{10}), con un perfil de temperatura vertical. Brightness temperature 12 μm se obtiene mediante la aplicación *SatDatos* (tabla 5.7). El perfil de temperatura vertical se obtiene a partir de las salidas del modelo WRF. El cálculo se realiza siguiendo los siguientes pasos:

1. Extraer perfiles verticales de temperatura y altitud. La temperatura se corresponde con la variable TK de las salidas del modelo WRF. La altitud se calcula a partir de las variables PH (perturbation geopotential) y PHB (base-state geopotential) de las salidas del modelo WRF, según el algoritmo 12.
2. Calcular la altura de la nube a partir de las variables del paso anterior y de la variable C_{10} , siguiendo los pasos del algoritmo 13.

5.4.3. Sistema de información meteorológica para vehículos

El sistema de información meteorológica para un vehículo, permite mostrar información meteorológica y alertas en tiempo real en la localización actual o en una localización próxima con el fin de proporcionar la ruta más segura posible. El sistema ha sido patentado por el GFA con el código ES2462552B1. Ofrece recomendaciones sobre como evitar y/o actuar ante las diferentes alertas. Además, el dispositivo permite transmitir alertas meteorológicas entre diferentes usuarios.

La figura 5.12 muestra la arquitectura del sistema. Incluye un interfaz (4), donde recibe una localización inicial y un destino, y calcula una ruta preliminar entre ambos. El dispositivo incluye un módulo de observación que recopila datos meteorológicos (1) que afectan a la ruta preliminar. El modulo de observación recoge datos de diversas fuentes: Estaciones meteorológicas cercanas (11), radar (12), red de descargas eléctricas (13) y satélite (14). Estos datos de observación son consultados en la plataforma GFA- \mathcal{M}^3 . También se toman datos de sensores de temperatura (15) y humedad (16) instalados en el propio vehículo. El sistema

Algoritmo 11 Corrección del efecto parallax.**Input:** $lon, lat, cloudHeight$ **Output:** $loncorr, latcorr$

```

1:  $satheight \leftarrow 36000 + 6378,077$ 
2:  $satlon \leftarrow \text{degrees2radians}(0,0)$ 
3:  $satlat \leftarrow \text{degrees2radians}(0,0)$ 
4:  $radius\_eq \leftarrow 6378,077$  ▷ Varius earth radius information
5:  $radius\_pole \leftarrow 6356,577$ 
6:  $radius\_ratio \leftarrow \frac{radius\_eq}{radius\_pole}$ 
7:  $mean\_radius \leftarrow 0,5 \cdot (radius\_eq + radius\_pole)$ 
8:  $zdiff \leftarrow 0,0$ 
9:  $alon \leftarrow \text{degrees2radians}(lon)$ 
10:  $alat \leftarrow \text{degrees2radians}(lat)$ 
11:  $satlat\_geod \leftarrow \arctan(\tan satlat \cdot radius\_ratio^2)$  ▷ Cartesian coordinates for the satellite
12:  $xsat \leftarrow satheight \cdot \cos satlat\_geod \cdot \sin satlon$ 
13:  $ysat \leftarrow satheight \cdot \sin satlat\_geod$ 
14:  $zsat \leftarrow satheight \cdot \cos satlat\_geod \cdot \cos satlon$ 
15:  $alat\_geod \leftarrow \arctan(\tan alat \cdot radius\_ratio^2)$  ▷ Cartesian coordinates of the surface point
16:  $radius\_surf \leftarrow \frac{radius\_eq}{\sqrt{(\cos alat\_geod)^2 + radius\_ratio^2 \cdot (\sin alat\_geod)^2}}$ 
17:  $xsurf \leftarrow radius\_surf \cdot \cos alat\_geod \cdot \sin alon$ 
18:  $ysurf \leftarrow radius\_surf \cdot \sin alat\_geod$ 
19:  $zsurf \leftarrow radius\_surf \cdot \cos alat\_geod \cdot \cos alon$ 
20:  $radius\_ratio\_local \leftarrow \left( \frac{radius\_eq + cloudHeight}{radius\_pole + cloudHeight} \right)^2$  ▷ Compute new radius ratio depending on height
21:  $xdiff \leftarrow xsat - xsurf$  ▷ Satellite minus surface location
22:  $ydiff \leftarrow ysat - ysurf$ 
23:  $zdiff \leftarrow zsat - zsurf$ 
24:  $e_1 \leftarrow xdiff^2 + radius\_ratio\_local \cdot ydiff^2 + zdiff^2$  ▷ Equation to solve for the line of sight at height Z
25:  $e_2 \leftarrow 2,0 \cdot (xsurf \cdot xdiff + radius\_ratio\_local \cdot ysurf \cdot ydiff + zsurf \cdot zdiff)$ 
26:  $e_3 \leftarrow xsurf^2 + zsurf^2 + radius\_ratio\_local \cdot ysurf^2 - (radius\_eq + cloudHeight)^2$ 
27:  $corr \leftarrow \frac{\sqrt{e_2^2 - 4,0 \cdot e_1 \cdot e_3} - e_2}{2,0}$ 
28:  $xcorr \leftarrow xsurf + corr \cdot xdiff$  ▷ Corrected surface coordinates
29:  $ycorr \leftarrow ysurf + corr \cdot ydiff$ 
30:  $zcorr \leftarrow zsurf + corr \cdot zdiff$ 
31:  $latcorr \leftarrow \arctan\left(\frac{ycorr}{\sqrt{xcorr^2 + zcorr^2}}\right)$ 
32:  $latcorr \leftarrow \arctan\left(\frac{\tan latcorr}{radius\_ratio^2}\right) \cdot \frac{180,0}{\Pi}$ 
33:  $loncoor \leftarrow \arctan(xcoor, zcoor) \cdot \frac{180,0}{\Pi}$ 
34: return  $loncoor, latcorr$ 

```

Algoritmo 12 Cálculo de la altitud a partir de las variables PH y PHB.

Input: PH, PHB ▷ input netCDF file
Output: $height$
1: $t \leftarrow \frac{PH+PHB}{9,81}$ ▷ Perturbation geopotential and Base-state geopotential on full (w) levels
2: $dimv \leftarrow \text{dimsizes}(t)$ ▷ Get dimensions of t
3: $height \leftarrow 0,5 \cdot (t[0 : dimv(0) - 2, :, :] + t[0 : dimv(0) - 1, :, :])$ ▷ On half (mass) levels
4: **return** $height$

Algoritmo 13 Cálculo de la altura de la nube.

Input: $C_{10}, TK, height$
Output: $cloudHeight$
1: $cloudHeight \leftarrow \text{new}(\text{dimsizes}(C_{10}, \text{float}))$ ▷ Get empty grid with C_{10} dimensions
2: **for** $i = 0$ to $\text{dimsizes}(cloudHeight) - 1$ **do**
3: $indez \leftarrow 0$
4: **for** $j = 0$ to $\text{dimsizes}(TK) - 1$ **do**
5: **if** $C_{10}(i) < TK(j)$ **then**
6: $index = j$
7: **end if**
8: **end for**
9: $cloudHeight(i) = \frac{height(i)+C_{10}(i)}{TK(index)}$
10: **end for**
11: **return** $cloudHeight/1000$ ▷ Return height in kms

cuenta con un módulo de predicción (2) que aporta estimaciones de una serie de variable meteorológicas. Los datos de predicción también se obtienen de la plataforma GFA- \mathcal{M}^3 . Un módulo de integración (3) define riesgos asociados a la ruta calculados en base a la información de los módulos de observación (1) y predicción (2). La interfaz (4) ofrece una ruta alternativa en base a los riesgos asociados a la ruta preliminar. El sistema puede integrarse con otros dispositivos (5), como por ejemplo un navegador Global Positioning Systems (GPS) instalado en el vehículo.

La figura 5.13 muestra el interfaz gráfico del sistema de información meteorológica para un vehículo. La primera imagen (arriba a la izquierda) muestra el menú de inicio, donde hay que seleccionar la ubicación actual (41), el destino (42) y, opcionalmente, seleccionar los ajustes deseados. La segunda imagen (arriba a la derecha) muestra el menú de opciones que ofrece el dispositivo después de calcular la ruta preliminar. Desde este menú se puede recalcular la ruta (51), ver las condiciones meteorológicas en el destino (52) o en la ubicación actual (53) en tiempo real, ver la ruta alternativa más segura (54) y las alertas meteorológicas en la ruta seleccionada (55). La tercera imagen (abajo a la izquierda) muestra el menú de selección de rutas alternativas. La cuarta imagen (abajo a la derecha) muestra las alertas meteorológicas en la ruta seleccionada.

