

# Iniciativa de Programa de e-Ciencia e IRIS-GRID

## Área Temática de Meteorología

Este documento es un primer y breve borrador para coordinar la actividad de preparación de la propuesta para el Área temática de la Meteorología en el Programa de e-Ciencia de red IRIS.

### Hoja de Estado del documento (REGISTRO DE CAMBIOS)

Version	Fecha	Cambios realizados	Autor
0.0	01-05-2003	Primer borrador de la propuesta, identificando problemas apropiados para GRID y grupos interesados.	José Manuel Gutiérrez (Universidad de Cantabria) gutierjm@unican.es

## 1. e-Ciencia y Tecnologías GRID en Meteorología

Históricamente, la Meteorología ha sido una de los principales usuarios de las nuevas tecnologías de la Computación, tanto en lo relativo a la capacidad de cálculo, como al almacenamiento de grandes volúmenes de información y a su rápida distribución mediante redes de alto rendimiento. En el pasado, muchas de las tareas involucradas en este área (integración de modelos numéricos de predicción, mantenimiento de bases de datos operativas, etc.) eran exclusivas de grandes centros meteorológicos que disponían de la tecnología necesaria. En la actualidad, la situación es distinta debido al abaratamiento de la tecnología, y diversos grupos de investigación públicos y privados llevan a cabo costosas simulaciones meteorológicas que utilizan distintas bases de datos para realizar tareas tan diversas como: estudios climatológicos y de cambio climático, pronóstico meteorológico local, predicción de viento para la gestión de parques eólicos, difusión de contaminantes en el mar y en la atmósfera, etc.

Sin embargo, el tipo de estudios que pueden emprender estos grupos en un tiempo limitado está todavía limitado por los recursos computacionales de que disponen. Asimismo, la colaboración entre distintos grupos para abordar proyectos comunes se ve dificultada por la heterogeneidad de bases de datos y formatos de uso común en este ámbito. Por tanto, la tecnología GRID puede proporcionar un doble beneficio en este Área permitiendo abordar problemas más complejos a los centros de investigación, reduciendo los tiempos necesarios para realizar estas tareas, y facilitando la colaboración y acceso/compartición de datos a los mismos.

La propuesta de Centros de e-Ciencia en la que se encuadra esta iniciativa parece el marco más apropiado para implantar esta tecnología, proporcionando un soporte técnico apropiado para emprender iniciativas en este Área.

## 2. El Grid en el Actual Plan de I+D+I

Grid Computing es prioridad temática dentro del Área de Tecnologías de la Información y de las Comunicaciones del Plan Nacional de I+D+I 2000-2003:

1.6. Computación de altas prestaciones

1.6.5. Grid Computing

Además también tiene relación con las siguientes prioridades:

## 1.6. Computación de altas prestaciones

### 1.6.4. Herramientas y entornos para el desarrollo de aplicaciones.

## 1.5. Sistemas en red

### 1.5.2. Sistemas basados en la Web. Servicios Web. XML.

Este tema también está priorizado en algún programa nacional que utiliza la computación como herramienta; por ejemplo, está relacionado con la prioridad temática del Programa Nacional De Física De Partículas y Grandes Aceleradores

## 2. Tecnologías en relación con el proyecto LHC Computing GRID del CERN

2.1 Desarrollo de tecnologías para cálculo en la red (Proyecto LHC Computing GRID del CERN). Se priorizarán los proyectos GRID coordinados con los del Programa Nacional de TIC.

