

PROMETEO: APLICACIÓN OPERATIVA DE *DOWNSCALING* ESTADÍSTICO PARA LA PREDICCIÓN DE FENÓMENOS LOCALES

Antonio S. COFIÑO*; Carmen SORDO**; José M. GUTIÉRREZ**;
Rafael CANO*** y Cristina PRIMO***

* *School of Computing and Mathematical Sciences, Liverpool John Moores University*

** *Dpto. de Matemática Aplicada y Ciencias de la Computación, Universidad de Cantabria*

*** *Centro Meteorológico Territorial de Cantabria y Asturias, Instituto Nacional de Meteorología*

RESUMEN

Esta comunicación, describe el prototipo desarrollado en colaboración con el Instituto Nacional de Meteorología para la predicción local basada en técnicas híbridas de *downscaling*, que combinan la información procedente de los modelos numéricos, con la información climatológica procedente de observaciones de fenómenos meteorológicos. El prototipo se encuentra funcionando de forma operativa, de forma que los usuarios puedan acceder diariamente a las predicciones ofrecidas, de una forma interactiva. Además el sistema cuenta con la posibilidad de acceder a predicciones pasadas y concretamente acceder a los distintos índices de validación como la frecuencia climatológica de eventos tales como, Precipitación, Nieblas, Tormentas, Rachas Máximas, Temperaturas etc. Estas predicciones son ofrecidas en la red secundaria de estaciones, gestionada por Instituto Nacional de Meteorología, a partir de las observaciones realizadas por colaboradores.

Palabras clave: *Downscaling*, predicción probabilística, aplicación web, predicción operativa, predicción local.

1. INTRODUCCIÓN

Los servicios meteorológicos utilizan modelos atmosféricos que simulan la dinámica de la atmósfera a gran escala, con una resolución equivalente a 50 o 100 km. Estos modelos permiten predecir con bastante pericia el comportamiento sinóptico de la atmósfera en el corto plazo (hasta 3 días de antelación). Sin embargo, estos modelos poseen menor pericia en la predicción de fenómenos que dependen de procesos físicos complejos, como la formación de nubes, evaporación, orografía, turbulencia, etc., los cuales ocurren a escalas menores que la resolución de los modelos y cuya física no es fácil de resolver. La predicción de fenómenos como la precipitación, nieblas, tormentas, etc, que dependen de estos procesos físicos, es un problema todavía a resolver, con una gran interés científico, social y económico.

En la literatura se ha propuesto diferentes métodos para abordar este problema:

- Técnicas de predicción estadística para series temporales, que se basan únicamente en la información procedente de las observaciones históricas.
- Técnicas de *downscaling* dinámico, que tienen como propósito aumentar la resolución de los modelos atmosféricos, anidándolos con modelos atmosféricos regionales o mesoescalares.
- Técnicas de *downscaling* estadístico, que son técnicas híbridas que combinan la salidas de los modelos numéricos con la información estadística (climatología) de las observaciones.

La predicción local con técnicas de aumento de resolución (*downscaling*) dinámica y estadística es una línea activa de investigación en la comunidad meteorológica. En COFIÑO (2004) se muestra como las técnicas híbridas son superiores a los otros métodos, y en concreto como algunas técnicas de agrupamiento pueden mejorar la pericia de eventos con una probabilidad climatológica baja (GUTIERREZ *et al.*, 2004a).

Todos estos resultados han permitido desarrollar un prototipo operativo denominado *Prometeo*, que permite organizar el flujo tanto de los datos como de los procesos que intervienen en los distintos experimentos de predicción local operativa. Como ejemplo global del sistema, que nos ayudará a entender su estructura, en la Sección 2 se describe la adaptación hecha en colaboración con el Instituto Nacional de Meteorología (INM) para la predicción local diaria en 2500 estaciones repartidas sobre toda España (versión experimental en <http://meteo.macc.unican.es/prometeo>). Esta aplicación genera diariamente predicciones operativas con los modelos que se utilizan en el INM (HIRLAM, con sus respectivas pasadas de la 00UTC y 12UTC con alcances de D+0 y D+1 y los alcances de D+2 y D+3 del modelo operativo del ECMWF cuya pasada es a las 12UTC). Con estas predicciones numéricas se realizan predicciones locales para las doce cuencas hidrográficas (diez en la Península, una para Baleares y otra para Canarias), cada una de las cuales tiene asociado su correspondiente patrón atmosférico. En total se realizan unas 162.500 predicciones diarias (5 salidas numéricas, 2.500 observatorios, y 13 predicciones para cada estación: 6 para precipitación y 7 meteoros). A la vista de la magnitud de los datos es evidente que para que el sistema funcione en modo operativo interactivamente es importante la organización y la estructura de los datos para un acceso rápido y eficiente.