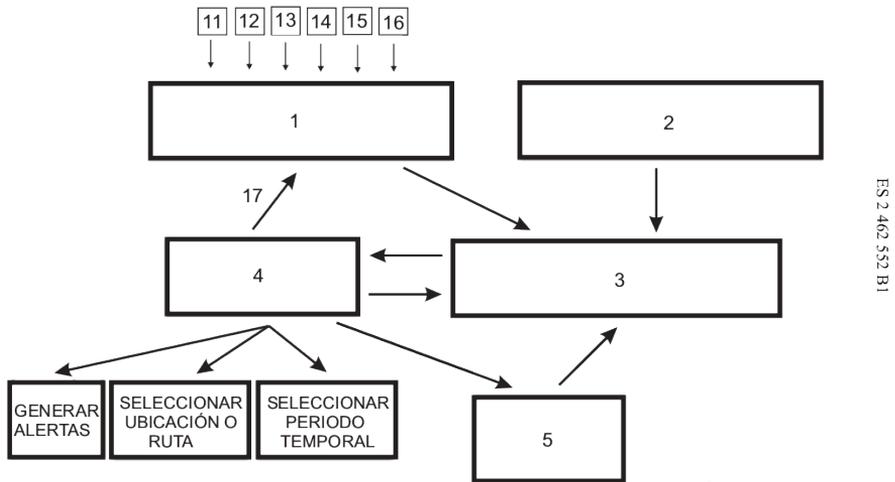


Figura 5.12: Arquitectura del sistema de información meteorológica para un vehículo (patente ES2462552B1).

5.5. Educación académica

El SGDM descrito en el capítulo 4 tiene un gran potencial en lo que a identificación y monitorización de riesgos meteorológicos, investigación científica y desarrollo de aplicaciones se refiere, como se ha visto en los apartados anteriores. Sin embargo, el GFA también pretende utilizarlo como una herramienta de apoyo a la enseñanza universitaria. En concreto, la plataforma GFA- \mathcal{M}^3 se utiliza en dos asignaturas del grado en Ciencias Ambientales y el Máster Oficial en Riesgos Naturales de la Universidad de León.

Meteorología y Climatología Los datos de la plataforma GFA- \mathcal{M}^3 se utilizan en la asignatura *Meteorología y Climatología* del segundo curso del grado en Ciencias Ambientales. En esta asignatura, entre otras habilidades, los estudiantes deben adquirir competencias en *Realización de predicciones meteorológicas en una localización determinada*. Esta parte del grado supone dos créditos ECTS, lo que convierte a esta competencia en una de las más importantes. La mayor dificultad a la hora de realizar una predicción es determinar qué modelo de predicción funciona mejor en una situación meteorológica determinada. Como resultado de esto último, los estudiantes deben analizar las predicciones de los modelos GFS, HIRLAM y WRF.

Las predicciones generadas por diferentes modelos son habitualmente similares, sin embargo, a medida que el horizonte temporal de la predicción se incremen-

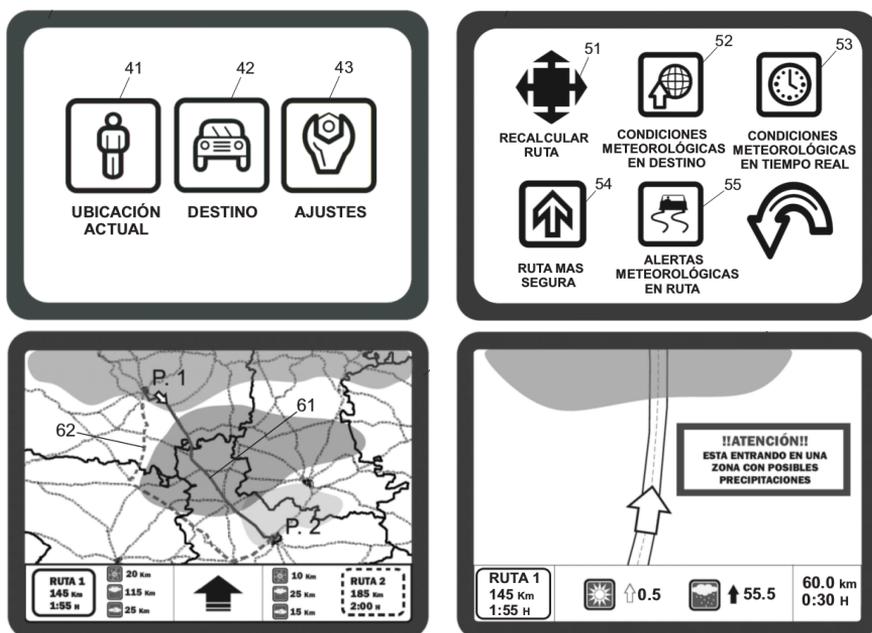


Figura 5.13: Interfaz gráfica del sistema de información meteorológica para un vehículo (patente ES246252B1).

ta, los modelos tienden a divergir, especialmente cuando se trata de precipitación y flujos de viento. La primera tarea que los estudiantes tienen que acometer en la asignatura de *Meteorología y Climatología*, consiste en evaluar qué modelo genera un pronóstico mejor en situaciones meteorológicas similares. Además, se pide a los estudiantes que hagan una predicción para los próximos días en diferentes partes de la Península Ibérica.

Para evaluar las predicciones del modelo WRF, los estudiantes deben comparar la salida del modelo con los datos de observación disponibles en la plataforma GFA- \mathcal{M}^3 . Para ello, utilizan el método propuesto en García-Ortega *et al.* (2012), que inicialmente fue desarrollado solamente con fines científicos. Esta forma de trabajar brinda a los estudiantes la oportunidad de realizar un aprendizaje autónomo.

Además de facilitar el aprendizaje autónomo, es importante destacar que esta información es de gran utilidad para evaluar el conocimiento adquirido por los estudiantes. La comparativa de las predicciones realizadas por los estudiantes con los datos contenidos en la plataforma GFA- \mathcal{M}^3 ofrece una visión objetiva de las habilidades de cada estudiante. Adicionalmente, esta información puede tratarse

en conjunto y obtener de este modo información estadística sobre la asignatura.

Observación y Vigilancia Atmosférica Los datos de la plataforma GFA- \mathcal{M}^3 también se han utilizado en la asignatura *Observación y Vigilancia Atmosférica* del Máster Oficial en Riesgos Naturales. En esta asignatura, los estudiantes aprenden a tomar decisiones relacionadas con la gestión de riesgos. Para ello, es muy importante que sean capaces de hacer el seguimiento de cualquier riesgo meteorológico con detalle. En la región de Castilla y León, la precipitación en forma de nieve, las heladas y las tormentas severas (con o sin granizo) son los más relevantes.

En resumen, el uso de los datos contenidos en la plataforma GFA- \mathcal{M}^3 y las herramientas/métodos de análisis en tiempo real de datos meteorológicos con fines educativos, facilita la evaluación del conocimiento adquirido por los estudiantes. También permite analizar sus resultados de manera global. Además, este método docente es especialmente útil para trabajar las competencias transversales *Aprendizaje y evaluación autónomos* y *Favorecer el pensamiento crítico* [ANECA (2004)], y las competencias específicas *Realizar predicciones meteorológicas* o *Gestión de riesgos meteorológicos*.

Capítulo 6

Evaluación de hipótesis, conclusiones y publicaciones

Este capítulo presenta las conclusiones alcanzadas y las contribuciones realizadas durante el desarrollo de este trabajo. Primero se expone una evaluación de las hipótesis planteadas en el capítulo 1, relacionando estas con las aplicaciones presentadas en el capítulo 5. Después, se exponen las conclusiones alcanzadas en relación a cada hipótesis. A continuación, se presentan una serie de líneas de investigación a desarrollar a partir de este trabajo. Por último, se enumeran las publicaciones producidas en el marco de esta tesis.