En febrero de 2001 la Comisión Europea presentó su propuesta para el VI Programa Marco, que abarca el periodo 2002-2006 ([www.mcyt.es/vipm/](http://www.mcyt.es/vipm/)). Este VI Programa Marco se ha planteado en el contexto del documento "*Hacia un Espacio Europeo de Investigación*". Como resultado del análisis que se realiza en el documento de enero de 2000, se propone la integración de la I+D europea para alcanzar masas críticas, lograr economías de escala y distribuir mejor los recursos. Se plantea la necesidad de

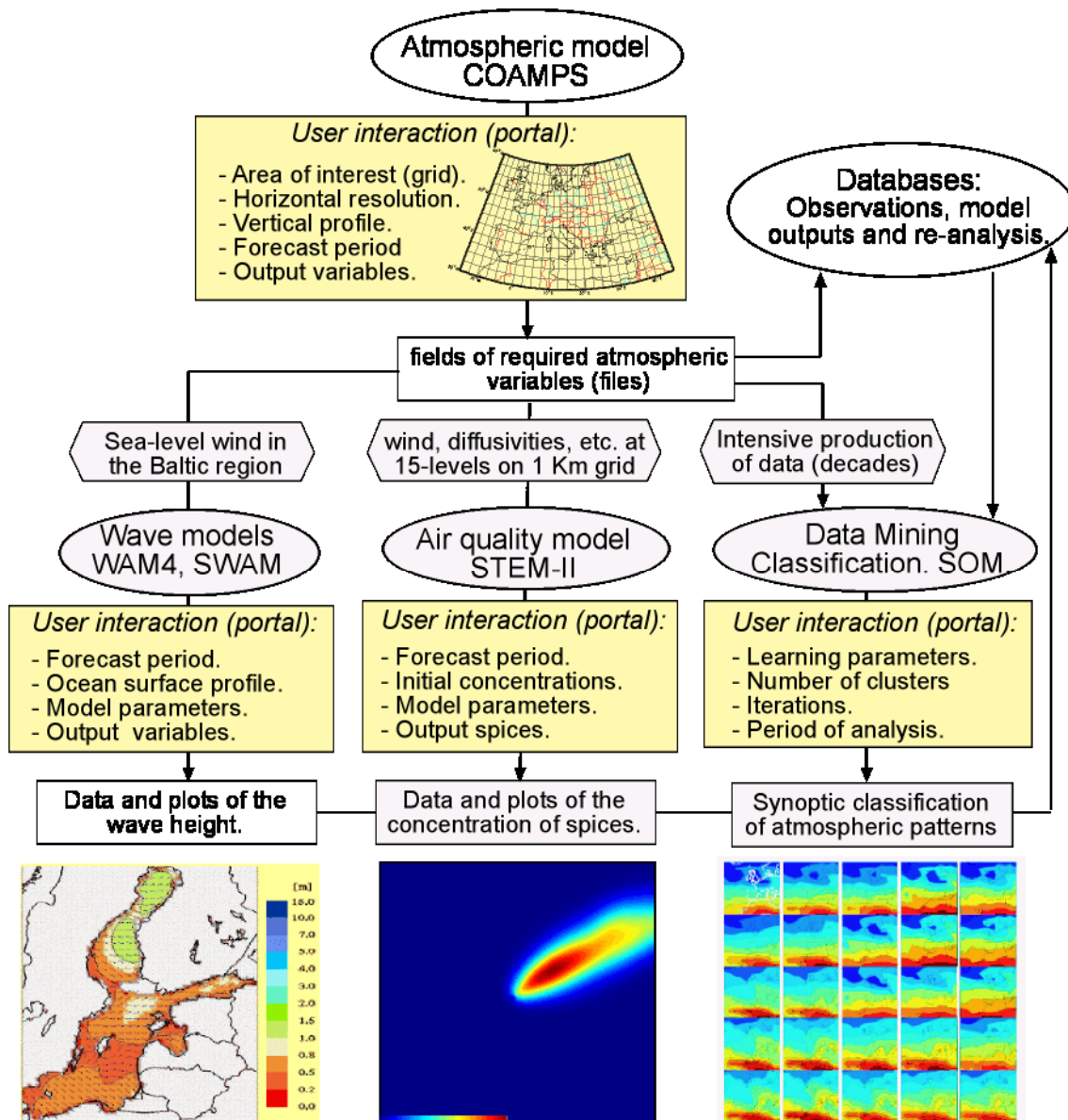
- (a) Interconectar centros de excelencia europeos y desarrollar *centros virtuales*,
- (b) apoyar la creación, funcionamiento y acceso a las grandes infraestructuras,

Existe por tanto una clara **sintonía entre estos objetivos y el paradigma que plantea el *Grid Computing***, que puede considerarse como la tecnología de base que habilita un acceso transparente a los recursos científicos distribuidos.