Aunque el prototipo descrito en esta comunicación corresponde a la predicción a corto plazo, actualmente esta siendo extendido al medio plazo haciendo uso de la predicción por conjuntos, y llegando a la predicción mensual-estacional de anomalías de fenómenos meteorológicos (COFIÑO, 2004 y GUTIÉRREZ *et al.*, 2004b).

2. ESTRUCTURA DE LA APLICACIÓN. SISTEMA OPERATIVO EN EL INM

En esta sección se describe la estructura del sistema *PROMETEO* (PROnóstico METEOrológico) utilizando como ejemplo la versión configurada para la predicción local diaria en el INM. Se comienza describiendo la fase de preproceso y las tareas involucradas, y se finaliza con el sistema de consulta Web para predicciones operativas. Para evitar problemas de confusión en la descripción del sistema, usaremos únicamente las salidas del modelo operativo y el re-análisis ERA-15 del ECMWF. Además, la base de datos de observaciones estará limitada a aproximadamente 2.500 estaciones operativas de la red secundaria del INM. La figura 1 muestra la estructura global de la aplicación, destacando sus cuatro módulos (A)-(D). En las siguientes secciones se describen en detalle cada uno de estos módulos.

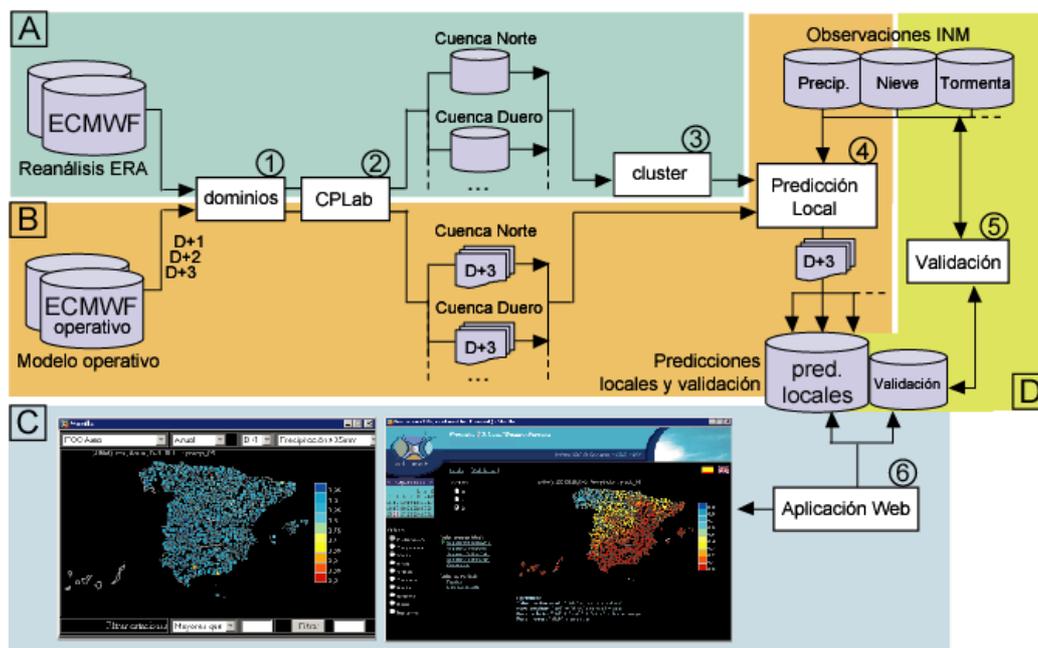


Fig. 1. Estructura global de la aplicación *PROMETEO* con las etapas: (A) Configuración, (B) producción operativa, (C) consulta a través de la Web, y (D) validación