6.1. Evaluación de las hipótesis

Este apartado presenta la evaluación de las hipótesis 1, 2, 3 y 4, planteadas previamente en el capítulo 1, en relación a las métricas 1, 2 y 3. En las siguientes secciones se exponen las conclusiones alcanzadas relacionando cada hipótesis con las métricas correspondientes.

6.1.1. Evaluación de la hipótesis 1

La hipótesis 1 propone que un SGDM permite realizar seguimiento en tiempo real de eventos meteorológicos. La métrica 1 propone validar esta hipótesis evaluando cuantitativa y cualitativamente el seguimiento de eventos meteorológicos realizado a partir de la información de la plataforma GFA- \mathcal{M}^3 .

Para validar la hipótesis, se plantea la posibilidad de utilizar la información recopilada en la plataforma para realizar seguimiento en tiempo real de ciertos eventos meteorológicos. En concreto, seguimiento de heladas, de precipitación y

de nevadas. Este tipo de eventos pueden llegar a suponer un riesgo importante para la sociedad y es vital minimizar su impacto. Para evaluar la hipótesis se presenta un episodio, ocurrido en la mitad norte de la Península Ibérica el 31 de enero de 2015, donde se registraron bajas temperaturas y precipitación.

Para el seguimiento de cada uno de los eventos mencionados, se utilizó la información de la plataforma GFA- \mathcal{M}^3 :

- Con respecto las heladas, el factor determinante es la temperatura. Gracias a los datos de temperatura contenidos en la plataforma se realizó un seguimiento cuantitativo de los valores de temperatura durante el episodio. Además, con esta información, se generaron gráficas similares a las mostradas en la figura 5.1, en el apartado 5.1, permitiendo un seguimiento cualitativo del evento en tiempo y espacio.
- La intensidad de agua precipitada durante el episodio, se midió cuantitativamente gracias a las estaciones meteorológicas cuyos datos recopila la plataforma. Estos datos permitieron un seguimiento del evento gracias a gráficas como las mostradas en la figura 5.2. Estas gráficas añaden un factor de calidad, ya que permitieron observar intuitivamente una precipitación leve, pero constante en el tiempo, que a la postre tuvo graves consecuencias.
- Para el seguimiento de nevadas, la combinación de dos variables, temperatura y precipitación, son el factor clave. Como ya se ha mencionado, gracias a los datos de temperatura y agua precipitada se realizó un seguimiento cuantitativo de estas variables durante todo el episodio. Además, la plataforma permite combinar ambas variables y producir gráficas como las mostradas en la figura 5.3, que permiten conocer, en tiempo real, las zonas donde se da precipitación en forma de nieve.

6.1.2. Evaluación de la hipótesis 2

La hipótesis 2 afirma que un SGDM permite obtener información para utilizarla después como datos de asimilación en modelos NWP. La métrica 2 propone validar esta hipótesis evaluando la utilidad de la información de la plataforma GFA- \mathcal{M}^3 como fuente de datos de asimilación para modelos NWP.

Para validar la hipótesis, se plantea la posibilidad de utilizar los datos de observación contenidos en la plataforma para ayudar a establecer las condiciones iniciales de un modelo NWP, en concreto el modelo WRF. Para demostrarlo, se plantea un método para extraer información de la plataforma y formatearla de acuerdo a las necesidades de WRFDA, el sistema de asimilación de datos del modelo WRF.

Utilizando los datos de la plataforma GFA- \mathcal{M}^3 , pueden obtenerse ficheros en formato LITTLE_R para su utilización como datos de asimilación con WRFDA.

El listado 5.2, en el apartado 5.2, muestra uno de estos ficheros obtenidos a partir de los datos del listado 5.1.

6.1.3. Evaluación de la hipótesis 3

La hipótesis 3 afirma que un SGDM permite para calibrar y validar algoritmos dentro del ámbito de la meteorología y la climatología. La métrica 3 propone validar esta hipótesis evaluando algunas contribuciones del GFA realizadas a partir de la información contenida en la plataforma GFA- \mathcal{M}^3 .

Se presentan tres herramientas que utilizan la información de la plataforma para validar algunos de los algoritmos utilizados por el GFA durante sus investigaciones, en concreto: *WRF precipitation evaluation tool*, *The Satellite Precipitation Evaluation Application*, y *Hail Precipitation Evaluation Application*.

La primera, permite evaluar los datos de precipitación obtenidos por el modelo WRF, comparándolos con datos de precipitación obtenidos a partir de las estaciones meteorológicas en superficie cuyos datos se almacenan en la plataforma GFA- \mathcal{M}^3 . La figura 5.4, en el apartado 5.3, muestra esta comparativa.

La segunda, *The Satellite Precipitation Evaluation Application*, evalúa los resultados obtenidos por el algoritmo MPE, que estima la precipitación a partir de datos de satélite, comparándolos también con los datos de precipitación obtenidos de estaciones meteorológicas. La figura 5.5 muestra los resultados del algoritmo MPE en un episodio concreto. Los datos de precipitación obtenidos a partir de estaciones meteorológicas en ese mismo episodio se muestran en la figura 5.6. La figura 5.7 permite apreciar las zonas donde el algoritmo MPE estima correctamente la precipitación comparando ambos resultados.

Por último, *Hail Precipitation Evaluation Application*, utiliza datos de radar para evaluar los resultados del algoritmo que permite identificar tormentas con presencia de granizo a partir de información de satélite. La figura 5.8 permite comparar información de radar, con los resultados del algoritmo propuesto por el GFA.

En resumen, las comparativas anteriores permiten evaluar las estimaciones de precipitación del modelo WRF y del algoritmo MPE, así como la detección de tormentas con presencia de granizo del algoritmo propuesto por el GFA.

6.1.4. Evaluación de la hipótesis 4

La hipótesis 4 afirma que un SGDM favorece el aprendizaje en el ámbito de la meteorología y la climatología. Para validar la hipótesis, se expone el uso que se ha hecho de la plataforma GFA- \mathcal{M}^3 durante la impartición de dos asignaturas (apartado 5.5): *Meteorología y Climatología y Observación y Vigilancia Atmosférica*.

Los resultados obtenidos y la satisfacción de los estudiantes han sido positivos en ambos casos de acuerdo a las encuestas generales que realiza la Universidad de León. Una línea de trabajo futura, es el planteamiento de una métrica que permita evaluar estos resultados (encuestas concretas, comparativa de resultados, etc.).

6.1.5. Desarrollo de aplicaciones

En la introducción, también se plantea la posibilidad de utilizar la información de un SGDM para desarrollar aplicaciones. Se presentan tres aplicaciones, presentadas en el apartado 5.4, que utilizan la información de la plataforma GFA- \mathcal{M}^3 . La primera herramienta, permite predecir la aparición tormentas severas con presencia de granizo. Se utilizó en Merino *et al.* (2013) para identificar este tipo de tormentas a partir de las variables obtenidas en un pronóstico del modelo WRF.

La segunda, PyHDT, calcula, en tiempo real, la probabilidad de que una tormenta presente precipitación en forma de granizo. PyHDT es una implementación del algoritmo propuesto en Merino *et al.* (2014c) para calcular la probabilidad de granizo a partir de información de satélite.

Por último, el sistema de información meteorológica para un vehículo, sujeto a la patente ES2462552B1, muestra información meteorológica y alertas en tiempo real con el fin de proporcionar la ruta más segura posible.

6.2. Conclusiones

En relación a la hipótesis 1, la información de las figuras 5.1, 5.2 y 5.3, presentadas en apartado 5.1, permitió monitorizar la evolución de los eventos mencionados durante el periodo de tiempo establecido, dejando patente la validez de la hipótesis. Podemos afirmar por tanto, que la plataforma GFA- \mathcal{M}^3 facilita el seguimiento de eventos meteorológicos en tiempo real en la mitad norte de la Península Ibérica. La misma idea es extrapolable a cualquier SGDM lo que permite afirmar la validez de la hipótesis 1.

Con respecto a la segunda hipótesis, en el apartado 5.2 se presenta un método para utilizar la información de la plataforma GFA- \mathcal{M}^3 para obtener datos de observación que puedan utilizarse con WRFDA. El listado 5.2 muestra un fichero con datos de observación en formato LITTLE_R, obtenido a partir de datos de observación de la plataforma, válido para ser utilizado por WRFDA. De nuevo, la misma idea es aplicable a otros SGDM. Por tanto, podemos afirmar que los datos de observación de un SGDM pueden utilizarse para obtener información de asimilación para modelos NWP, dando validez a la hipótesis 2.