Así, el VI Programa Marco se basa en los principios de concentración de esfuerzos, de mayor coordinación entre las actividades de I+D nacionales y europeas, y en la descentralización de la gestión. Grid computing ha sido una de las líneas temáticas priorizada en las primeras llamadas a proyectos del VI programa marco, en concreto dentro de la denominación "Complex Problem Solving" donde la Meteorología y Biología tienen un papel predominante hoy día.

## 3. Experiencias GRID Previas en Meteorología

Los primeros desarrollos de aplicaciones meteorológicas en Europa en entornos GRID se están llevando a cabo en el proyecto Europeo del V Programa Marco CROSS-GRID (<http://www.crossgrid.org>), con una activa participación de grupos Españoles en el ámbito de la dispersión de contaminantes en la atmósfera y en la implementación de herramientas de minería de datos para bases de datos meteorológicas. Fruto de estas iniciativas se están resolviendo distintos problemas de migración y adaptación de las técnicas y productos existentes al nuevo entorno GRID, facilitando el trabajo para futuras iniciativas en este campo. Otra de las experiencias piloto en este proyecto es la paralelización y migración a GRID del modelo COAMPS "Coupled Ocean/Atmosphere Mesoscale Prediction System" en forma de servicio web en el que el usuario pueda seleccionar interactivamente un área de interés, la resolución horizontal y vertical, y un período de predicción, obteniendo los campos meteorológicos solicitados resultado de la integración del modelo. Está planeado integrar este servicio web con aplicaciones específicas de cálculo de contaminantes en la atmósfera, cálculo de oleaje, y procesamiento y análisis de proyectos de re-análisis regionales. En la figura siguiente se muestran más detalles.



## 4. Características de los Procesos y Datos Involucrados

En esta sección describimos el tipo de aplicaciones y el tipo y volumen de datos involucrados en la actividad en el área de la Meteorología.

### 4.1. Procesos que requieren computación de altas prestaciones

La integración de cualquier modelo numérico de circulación atmosférica u oceánica es un proceso computacionalmente costoso, tanto en tiempo de cómputo como almacenamiento de información. Estos modelos constituyen la herramienta fundamental para numerosas aplicaciones prácticas. Por otra parte, la aplicación de distintas técnicas estadísticas (correlación canónica, regresión múltiple, técnicas de clasificación, etc.) a campos atmosféricos almacenados requiere también un enorme esfuerzo computacional debido a las dimensiones de los datos.

En la actualidad las técnicas de predicción basadas en conjuntos requieren aún mayor esfuerzo computacional ya que los modelos se integran varias veces, a partir de condiciones iniciales perturbadas. Este tipo de aplicaciones es un ejemplo típico de aplicación paramétrica, que puede explotarse fácilmente en un entorno GRID obteniendo una mayor productividad.

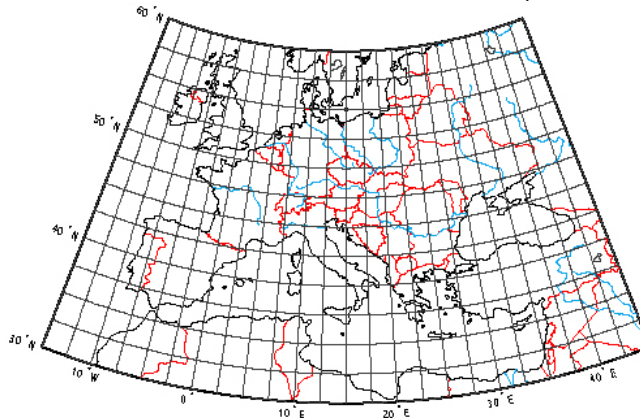
Por otra parte, los proyectos de reanálisis son necesarios para llevar a cabo estudio climatológicos en zonas de interés. En este caso, los modelos se integran durante largos períodos de tiempo guardando principalmente los análisis obtenidos; esta tarea da lugar a enormes bases de datos que definen las climatología de la zona de interés (a la resolución dada por el modelo integrado). Éste es otro ejemplo de aplicación paramétrica en la que el entorno GRID permitiría incrementar la productividad.