2.1. Configuración e Inicialización

En esta sección se describe la fase inicial de configuración y adaptación del sistema a un entorno concreto (zona geográfica, variables, bases de datos, etc.). El sistema está basado en el método descrito en COFIÑO (2004) y, por tanto, a ella se remite al lector para consultar detalles sobre los algoritmos. En esta sección se describen los elementos que intervienen para configurar la aplicación para la predicción local en la red secundaria del INM a partir de las predicciones numéricas del ECMWF. En este módulo se utiliza la base de datos de re-análisis ERA-15 del ECMWF (<http://www.ecmwf.int/research/era/ERA-15/>), que comprende el período 1979-1994 (5.569 días en total). Primero, se define el patrón que caracterizará el estado de la atmósfera para el problema en cuestión (tarea 1 en la figura 2; a continuación en la tarea 2 se aplica el algoritmo de componentes principales (COFIÑO, 2004) para reducir la dimensión de los patrones y, finalmente, se aplica la técnica de agrupamiento apropiada sobre cada uno de los vectores de componentes principales obtenidos (tarea 3). Este módulo es el que contiene el núcleo del sistema y será necesario ejecutarlo cada vez que queramos para re-adaptarlo a un nuevo problema o a una nueva situación. Esta tarea se realizará una única vez, previamente a la fase operativa de la aplicación (la obtención diaria de predicciones).

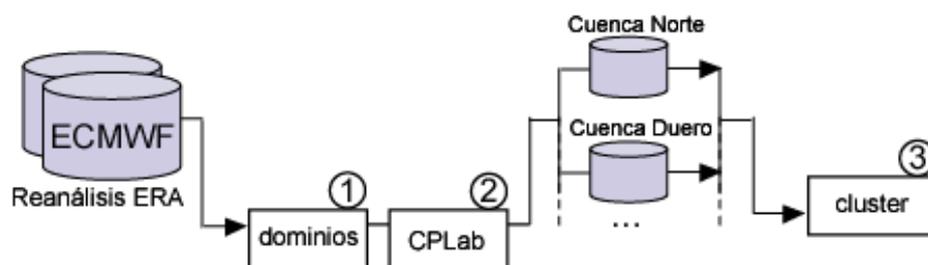


Fig. 2. Módulo de configuración e inicialización, con la definición del *patrón atmosférico* abarcando un dominio dado (tarea 1), la compresión de información aplicando componentes principales (2), y la ejecución de la técnica de agrupamiento apropiada (3)

La figura 2 muestra esquemáticamente este módulo. En esta figura podemos ver la base de datos de re-análisis y los 3 módulos siguientes:

1. *Dominios*. Es el encargado de extraer los datos necesarios para definir el patrón atmosférico a partir de la base de datos de re-análisis y aglutinar la información en forma vectorial. En COFIÑO (2004) se muestra que los mejores resultados de predicción local se obtienen con un patrón 4D de área limitada. Por ello, consideramos 12 dominios distintos (uno para cada cuenca) tal como se muestra en la Fig. 3. Para ello se ha de especificar región geográfica de interés (cuadrícula longitud-latitud, resolución y nodos a incluir), las variables, las horas de análisis, y los niveles de presión. Esta información se usa como entrada en el módulo “dominio”. Por ejemplo para las diferentes cuencas hidrográficas se ha definido un patrón atmosférico de 1° de resolución en longitud y latitud; tres niveles en altura 1000mb, 850mb, y 500mb; cinco horas de análisis 06, 12, 18, 24, y 0 UTC; y cinco variables Temperatura (T), geopotencial (Z), coordenadas del viento (U y V) y humedad relativa (H).
2. *CPLab*. Una vez se ha “filtrado” la base de datos de re-análisis extrayendo los patrones con el módulo *dominio*, se aplica un análisis de componentes principales con el objetivo de reducir la dimensión de los patrones. Dado que esta dimensión puede ser muy elevada y los patrones pueden contener mucha información redundante (debido a las grandes correlaciones espaciales que existen en los datos atmosféricos), es probable obtener matrices de covarianza mal condicionadas, que dificulten la aplicación de técnicas estándar. En este módulo se ha utilizado una técnica iterativa para calcular los primeros valores singulares. Antes de aplicar este análisis es necesario estandarizar los datos para evitar que las variables con mayor magnitud dominen al resto. Como resultado de este proceso obtenemos datos necesarios para poder aplicar la misma transformación (matriz de transformación, media y desviación de los datos) a las salidas de otros modelos numéricos, como por ejemplo el operativo del ECMWF, que se usa en la etapa de explotación de la aplicación.
3. *Cluster (k-medias/SOM)*. Este módulo es el encargado de realizar y almacenar los resultados de agrupar los datos de re-análisis después de haber sido filtrados y comprimidos. En este módulo se definen el número de grupos que vamos a realizar, y los parámetros de entrenamiento del algoritmo de agrupamiento (para *k-medias*); en el caso de utilizar una SOM también es necesario indicar la topología y el parámetro de vecindad. Debido a la exigencia

computacional de esta tarea, es necesario guardar los distintos agrupamientos realizados (centros, etc.), junto con información auxiliar necesaria par su posterior uso. Por otra parte, los problemas de inicialización en los algoritmos iterativos de agrupamiento, hacen conveniente realizar varios entrenamientos y evaluarlos posteriormente con el módulo de validación para contrastar la calidad de cada uno de ellos.