En el apartado 5.3, se presentan tres ejemplos de como la información de la plataforma GFA- \mathcal{M}^3 se ha utilizado para validar las estimaciones de precipitación

pronosticadas por el modelo WRF y por el algoritmo MPE, así como la probabilidad de granizo calculada según el algoritmo desarrollado por el GFA. Estos tres ejemplos permiten afirmar que la plataforma GFA- \mathcal{M}^3 facilita la validación y calibración de algoritmos en el ámbito de la meteorología y la climatología, aplicados a la mitad norte de la Península Ibérica. Una vez más, esto es extrapolable a cualquier SGDM, lo que da validez a la hipótesis 3.

Por último, en relación a la hipótesis 4, la experiencia llevada a cabo en las asignaturas de Meteorología y Climatología y Observación y Vigilancia Atmosférica, permite afirmar que la información de la plataforma GFA- \mathcal{M}^3 sirve como apoyo a la educación académica.

Este trabajo también plantea la posibilidad de que un SGDM sirva como fuente de datos meteorológicos para otras aplicaciones desarrolladas por terceros. Para demostrarlo, se han presentado tres herramientas que hacen uso de la información meteorológica de la plataforma GFA- \mathcal{M}^3 : una que permite predecir tormentas severas con presencia de granizo a partir de la información del modelo WRF; otra herramienta para calcular la probabilidad de granizo a partir de datos de satélite; y un sistema de información meteorológica para un vehículo, esta última sujeta a la patente ES2462552B1. Por lo anterior, podemos afirmar que un SGDM favorece el desarrollo de aplicaciones que utilicen datos meteorológicos.

6.3. Trabajos futuros

Establecidas las bases de un SGDM, formalizado el modelo \mathcal{M}^3 e implementado la plataforma GFA- \mathcal{M}^3 , hay varias líneas de trabajo que se pueden seguir a partir de esta tesis. La primera sería incorporar nuevas fuentes de datos, lo cual no solamente supondría disponer de un volumen de datos mayor, sino que además podría permitir incrementar la cobertura espacial del SGDM. Incorporar nuevas fuentes implica dos cosas: primero, contactar con otros organismos y particulares, en el caso de que la nueva fuente de datos pertenezca a terceros, o bien configurar y poner en producción el nuevo sistema de observación, en el caso de tratarse de instrumentación propia; después, habrá que analizar el formato de los datos de la nueva fuente e implementar las funcionalidades necesarias si procede. En estos momentos se trabaja por incorporar los datos de un detector de rayos emplazado en las instalaciones del GFA, en el IMARENABIO de la Universidad de León.

Además de incluir nuevas fuentes y formatos de entrada en el repositorio, es interesante ampliar el número de formatos en los cuales pueden obtenerse datos directamente de la plataforma, incluyendo por ejemplo: HRIT/LRIT, MDV, GML, JSON, LITTLE_R, etc.

Otra de las líneas de investigación se planteó en el apartado 2.3.1, en relación a la definición del repositorio. En la plataforma GFA- \mathcal{M}^3 se ha optado por un esquema relacional clásico, por los motivos explicados en el capítulo 2, sin

embargo, hay otras alternativas no estructuradas y/o distribuidas con características diferentes que pueden resultar útiles en un SGDM. Estas soluciones pueden aportar funcionalidades no contempladas en la solución actual: integración de información no estructurada, repositorio distribuido, etc. En cualquier caso, sería necesaria una evaluación previa del rendimiento.

Por otro lado, dado que un SGDM está expuesto, desde el punto de vista de las comunicaciones, habría que revisar cuestiones relacionadas con la ciberseguridad, en previsión de un posible ataque. Esto es de especial relevancia si se utilizan los datos del SGDM para la gestión de riesgos meteorológicos.

Por último, otra de las líneas de trabajo futuro es la mejora de la presentación y el acceso de los datos. Sería conveniente construir un portal de acceso a la información (Open Data) que simplificara el acceso a los datos. Actualmente solamente puede hacerse a través del sitio web del GFA completando el formulario web mostrado en la sección 4.1.3.

6.4. Publicaciones

El presente trabajo a dado lugar a una serie de publicaciones en revistas científicas y congresos internacionales que se detallan a continuación.

6.4.1. Publicaciones

- Merino, A., García-Ortega, E., López, L., Sánchez, J. L., y Guerrero Higuera, A. M. (2013). Synoptic environment, mesoscale configurations and forecast parameters for hailstorms in Southwestern Europe. *Atmospheric Research*, 122:183–198

6.4.2. Participación en congresos internacionales

- Merino, A., Campos, L., López, L., García-Ortega, E., Sánchez, J. L., Marcos, J. L., y Guerrero-Higuera, A. M. (2012a). Snow in Castile-León: trends and variability. En *European Geosciences Union, General Assembly*. Geophysical Research Abstracts, Vol. 14, EGU2012-11363
- Merino, A., García, C., López, L., García-Ortega, E., Sánchez, J. L., Marcos, J. L., y Guerrero-Higuera, A. M. (2012b). Trends in hailstorm frequency in Castile-León (Spain). En *European Geosciences Union, General Assembly*. Geophysical Research Abstracts, Vol. 14, EGU2012-11273-1
- Merino, A., Guerrero-Higuera, A. M., de Galisteo, J. O., López, L., García-Ortega, E., Nafría, D. A., y Sánchez, J. L. (2012c). Integrated Meteorological Observation Network in Castile-León (Spain). En *European Geosciences Union, General Assembly*. Geophysical Research Abstracts, Vol. 14, EGU2012-11273-1

- cies Union, General Assembly*. Geophysical Research Abstracts, Vol. 14, EGU2012-11433
- Guerrero-Higueras, A. M., García-Ortega, E., Matellán-Olivera, V., y Sánchez, J. L. (2012). Procesamiento paralelo de los pronósticos meteorológicos del modelo WRF mediante NCL. En *Actas XXII Jornadas de Paralelismo (JP2012)*. ISBN: 978-84-694-1791-1, pp. 55-60
 - López, L., Hermida, L., Merino, A., Campos, L., Sánchez, J. L., García-Ortega, E., Guerrero-Higueras, A. M., y Acosta, S. (2012). Riesgo de nevadas en Castilla y León: tendencia climática vs. percepción social. En *VIII CONGRESO Internacional AEC: Cambio climático. Extremos e impactos*.
 - Guerrero-Higueras, A. M., López, L., Merino, A., Sánchez, J. L., Matía, P., Lorente, J. M., Hermida, L., Nafría, D., de Galisteo, J. P. O., Marcos, J. L., y García-Ortega, E. (2013b). Nowcasting of meteorological risks during the winter season using the “Integrated Meteorological Observation Network in Castilla y León, (Spain)”. En *European Geosciences Union, General Assembly*. Geophysical Research Abstracts, Vol. 15, EGU2013-10237
 - López, L., Guerrero-Higueras, A. M., Sánchez, J. L., Matía, P., de Galisteo, J. P. O., Rodríguez, V., Lorente, J. M., Merino, A., Hermida, L., García-Ortega, E., y Fernández-Manso, O. (2013). Evaluation of meteorological alert chain in Castilla y León (Spain): How can the meteorological risk managers help researchers? En *European Geosciences Union, General Assembly*. Geophysical Research Abstracts, Vol. 15, EGU2013-10028
 - Guerrero-Higueras, A. M., Merino-Suances, A., López, L., y Sánchez, J. L. (2013d). Identification of summer hailstorm from MSG data using Python. En *Third Symposium on Advances in Modeling and Analysis Using Python. AMS 93rd annual meeting 2013*.
 - Guerrero-Higueras, A. M., García-Ortega, E., y Sánchez, J. L. (2013a). Schedule WRF executions in parallel environments using Python. En *Third Symposium on Advances in Modeling and Analysis Using Python. AMS 93rd annual meeting 2013*.
 - Guerrero-Higueras, A. M., Merino, A., López, L., Sánchez, J. L., Matellán-Olivera, V., García-Ortega, E., Marcos, J. L., Gascón, E., y Hermida, L. (2013c). GFA meteorological and climatological open data for Castilla y León. En *EUMETSAT Meteorological Satellite Conference & 19th American Meteorological Society AMS Satellite Meteorology, Oceanography, and Climatology Conference*.
 - Guerrero-Higueras, A. M., López, L., Matellán-Olivera, V., Merino, A., y Sánchez, J. L. (2014). Atmospheric Physics Group Open Data (GFA Open Data): Meteorological data and tools for learning analytics. En *Proceedings*