## 4.2. Bases de datos

Una de las principales características de esta área es la gran cantidad y diversidad de datos que se manejan en cualquier aplicación real. En esta sección tratamos de describir las fuentes más importantes.

### 4.2.1. Salidas de modelos numéricos

Los modelos numéricos de circulación atmosférica son la principal herramienta en la Meteorología moderna. Estos modelos proporcionan una predicción del estado de la atmósfera a distintas resoluciones temporales (de horas a meses) y espaciales (entre los cientos y las decenas de Km de resolución horizontal). Por ejemplo, la figura siguiente muestra una rejilla horizontal utilizada para simular la dinámica mensual de la atmósfera sobre Europa, con una resolución horizontal de 300 Km (esta rejilla tiene otra componente vertical, con 15 niveles de altura distintos).

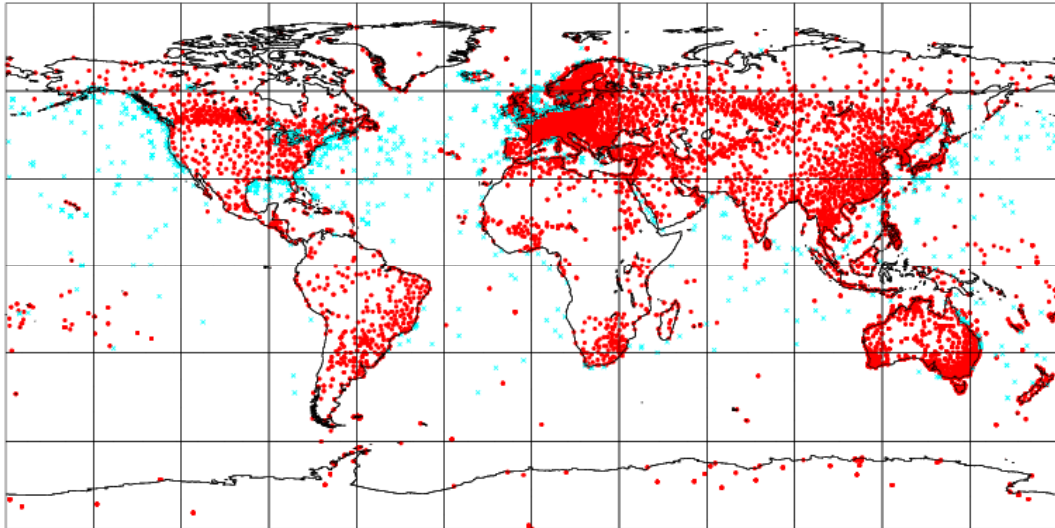


En cada uno de estos puntos de la rejilla 3D se simula el valor de distintas variables primitivas (temperatura, humedad, geopotencial, viento) y derivadas (espesores, vorticidad, precipitación, etc.). Todo este volúmen de información se genera en cada paso de integración y se almacena horaria o diariamente, dependiendo de la resolución temporal de la integración. Por tanto el volúmen de información generado en cada simulación es enorme.

Aparte del uso operativo de estos modelos, en la actualidad resultan de gran utilidad los proyectos de *Re-análisis*, donde un mismo modelo es integrado un largo período de tiempo (decenas de años) para obtener salidas homogéneas de un mismo modelo válidas para estudios Climatológicos o Estadísticos. El volúmen de datos almacenados en estas bases de datos de reanálisis es del orden de Tera Bytes. Algunas de estas bases de datos de reanálisis son públicas, como la del NCEP/NCAR y otras están disponibles para trabajos de investigación (como la del ECMWF).