En la aplicación operativa en el INM, los procesos anteriores se han aplicado a cada una de las 12 cuencas hidrográficas, almacenando las CPs del re-análisis, la matriz de transformación, y los agrupamientos realizados.

Esta fase de configuración e inicialización, es la que exige una mayor capacidad de cómputo y de almacenamiento. Toda esta información almacenada, con el objetivo de hacer el sistema interactivo, es utilizada en el siguiente proceso de explotación operativa.

2.2. Explotación Operativa

Este módulo es el encargado de obtener las predicciones locales. En la figura 3 se puede observar su estructura, incluyendo su tarea principal *predicción local*. Este módulo requiere la configuración e inicialización previa del sistema, que conlleva la definición del dominio (o dominios) y de las correspondientes componentes principales y del agrupamiento a utilizar (línea discontinua de la figura).

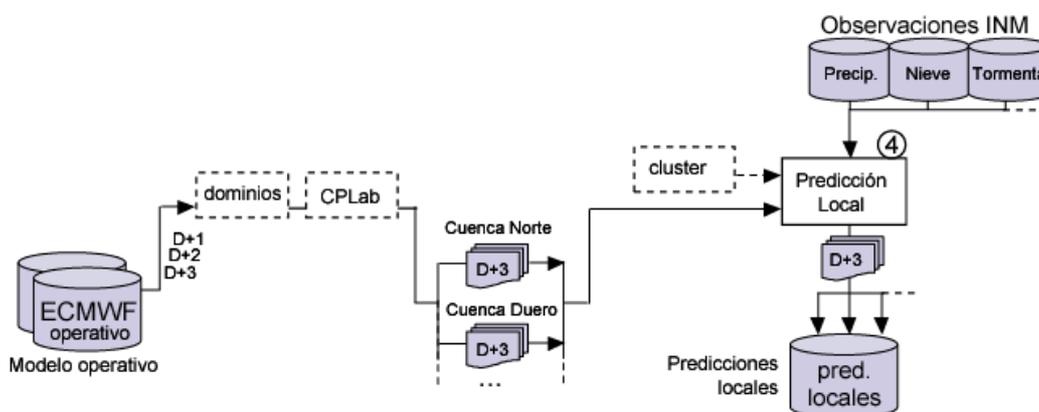


Fig. 3. Módulo de explotación Operativa de la aplicación *Prometeo*

En esta fase se precisan las salidas diarias (predicciones numéricas) de un modelo operativo. En esta aplicación se utiliza el modelo operativo del ECMWF en la pasada de las 12 UTC y para un alcance de 10 días, con salida cada 6 horas, y un dominio espacial global. De estas predicciones se extraen las mismas variables, niveles y área geográfica previamente especificadas en el módulo *dominios*. En este caso sólo es necesario disponer de la predicción para el día de interés y, por tanto, no es necesaria una base de datos, sino sólo un fichero de entrada. Sin embargo, nuestro objetivo es desarrollar una herramienta interactiva que permita explorar y validar un periodo amplio de predicciones. En esta aplicación se han incluido los dos últimos años (2002 y 2003) de predicciones (unos 640 días) en la base de datos *ECMWF operativo*. Esta base de datos se

completa diariamente con las nuevas predicciones obtenidas en días sucesivos. Por tanto, por cada día tendremos 10 días de alcance de predicción, lo que equivale a la misma magnitud de datos que en el reanálisis (100 gigabytes). Debido a la magnitud de información manejada los datos se encuentran almacenados en su formato original (GRIB) para impedir duplicidad de información; además su lectura y decodificación se encuentra optimizada (indexada) para un rápido acceso a esta (detalles sobre este formato y códigos de codificación-decodificación pueden encontrarse en wesley.wwb.noaa.gov/grib.html)