TEEM'14, Second International Conference on Technological Ecosystems for Enhancing Multiculturality. ISBN: 978-1-4503-2896-8, pp. 415-420

- Merino, A., Guerrero-Higueras, A. M., López, L., Gascón, E., Sánchez, J. L., Lorente, J. M., Marcos, J. L., Matía, P., de Galisteo, J. P. O., Nafría, D., Fernández-González, S., Weigand, R., Hermida, L., y García-Ortega, E. (2014b). Development of tools for evaluating rainfall estimation models in real-time using the Integrated Meteorological Observation Network in Castilla y León (Spain). En *European Geosciences Union, General Assembly*. Geophysical Research Abstracts, Vol. 16, EGU2014-10234
- López, L., Sánchez, J. L., García-Ortega, E., Merino, A., Gascón, E., Fernández González, S., Hermida, L., Guerrero-Higueras, A. M., Weigand, R., y Marcos, J. L. (2015). Sistema de información sobre riesgos meteorológicos para un vehículo: predicción de rutas seguras y sistema de aviso entre usuarios y gestores de riesgo. En *XV Congreso Español Sistemas Inteligentes de transporte*.
- López, L., Merino, A., Guerrero-Higueras, A. M., García-Ortega, E., Gascón, E., Fernández González, S., Hermida, L., Sánchez, J. L., y Marcos, J. L. (2016). Nowcasting de riesgos agrometeorológicos. En *II Simposio Nacional de Ingeniería Hortícola. Automatización y TICs en Agricultura*.

6.4.3. Otras contribuciones

Además de las contribuciones anteriores, en el apéndice A se enumeran una serie de publicaciones y proyectos para los cuales fueron utilizados datos de la plataforma GFA- \mathcal{M}^3 .

Apéndice A

Otras contribuciones

Las publicaciones y proyectos se presentan en este capítulo guardan relación con la plataforma GFA- \mathcal{M}^3 , el SGDM descrito en el capítulo 4.

A.1. Publicaciones

Durante las investigaciones que dieron lugar a las publicaciones enumeradas a continuación, en las cuales el autor no es firmante, se utilizó información de la plataforma GFA- \mathcal{M}^3 :

- García-Ortega, E., Merino, A., López, L., y Sánchez, J. L. (2012). Role of mesoscale factors at the onset of deep convection on hailstorm days and their relation to the synoptic patterns. *Atmospheric Research*, 114-115:91–106
- Merino, A., López, L., Sánchez, J. L., García-Ortega, E., Cattani, E., y Levizzani, V. (2014c). Daytime identification of summer hailstorm cells from MSG data. *Natural Hazards and Earth System Sciences*, 14(4):1017–1033
- Merino, A., Fernández-González, S., Hermida, L., López, L., Sánchez, J. L., García-Ortega, E., y Gascón, E. (2014a). Snowfall in the northwest Iberian Peninsula: Synoptic circulation patterns and their influence on snow day trends. *Scientific World Journal*, 2014
- Fernández-González, S., Valero, F., Sánchez, J. L., Gascón, E., López, L., García-Ortega, E., y Merino, A. (2015). Numerical simulations of snowfall events: sensitivity analysis of physical parameterizations. *Journal of Geophysical Research: Atmospheres*, 120(19):10130–10148

- Gascón, E., Merino, A., Sánchez, J. L., Fernández-González, S., García-Ortega, E., López, L., y Hermida, L. (2015). Spatial distribution of thermodynamic conditions of severe storms in southwestern Europe. *Atmospheric Research*, 164-165:194–209
- Merino, A., Fernández-Vaquero, M., López, L., Fernández-González, S., Hermida, L., Sánchez, J. L., García-Ortega, E., y Gascón, E. (2016). Large-scale patterns of daily precipitation extremes on the Iberian Peninsula. *International Journal of Climatology*, 36(11):3873–3891
- García-Ortega, E., Lorenzana, J., Merino, A., Fernández-González, S., López, L., y Sánchez, J. L. (2017). Performance of multi-physics ensembles in convective precipitation events over northeastern Spain. *Atmospheric Research*, 190:55–67

A.2. Proyectos

En el marco de los siguientes proyectos se identificaron algunas de las fuentes de información de la plataforma GFA- \mathcal{M}^3 :

- Optimización de las herramientas de identificación y seguimiento de riesgos meteorológicos mediante el satélite Meteosat Second Generation (MSG) a través de la implantación de la red de observadores meteorológicos de Castilla y León (LE220A11-2). Financiado por la Junta de Castilla y León. Investigador principal: Laura López Campano. Cuantía: 26.000 €. Duración: 3 años. Fecha de inicio: 1-ene-2011, fecha final: 31-oct-2013.
- Mejora de la predicción de precipitación mediante el modelo WRF y el simulador de satélite GPM de la NASA (LE176A11-2). Financiado por la Junta de Castilla y León. Investigador principal: Eduardo García Ortega. Cuantía: 30.000 €. Duración: 3 años. Fecha de inicio: 1-ene-2011, fecha final: 31-dic-2013.
- Plataforma Micrometeo.com (IPT-3100002010-22). Financiado por el Ministerio de Ciencia e Innovación, programa INNPACTO. Investigador principal: José Luis Sánchez Gómez. Cuantía: 600.000 €. Duración: 3 años y medio. Fecha de inicio: 22-jun-2019, fecha final: 31-dic-2013.
- Teledetección aplicada a la elaboración de mapas de peligrosidad de granizo en tiempo real y mapas de daños en cultivos e infraestructuras. Financiado por la Fundación MAPFRE. Investigador Principal: Laura López Campano.
- Apoyo tecnológico y de servicios, asesoría técnica y científica relacionado con la previsión, tipo e incidencia de las precipitaciones (2008/00261/001). Financiado por el Gobierno de Aragón. Cuantía: 60.000 €. Duración: 4 años. Fecha de inicio: 1-dic-2008, fecha final: 31-dic-2013.

Bibliografía

- ANECA (2004). *Libro blanco. Título de grado en Ciencias Ambientales*. Agencia Nacional de Evaluación de la Calidad y Acreditación.
- Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., e Ives, Z. (2007). Dbpedia: A nucleus for a web of open data. En *The semantic web*, pp. 722–735. Springer.
- Bailey, R., Gronell, A., Phillips, H., Tanner, E., , y G.Meyers (1994). CSIRO Marine Laboratories Report 221: Quality Control Cookbook for XBT Data. [Online; accessed 2017-02-02].
- Bauer, P., Thorpe, A., y Brunet, G. (2015). The quiet revolution of numerical weather prediction. *Nature*, 525(7567):47–55.
- Bolton, D. (1980). The computation of equivalent potential temperature. *Monthly weather review*, 108(7):1046–1053.
- Cattell, R. (2011). Scalable SQL and NoSQL data stores. *Acm Sigmod Record*, 39(4):12–27.
- Chen, M., Xie, P., Janowiak, J. E., y Arkin, P. A. (2002). Global land precipitation: A 50-yr monthly analysis based on gauge observations. *Journal of Hydrometeorology*, 3(3):249–266.
- Chen, P. (1976). The Entity-Relationship Model: Toward a Unified View of Data. *ACM Transactions on Database Systems*, 1:9–36.
- Chleusener, R. A. y Jennings, P. C. (1960). An energy method for relative estimates of hail intensit. *Bull. Am. Meteorol. Soc.*, 41:372–376.
- Christensen, O. B., Drews, M., Christensen, J. H., Dethloff, K., Ketelsen, K., Hebestadt, I., y Rinke, A. (2007). The HIRHAM Regional Climate Model. Version 5 (beta). Technical report, Danish Climate Centre, Danish Meteorological Institute.

- Chuvieco, E. (2016). *Fundamentals of Satellite Remote Sensing: An Environmental Approach*. CRC press.
- Codd, E. F. (1970). A relational model of data for large shared data banks. *Communications of the ACM*, 13(6):377–387.
- Coiffier, J. (2011). *Fundamentals of numerical weather prediction*. Cambridge University Press.
- Couclelis, H. (1992). People manipulate objects (but cultivate fields): beyond the raster-vector debate in GIS. En *Theories and methods of spatio-temporal reasoning in geographic space*, pp. 65–77. Springer.
- Daley, R. (1993). *Atmospheric data analysis*. Nmero 2. Cambridge university press.
- Davies, J. M. (1993). Hourly Helicity, Instability, and EHI in Forecasting Supercell Tornadoes. En *17th Conf. on Severe Local Storm*, pp. 107–111. Amer. Meteor. Soc.
- Dessens, J., Berthet, C., y Sanchez, J. L. (2015). Change in hailstone size distributions with an increase in the melting level height. *Atmospheric Research*, 158:245–253.
- Dixon, M. (2006). MDV Format, Interface Control Document (ICD). Technical report, National Center for Atmospheric Research (NCAR), Research Applications Laboratory (RAL).
- Dixon, M. y Wiener, G. (1993). TITAN: thunderstorm identification, tracking, analysis, and nowcasting - a radar-based methodology. *Journal of Atmospheric & Oceanic Technology*, 10(6):785–797.
- Droegemeier, K. K., Lazarus, S. M., y Davies-Jones, R. (1993). The influence of helicity on numerically simulated convective storms. *Monthly Weather Review*, 121(7):2005–2029.
- Dudhia, J. (1993). A nonhydrostatic version of the Penn State/NCAR mesoscale model: validation tests and simulation of an Atlantic cyclone and cold front. *Mon. Weather Rev.*, 121:1493–1513.
- Dudhia, J., Gill, D., Manning, K., Wang, W., y Bruyere, C. (2015). PSU/NCAR Mesoscale Modeling System Tutorial Class Notes and Users' Guide (MM5 Modeling System Version 3). <http://www2.mmm.ucar.edu/mm5/documents/tutorial-v3-notes.html>. [Online; accessed 25-November-2016].

- Epstein, E. S. (1969). Stochastic dynamic prediction 1. *Tellus*, 21(6):739–759.
- Fernández-González, S., Valero, F., Sánchez, J. L., Gascón, E., López, L., García-Ortega, E., y Merino, A. (2015). Numerical simulations of snowfall events: sensitivity analysis of physical parameterizations. *Journal of Geophysical Research: Atmospheres*, 120(19):10130–10148.
- Fukao, S., Hamazu, K., y Doviak, R. J. (2014). *Radar for meteorological and atmospheric observations*. Springer.
- García-Ortega, E., Lorenzana, J., Merino, A., Fernández-González, S., López, L., y Sánchez, J. L. (2017). Performance of multi-physics ensembles in convective precipitation events over northeastern Spain. *Atmospheric Research*, 190:55–67.
- García-Ortega, E., Merino, A., López, L., y Sánchez, J. L. (2012). Role of mesoscale factors at the onset of deep convection on hailstorm days and their relation to the synoptic patterns. *Atmospheric Research*, 114-115:91–106.
- Gascón, E. (2015). *Analysis and forecasting of convective precipitation using WRF mesoscale and multichannel microwave radiometer*. Tesis doctoral, University of León.
- Gascón, E., Merino, A., Sánchez, J. L., Fernández-González, S., García-Ortega, E., López, L., y Hermida, L. (2015). Spatial distribution of thermodynamic conditions of severe storms in southwestern Europe. *Atmospheric Research*, 164-165:194–209.
- Gil, B. (2009). *Modelling of hail precipitation*. Tesis doctoral, University of León.
- Goodchild, M. F. y Kemp, K. K. (1990). *NCGIA Core Curriculum: Introduction to GIS*, volumen 1. National Center for Geographic Information and Analysis, University of California at Santa Barbara.
- Grell, G., Dudhia, J., y Stauffer, D. (1995). A description of the fifth-generation Penn State/NCAR Mesoscale Model (MM5). Technical report, NCAR Tech. Note NCAR/TN-398+STR, 117pp.
- Guerrero-Higueras, A. M., García-Ortega, E., Matellán-Olivera, V., y Sánchez, J. L. (2012). Procesamiento paralelo de los pronósticos meteorológicos del modelo WRF mediante NCL. En *Actas XXII Jornadas de Paralelismo (JP2012)*. ISBN: 978-84-694-1791-1, pp. 55-60.

- Guerrero-Higueras, A. M., García-Ortega, E., y Sánchez, J. L. (2013a). Schedule WRF executions in parallel environments using Python. En *Third Symposium on Advances in Modeling and Analysis Using Python. AMS 93rd annual meeting 2013*.
- Guerrero-Higueras, A. M., López, L., Matellán-Olivera, V., Merino, A., y Sánchez, J. L. (2014). Atmospheric Physics Group Open Data (GFA Open Data): Meteorological data and tools for learning analytics. En *Proceedings TEEM'14, Second International Conference on Technological Ecosystems for Enhancing Multiculturality*. ISBN: 978-1-4503-2896-8, pp. 415-420.
- Guerrero-Higueras, A. M., López, L., Merino, A., Sánchez, J. L., Matía, P., Lorente, J. M., Hermida, L., Nafría, D., de Galisteo, J. P. O., Marcos, J. L., y García-Ortega, E. (2013b). Nowcasting of meteorological risks during the winter season using the “Integrated Meteorological Observation Network in Castilla y León, (Spain)”. En *European Geosciences Union, General Assembly. Geophysical Research Abstracts*, Vol. 15, EGU2013-10237.
- Guerrero-Higueras, A. M., Merino, A., López, L., Sánchez, J. L., Matellán-Olivera, V., García-Ortega, E., Marcos, J. L., Gascón, E., y Hermida, L. (2013c). GFA meteorological and climatological open data for Castilla y León. En *EUMETSAT Meteorological Satellite Conference & 19th American Meteorological Society AMS Satellite Meteorology, Oceanography, and Climatology Conference*.
- Guerrero-Higueras, A. M., Merino-Suances, A., López, L., y Sánchez, J. L. (2013d). Identification of summer hailstorm from MSG data using Python. En *Third Symposium on Advances in Modeling and Analysis Using Python. AMS 93rd annual meeting 2013*.
- Han, J., Haihong, E., Le, G., y Du, J. (2011). Survey on NoSQL database. En *Pervasive computing and applications (ICPCA), 2011 6th international conference on*, pp. 363–366. IEEE.
- Hannesen, R., Dotzek, N., y Handwerker, J. (2000). Radar analysis of a tornado over hilly terrain on 23 July 1996. *Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere*, 25(10-12):1079–1084.
- Hart, J. y Korotky, W. (1991). *The SHARP Workstation_v 1.50: A Skew T/Hodograph Analysis and Research Program for the IBM and Compatible PC: User's Manual*. NOAA/NWS Forecast Office, Charleston, WV.
- Heinemann, T., Lattanzio, A., y Roveda, F. (2002). The Eumetsat multi-sensor precipitation estimate (MPE). En *Second International Precipitation Working*