Este módulo involucra las siguientes tareas:

1. *Dominio, CPLab.* Al igual que en la fase de configuración, se filtran los datos operativos sobre la región geográfica indicada. Ahora hay que tener en cuenta que para cada día tenemos distintos alcances (en la figura se indican los alcances D+1, D+2 y D+3, correspondientes a las predicciones hechas un día para cada uno de los tres días siguientes). Las componentes principales de los patrones atmosféricos resultantes son obtenidas aplicando las transformaciones calculadas en la fase de configuración. Por tanto, este módulo se limita a un simple cálculo en la fase operativa. Un aspecto importante de esta fase es que ahora no es necesario almacenar las salidas del módulo *CPLab* ya que sólo se calculan en tiempo real para obtener la predicción.

2. *Predicción local.* Esta es la tarea central de la aplicación ya que en ella se combinan los resultados del agrupamiento (que condensan los datos de re-análisis) con las salidas de los modelos numéricos (en forma de vectores de CPs) y con las observaciones (almacenadas en bases de datos). Todo ello es necesario para obtener una predicción local a partir del método descrito en GUTIÉRREZ *et al.* (2004a). En este módulo también se realizan las distintas tareas de inferencia probabilística y numérica. Por ejemplo, se obtienen predicciones probabilísticas de la precipitación para distintos umbrales y se obtiene una predicción numérica de la cantidad estimada de precipitación (usando tanto la media como un percentil superior para compensar la pérdida de resolución del grupo; GUTIÉRREZ *et al.* (2004a) para más detalles).

Un fichero de configuración define los observatorios de la base de datos de observaciones donde se desea obtener una predicción (2.500 en la aplicación del INM), y asocia cada uno de ellos a un dominio (en este caso, la cuenca hidrográfica a la que pertenece). Estas predicciones se almacenan posteriormente en la base de datos de *predicciones locales* para un posterior acceso y tratamiento.

Una vez que hemos obtenido las predicciones de nuestro sistema, sería deseable poder acceder a ellas de forma interactiva. En la siguiente subsección se explica el módulo encargado de establecer un sistema de acceso interactivo a esta base de datos de predicciones locales, y como veremos más adelante, de acceso a los resultados de las validaciones de estas predicciones.

3. ACCESO WEB A LAS PREDICCIONES

Una forma simple de acceder a esta información es a través de Internet utilizando un navegador. Para ello, se ha diseñado una aplicación Web para acceder a esta información usando cualquier navegador de internet (Fig. 4).