- group (IPWG) Meeting*. Geophysical Research Abstracts, Vol. 16, EGU2014-10234.
- Huntrieser, H., Schiesser, H. H., Schmid, W., y Waldvogel, A. (1997). Comparison of traditional and newly developed thunderstorm indices for Switzerland. *Weather and Forecasting*, 12(1):108–123.
- Huuskonen, A., Saltikoff, E., y Holleman, I. (2014). The operational weather radar network in europe. *Bulletin of the American Meteorological Society*, 95(6):897–907.
- J. Tapiador, F., Turk, F. J., Petersen, W., Hou, A. Y., García-Ortega, E., Machado, L. A. T., Angelis, C. F., Salio, P., Kidd, C., Huffman, G. J., y de Castro, M. (2012). Global precipitation measurement: Methods, datasets and applications. *Atmospheric Research*, 104-105:70–97.
- Källén, E. (1996). *Hirlam documentation manual: System 2.5*. Swedish Meteorological and Hydrological Institute.
- Kalnay, E. (2003). *Atmospheric modeling, data assimilation and predictability*. Cambridge university press.
- Kalnay, E., Kanamitsu, M., y Baker, W. E. (1990). Global numerical weather prediction at the National Meteorological Center. *Bulletin of the American Meteorological Society*, 71(10):1410–1428.
- Kaltenböck, R. (2004). The outbreak of severe storms along convergence lines northeast of the Alps. Case study of the 3 August 2001 mesoscale convective system with a pronounced bow echo. *Atmospheric Research*, 70(1):55–75.
- Kaltenböck, R., Diendorfer, G., y Dotzek, N. (2009). Evaluation of thunderstorm indices from ECMWF analyses, lightning data and severe storm reports. *Atmospheric Research*, 93(1-3):381–396.
- Kanamitsu, M. (1989). Description of the NMC global data assimilation and forecast system. *Weather and Forecasting*, 4(3):335–342.
- Kanamitsu, M., Alpert, J. C., Campana, K. A., Caplan, P. M., Deaven, D. G., Iredell, M., Katz, B., Pan, H.-L., Sela, J., y White, G. H. (1991). Recent changes implemented into the global forecast system at NMC. *Weather and Forecasting*, 6(3):425–435.
- Lafore, J. P., Stein, J., Asencio, N., Bougeault, P., Ducrocq, V., Duron, J., Fischer, C., Hérel, P., Mascart, P., Masson, V., Pinty, J. P., Redelsperger, J. L.,

- Richard, E., , y de Arellano, J. V.-G. (1998). The Meso-NH Atmospheric Simulation System. Part I: Adiabatic formulation and control simulations. *Ann. Geophys.*, 16:90–109.
- Leith, C. E. (1974). Theoretical skill of Monte Carlo forecasts. *Monthly Weather Review*, 102(6):409–418.
- Levizzani, V., Bauer, P., y Turk, F. J. (2007). *Measuring precipitation from space: EURAINSAT and the future*, volumen 28. Springer Science & Business Media.
- Lezaun, M. (2002). Predicciones del tiempo y matemáticas. *Boletín de la Sociedad Española de Matemática Aplicada*, 22:59–98.
- Long, A. B., Matson, R. J., y Crow, E. L. (1980). The hailpad: Material, data reduction and calibration. *J. Appl. Meteor.*, 19:1300–1313.
- Lorenz, E. (1993). *The Essence of Chaos*. University of Washington Press.
- Lorenz, E. N. (1963). Deterministic nonperiodic flow. *Journal of the atmospheric sciences*, 20(2):130–141.
- López, L., Guerrero-Higueras, A. M., Sánchez, J. L., Matía, P., de Galisteo, J. P. O., Rodríguez, V., Lorente, J. M., Merino, A., Hermida, L., García-Ortega, E., y Fernández-Manso, O. (2013). Evaluation of meteorological alert chain in Castilla y León (Spain): How can the meteorological risk managers help researchers? En *European Geosciences Union, General Assembly*. Geophysical Research Abstracts, Vol. 15, EGU2013-10028.
- López, L., Hermida, L., Merino, A., Campos, L., Sánchez, J. L., García-Ortega, E., Guerrero-Higueras, A. M., y Acosta, S. (2012). Riesgo de nevadas en Castilla y León: tendencia climática vs. percepción social. En *VIII CONGRESO Internacional AEC: Cambio climático. Extremos e impactos*.
- López, L., Merino, A., Guerrero-Higueras, A. M., García-Ortega, E., Gascón, E., Fernández González, S., Hermida, L., Sánchez, J. L., y Marcos, J. L. (2016). Nowcasting de riesgos agrometeorológicos. En *II Simposio Nacional de Ingeniería Hortícola. Automatización y TICs en Agricultura*.
- López, L., Sánchez, J. L., García-Ortega, E., Merino, A., Gascón, E., Fernández González, S., Hermida, L., Guerrero-Higueras, A. M., Weigand, R., y Marcos, J. L. (2015). Sistema de información sobre riesgos meteorológicos para un vehículo: predicción de rutas seguras y sistema de aviso entre usuarios y gestores de riesgo. En *XV Congreso Español Sistemas Inteligentes de transporte*.
- Männik, A. y Room, R. (2001). *Nonhydrostatic adiabatic kernel for HIRLAM: part II: anelastic, hybrid-coordinate, explicit-Eulerian model*. SMHI.

- Marshall, J. L., Lee, J., Jung, J., Gregory, P., y Roux, B. (2013). The considerable impact of earth observations from space on numerical weather prediction. *AUSTRALIAN METEOROLOGICAL AND OCEANOGRAPHIC JOURNAL*, 63(4):497–500.
- Merino, A. (2013). *Analysis, Identification and Forecast of hail precipitation events in the Iberian Peninsula*. Tesis doctoral, University of León.
- Merino, A., Campos, L., López, L., García-Ortega, E., Sánchez, J. L., Marcos, J. L., y Guerrero-Higueras, A. M. (2012a). Snow in Castile-León: trends and variability. En *European Geosciences Union, General Assembly*. Geophysical Research Abstracts, Vol. 14, EGU2012-11363.
- Merino, A., Fernández-González, S., Hermida, L., López, L., Sánchez, J. L., García-Ortega, E., y Gascón, E. (2014a). Snowfall in the northwest Iberian Peninsula: Synoptic circulation patterns and their influence on snow day trends. *Scientific World Journal*, 2014.
- Merino, A., Fernández-Vaquero, M., López, L., Fernández-González, S., Hermida, L., Sánchez, J. L., García-Ortega, E., y Gascón, E. (2016). Large-scale patterns of daily precipitation extremes on the Iberian Peninsula. *International Journal of Climatology*, 36(11):3873–3891.
- Merino, A., García, C., López, L., García-Ortega, E., Sánchez, J. L., Marcos, J. L., y Guerrero-Higueras, A. M. (2012b). Trends in hailstorm frequency in Castile-León (Spain). En *European Geosciences Union, General Assembly*. Geophysical Research Abstracts, Vol. 14, EGU2012-11273-1.
- Merino, A., García-Ortega, E., López, L., Sánchez, J. L., y Guerrero Higueras, A. M. (2013). Synoptic environment, mesoscale configurations and forecast parameters for hailstorms in Southwestern Europe. *Atmospheric Research*, 122:183–198.
- Merino, A., Guerrero-Higueras, A. M., de Galisteo, J. O., López, L., García-Ortega, E., Nafría, D. A., y Sánchez, J. L. (2012c). Integrated Meteorological Observation Network in Castile-León (Spain). En *European Geosciences Union, General Assembly*. Geophysical Research Abstracts, Vol. 14, EGU2012-11433.
- Merino, A., Guerrero-Higueras, A. M., López, L., Gascón, E., Sánchez, J. L., Lorente, J. M., Marcos, J. L., Matía, P., de Galisteo, J. P. O., Nafría, D., Fernández-González, S., Weigand, R., Hermida, L., y García-Ortega, E. (2014b). Development of tools for evaluating rainfall estimation models in real-time using the Integrated Meteorological Observation Network in Castilla

- y León (Spain). En *European Geosciences Union, General Assembly*. Geophysical Research Abstracts, Vol. 16, EGU2014-10234.
- Merino, A., López, L., Sánchez, J. L., García-Ortega, E., Cattani, E., y Levizzani, V. (2014c). Daytime identification of summer hailstorm cells from MSG data. *Natural Hazards and Earth System Sciences*, 14(4):1017–1033.
- Molloy, J. C. (2011). The open knowledge foundation: open data means better science. *PLoS Biol*, 9(12):e1001195.
- Moncrieff, M. W. (1981). A theory of organized steady convection and its transport properties. *Quarterly Journal, Royal Meteorological Society*, 107(451):29–50.
- Mullen, S. L. y Baumhefner, D. P. (1988). Sensitivity of numerical simulations of explosive oceanic cyclogenesis to changes in physical parameterizations. *Monthly weather review*, 116(11):2289–2329.
- OGC (2008). OGC KML. Technical report, Open Geospatial Consortium, Inc. OGC 07-147r2.
- Plummer, N., Lipa, W., Palmer, S., Prank4, G., Shortridge, J., y Stuber, D. (2007). Directrices sobre la gestión de datos climáticos. Technical report, Organización Meteorológica Mundial. WMO/TD N° 1376.
- PraPro (2009). PraPro: Browsing tools for satellite data. [Online; accessed 2016-08-03].
- Rasmussen, E. y Blanchard, D. (1998). A baseline climatology of sounding-derived supercell and tornado forecast parameters. *Weather Forecast.*, 13:1148–1164.
- Regueiro, M. A., Viqueira, J. R. R., Cortizas, C., Díaz, P., Méndez, X., Touriño, J., Parapar, J., y Landeira, E. (2012). MeteoSIX: Difusión de datos meteorológicos y oceanográficos en MeteoGalicia. *Terceras Jornadas Ibéricas de las Infraestructuras de Datos Espaciales (JIIDE 2012)*, 10:17–2012.
- Rew, R., Davis, G., Emmerson, S., Davies, H., Hartnett, E., Heimbigner, D., y Fisher, W. (2015). *Network Common Data Form (NetCDF)*. UCAR/Unidata, Boulder, CO. <http://doi.org/10.5065/D6H70CW6>.
- Schmetz, J. y Menzel, W. P. (2015). A look at the evolution of meteorological satellites: Advancing capabilities and meeting user requirements. *Weather, Climate, and Society*, 7(4):309–320.

- Schmetz, J., Pili, P., Tjemkes, S., Just, D., Kerkmann, J., Rota, S., y Ratier, A. (2002). An introduction to meteosat second generation (MSG). *B. Am. Meteorol. Soc.*, 79:2457–2476.
- Shafranovich, Y. (2005). Common Format and MIME Type for Comma-Separated Values (CSV) File. Technical report, RFC Editor. RFC 4180, doi:10.17487/RFC4180.
- Showalter, A. K. (1953). A stability index for thunderstorm forecasting. *B. Am. Meteorol. Soc.*, pp. 250–252.
- Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Duda, M. G., Huang, X. Y., Wang, W., , y Power, J. G. (2008). A Description of the Advanced Research WRF Version 3. Technical report, NCAR Tech. Note NCAR/TN-475+STR, 113 pp. doi:10.5065/D68S4MVH.
- Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., y Wang, J. G. P. (2005). A description of the Advanced Research WRF Version 2. Technical report, NCAR Tech. Note NCAR/TN-468+STR, 88 pp. doi:10.5065/D6DZ069.
- Stensrud, D. J., Bao, J.-W., y Warner, T. T. (2000). Using initial condition and model physics perturbations in short-range ensemble simulations of mesoscale convective systems. *Monthly Weather Review*, 128(7):2077–2107.
- Stinson, B. (2001). *PostgreSQL Essential Reference*. New Riders Publishing, Thousand Oaks, CA, USA.
- Tapiador, F. J., Tao, W. K., Shi, J. J., Angelis, C. F., Martinez, M. A., Marcos, C., Rodriguez, A., y Hou, A. (2012). A comparison of perturbed initial conditions and multiphysics ensembles in a severe weather episode in Spain. *Journal of Applied Meteorology and Climatology*, 51(3):489–504.
- Vicente, G. A., Davenport, J. C., y Scofield, R. A. (2002). The role of orographic and parallax corrections on real time high resolution satellite rainfall rate distribution. *Int. J. Remote Sens.*, 203:221–330.
- Warner, T. T. (2010). *Numerical weather and climate prediction*. Cambridge University Press.
- WMO (2004). *Manual on Codes vols I.1 and I.2*. Geneva, Switzerland: World Meteorological Organization. WMO Publ. No 306, p 23.
- WMO (2011). MCH: Meteorological, Climatological and Hydrological Database Management System. [Online; accessed 2017-02-09].

- Wolf, R. (1999). LRIT/HRIT Global Specification. Technical report, Coordination Group for Meteorological Satellites. Doc. No. CGMS 03. Issue 2.6.
- Woods, A. (2005). *Medium-range weather prediction: The European approach*. Springer Science & Business Media.
- Zhang, D. L., and N. L. Seaman, H. R. C., Warner, T. T., y Fritsch, J. M. (1986). A two-way interactive nesting procedure with variable terrain resolution. *Mon. Weather Rev*, 114:1330–1339.

Acrónimos

AEMET Agencia Estatal de Meteorología.

ASCII American Standard Code for Information Interchange.

API Application Programming Interface.

BMP BitMaP.

EUMETSAT European Organisation for the Exploitation of Meteorological Satellites.

CAPE Convective Available Potential Energy.

CESGA Centro de Supercomputación de Galicia.

CGMS Coordination Group for Meteorological Satellites.

CHD Confederación Hidrográfica del Duero.

CHE Confederación Hidrográfica del Ebro.

CI Índice de inestabilidad convectiva.

CRIS Current Research Information System.

CSV Comma Separated Values.

DOM Document Object Model

DBMS Data Base Management System.

ECMWF European Center for Medium range Weather Forecasting.

EHI Energy–Helicity Index.

FCSCCL Fundación del Centro de Supercomputación de Castilla y León.

- FDDA** Four-Dimensional Data Assimilation.
- FTP** File Transfer Protocol.
- GFA** Grupo de Física de la Atmósfera.
- GFA- \mathcal{M}^3** Meteorological data Management Model del GFA.
- GFS** Global Forecasting System.
- GIF** Graphics Interchange Format.
- GIS** Geographic Information System.
- GML** Geography Markup Language.
- GOES** Geostationary Operational Environmental Satellite.
- GPM** Global Precipitation Measurement.
- GPS** Global Positioning Systems.
- GrADS** Grid Analysis and Display System.
- HIRLAM** High Resolution Limited Area Model.
- HRIT** High Rate Information Transmission.
- HRV** High-Resolution Visible.
- HTTP** Hypertext Transfer Protocol.
- IDS** Intrusion Detection System.
- IMARENABIO** Instituto de Medio Ambiente, Recursos Naturales y Biodiversidad.
- ITACyL** Instituto Tecnológico Agrario de Castilla y León.
- IWV** Integrated Water Vapor.
- JPEG** Joint Photographic Experts Group.
- JSON** JavaScript Object Notation.
- KML** Keyhole Markup Language.
- LAM** Limited Area Models.
- LRIT** Low Rate Information Transmission.

LWC Liquid Water Content.

LWP Liquid Water Path.

M³ Meteorological data Management Model.

MCH Meteorological, Climatological and Hydrological Database Management System.

MDV Meteorological Data Volume.

MINECO Ministerio de Economía, Industria y Competitividad.

MM5 Fifth-Generation NCAR/Penn State Mesoscale Model.

MMWR Multichannel MicroWave Radiometer.

MP3000A Microwave Profiles 3000A.

MPE Multi-sensor Precipitation Estimate.

MPI Message Passing Interface.

MSG Meteosat Second Generation.

NASA National Aeronautics and Space Administration.

NCAR National Center for Atmospheric Research.

NCL NCAR Command Language.

netCDF Network Common Data Form.

NOAA National Oceanic and Atmospheric Administration.

NoSQL Not Only SQL.

NWP Numerical weather prediction.

OGC Open Geospatial Consortium.

OpenMP Open Multi-Processing.

PHP Hypertext Preprocessor.

PNG Portable Network Graphics.

RAL Research Applications Laboratory.

RdOCyL Red de observadores de Castilla y León.

- SAIH** Sistema Automático de Información Hidrológica.
- SEVIRI** Spinning Enhanced Visible and InfraRed Imager.
- SGDM** Sistema de Gestión de Datos Meteorológicos.
- SGDC** Sistema de Gestión de Datos Climáticos.
- SGE** Sun Grid Engine.
- SI** Showalter Index.
- SiAR** Sistema de Información Agroclimática para el Regadío.
- SRH₀₋₃** Storm-relative helicity between 0 and 3 km.
- SSH** Secure SHell.
- SQL** Structured Query Language.
- TIFF** Tagged Image File Format.
- UCAR** University Corporation for Atmospheric Research.
- UDC** Universidad de La Coruña.
- USC** Universidad de Santiago de Compostela.
- UTM** Universal Transverse Mercator.
- WKT** Well Known Text.
- WMO** World Meteorological Organization.
- WVC** Water Vapor Content.
- WRF** Weather Research and Forecasting.
- WRFDA** Weather Research and Forecasting model data assimilation system.
- WPS** WRF Pre-processing System.
- WSDL** Web Services Description Language.
- WVFD** Divergencia del flujo de vapor de agua.
- XML** eXtensible Markup Language.