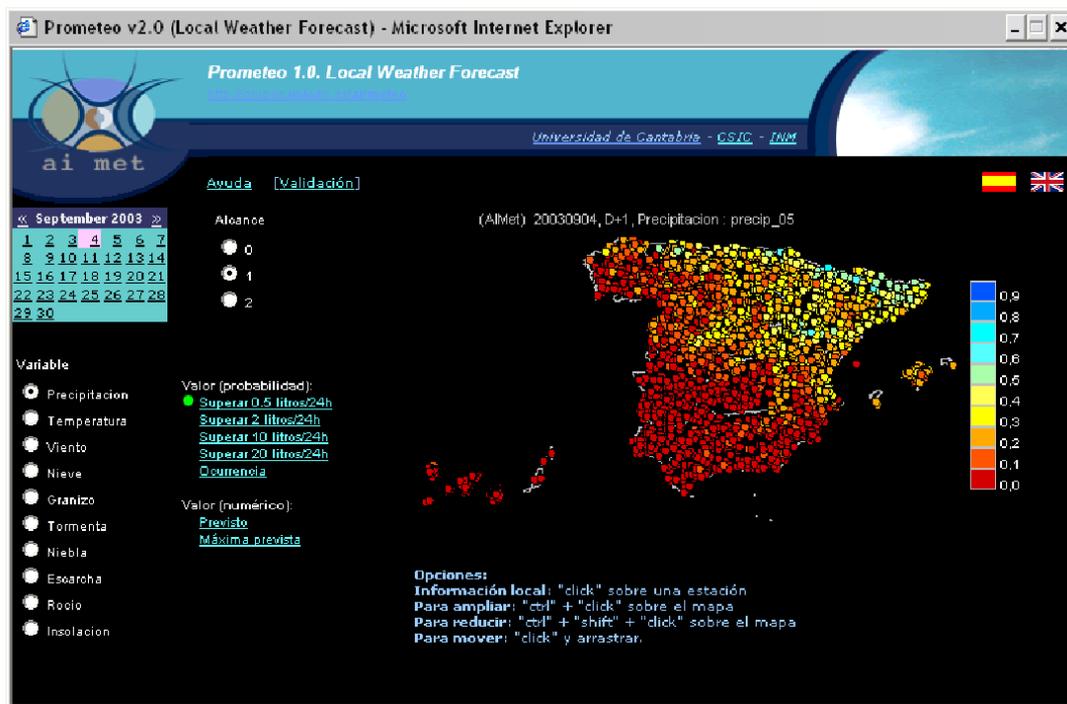


Fig. 4. Aplicación Web de acceso a predicciones locales del sistema *Prometeo* adaptado a la predicción local en la red secundaria del INM, meteo.macc.unican.es/prometeo/

Se ha hecho uso de tecnología de acceso dinámico a bases de datos a través de internet. En particular se ha usado tecnología JAVA <http://www.sun.com/java>, tanto servlets como JSP (Java Server Pages), debido a su difusión, funcionalidad, y carácter abierto de su código. El software de conexión utilizado ha sido el desarrollado por el grupo de trabajo Apache (<http://www.apache.org>) y se denomina Jakarta. Esta tecnología ha permitido desarrollar una aplicación WEB (módulo 5) con el objetivo de que el usuario pueda interactuar con los datos de la predicción local. En la figura 4 se muestra un ejemplo de la página web, en la que se puede seleccionar la información deseada para una fecha concreta y mostrarla en un mapa dinámico donde se muestran las probabilidades con un código de colores (en el ejemplo mostrado se muestra la probabilidad de que la precipitación supere los 0,1mm). Además se puede acceder a la información puntual para una estación concreta sin más que pulsar sobre ella con el ratón.

A continuación se muestran distintos ejemplos de funcionamiento del sistema. En la figura 5 se muestran las predicciones de nieve, granizo, y la predicción numérica de la temperatura máxima, donde se observan valores extremadamente bajos (tener en cuenta que se trata del máximo de temperatura en todo un día) en muchas localidades. Cabe también destacar la predicción realizada para las islas Baleares en las que se observaron nevadas en localidades próximas al mar.

Otro aspecto interesante del sistema es su reflejo de la dinámica y evolución de distintas situaciones sinópticas (como las entradas de frentes en la península). En la figura 7 se puede ver la evolución de la probabilidad de lluvia débil, desde el 24 de marzo del 2003 al 1 de abril del 2003. En ese período puede observarse (a través de sus efectos en la probabilidad de lluvia) la entrada de un frente por el noroeste de la península y su posterior evolución hasta que finalmente desaparece.

Todos los mapas mostrados en esta Sección se han obtenido directamente de la página Web del sistema y las predicciones han sido realizadas en modo operativo, que también pueden ser consultadas en tiempo pasado para comprobar los distintos aspectos de la pericia del modelo. Fruto de esta comprobación también surge la necesidad de realizar y almacenar distintos elementos de validación, los cuales ofrecen valiosa información que permite evaluar las distintas configuraciones dependientes de los parámetros del sistema, y que es utilizada para mejorar las predicciones.

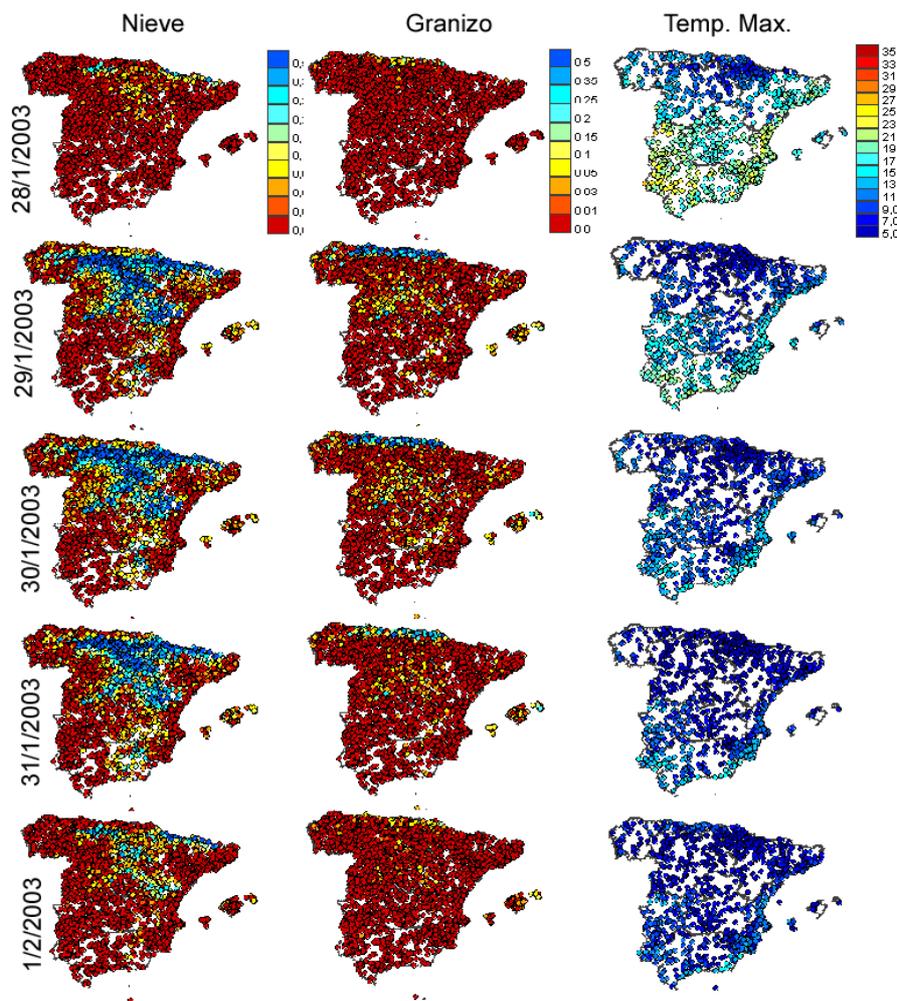


Fig. 5. Predicciones del 28/1/2003 al 1/2/2003 a D+1. En columnas se muestran los eventos $P(\text{Nieve})$, $P(\text{Granizo})$ y valor de temperatura máxima prevista

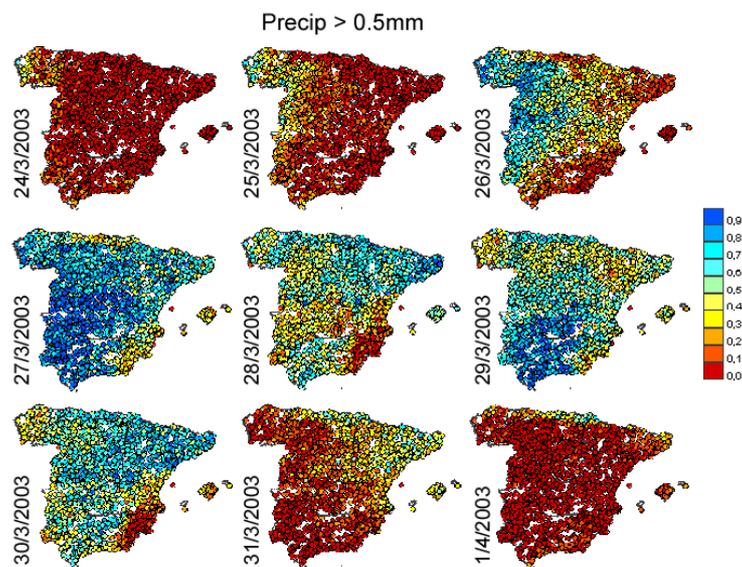


Fig. 6. Predicciones probabilísticas de $P(\text{Precip} > 0.5\text{mm})$ entre el 24/3/2003 y el 1/4/2003 a D+1

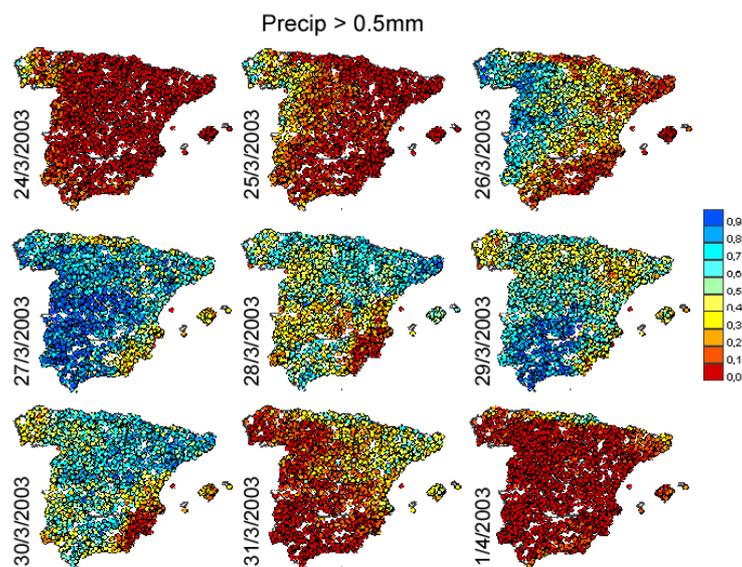


Fig. 7. Predicciones probabilísticas de $P(\text{Precip} > 0.5\text{mm})$ entre el 24/3/2003 y el 1/4/2003 a D+1

3.1. Proceso de validación. Retro-alimentación del Sistema

Una parte importante de la aplicación es el módulo de validación. Este módulo tiene un valor tanto informativo (mostrando la calidad de las distintas predicciones), como técnico (proporcionando una medida de calidad para tomar decisiones en el momento de construir el sistema; por ejemplo, a la hora de seleccionar el número de vecinos óptimo, o un agrupamiento eficiente para el algoritmo. Para que las validaciones sean representativas, es necesario considerar un periodo suficientemente largo de predicciones operativas. Para ello, el módulo de validación puede ejecutar el módulo de producción operativa para un período dado y almacenar los resultados en la base de datos; o bien

puede trabajar con un conjunto de fechas ya almacenadas en la base de datos. Como se comentó anteriormente, en el sistema se han cargado las predicciones de los años 2002-2003 y en base a ellas puedan realizarse las validaciones oportunas. En la figura 8 se muestra la estructura de este módulo, que accede las predicciones locales y las valida frente a las observaciones almacenadas en la base de datos. Después de aplicar distintos índices de validación (Brier Score, Brier Skill Score, fiabilidad, resolución, ROC Área, error cuadrático) almacena los resultados en la base de datos *validación* para cada uno de los alcances del modelo numérico. Los resultados de validación son posteriormente promediados por estación (Invierno, Primavera, Verano u Otoño) y también se almacenan las medias anuales. Además, también se almacena la climatología del evento, para que el usuario pueda comprobar cual es la frecuencia del evento.

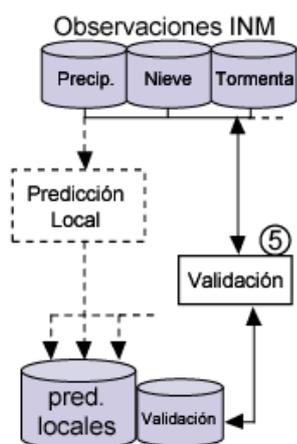


Fig. 8. Módulo de validación de la aplicación *Prometeo*

Estas validaciones son usadas para decidir que estaciones, de todas las que el INM dispone de información, pueden ser usadas con capacidad predictiva, y por tanto saber cuáles son buenas estaciones de observación. No sólo eso, si no que a un usuario experto le puede interesar conocer qué configuración atmosférica de patrones ofrece mejores resultados de validación. Todo esto se calcula de forma simple utilizando este módulo.

En esta sección hemos visto un ejemplo de implementación realizado con el INM, pero todo este sistema posee un núcleo central que permite diseñar esquemas adaptables a otras regiones y datos. Este sistema facilita en gran medida esta labor ya que esta diseñada para aprovechar el esquema de Prometeo descrito, y además ofrecer validaciones en tiempo real a partir de datos de modelos operativos, en tiempo real.

4. REFERENCIAS

- COFIÑO, A.S. (2004). *Técnicas Estadísticas y Neuronales de Agrupamiento Adaptativo para la Predicción Probabilística de Fenómenos Meteorológicos Locales. Aplicación en el Corto Plazo y en la Predicción Estacional*. Tesis Doctoral, Universidad de Cantabria.
- GUTIERREZ, J.M.; CANO, R.; COFIÑO, A.S. and RODRÍGUEZ, M.A. (2004a). "Clustering methods for statistical downscaling in short-range weather forecast". *Monthly Weather Review* (en prensa).
- GUTIÉRREZ, J.M.; CANO, R.; COFIÑO, A.S. and SORDO, C. (2004b). "Analysis and Downscaling Multi-Model Seasonal Forecasts using Self-Organizing Maps". *Tellus A* (en prensa).