### 4.2.2. Observaciones y datos de estaciones meteorológicas

Las estaciones situadas en tierra, en barcos y en aviones, proporcionan observaciones de variables atmosféricas en todo el mundo. Por ejemplo, la siguiente figura muestra la red de observatorios Synop y observatorios situados en barcos utilizada por el ECMWF para la predicción meteorológica diaria.



En España el INM dispone de observaciones de precipitación y meteoros en más de 6000 estaciones a lo largo y ancho de la geografía, así como de una red de más de 200 estaciones completas con información sobre viento, insolación y temperaturas, además de la precipitación y meteoros. A su vez, distintos servicios meteorológicos autonómicos, Universidades y centros de investigación disponen de datos específicos complementarios de distintas resoluciones espaciales y temporales (datos minutales de viento para generadores eólicos, etc.). El número de estaciones automatizadas que se instalan para distintas aplicaciones (estudio de cuencas hidrológicas, etc.) crece cada día proporcionando una tupida red de datos, algunos de los cuales son públicos y otros privados para los que el entorno GRID tendrá que proporcionar medidas de privacidad apropiadas.

#### **4.2.3. Datos de satélites y radares**

Aparte de los datos ya mencionados, existe numerosa información pública relativa a mediciones de satélites y radares, que proporcionan datos interesantes sobre la dinámica de las nubes y la fenomenología asociada. En el corto plazo estos datos (imágenes, etc.) pasarán a formar parte de la cadena operativa, siendo asimilados por los modelos, o siendo postprocesados con técnicas inteligentes de minería de datos.

## **5. Definición de Posibles Proyectos Piloto**

La iniciativa propuesta en el Área de la Meteorología está vertebrada sobre distintas aplicaciones piloto con las que se pretende demostrar los beneficios de la tecnología GRID en este Área. La implementación de estas aplicaciones requerirá una estrecha colaboración con los grupos encargados del desarrollo de middleware (software intermedio entre las aplicaciones y el entorno GRID) que proporcionarán las herramientas apropiadas para que el acceso a datos, ejecución de procesos en distintos clusters, etc., sean tareas casi-transparentes para los grupos que desarrollen aplicaciones.

Por otra parte, una iniciativa de este tipo tiene el beneficio añadido de formentar el **trabajo colaborativo** entre distintos grupos de investigación en Meteorología Españoles (compartición segura de observaciones, salidas de modelos numéricos, reanálisis, predicciones operativas, etc.), así como el desarrollo de proyectos comunes aunando esfuerzos computacionales.

La propuesta de aplicaciones piloto es la siguiente:

1) Un problema de enorme interés en la actualidad es la **predicción por conjuntos**. En la actualidad la tecnología GRID está suficientemente madura para permitir la ejecución de aplicaciones paramétricas, permitiendo incrementar su productividad. La predicción por conjuntos es un ejemplo típico de este tipo, ya que se trata de ejecutar la misma aplicación (modelo atmosférico) con distintas condiciones iniciales (parámetros); por tanto, un primer ejemplo realista de aplicación Meteorológica en el entorno GRID sería el desarrollo de un prototipo de sistema de predicción por conjuntos que integrase los distintos modelos y métodos de perturbación utilizados por los distintos grupos nacionales.

2) **Integración de modelos de área limitada en regiones de interés:** Un denominador común a los distintos grupos que realizan desarrollo e investigación en el ámbito de la Meteorología es la necesidad de simular la circulación de la atmósfera en determinadas condiciones (incluso diariamente, de forma operativa). En la actualidad existen distintas simulaciones globales de baja resolución disponibles, tanto de reanálisis (integraciones en tiempo pasado para un período largo de tiempo, 10-50 años), como operativas (que proporcionan los campos atmosféricos previstos con unos días, semas, o incluso meses de antelación). En la mayoría de las ocasiones, la resolución de estos modelos no es suficiente para analizar diversos problemas locales y, por tanto, es necesario integrar modelos de mayor resolución sobre una zona de interés concreta (modelos de área limitada, o modelos regionales), utilizando como condiciones iniciales y/o de contorno de los modelos globales.

Uno de los modelos regionales de dominio público más populares es el MM5 (por ejemplo, en la Península distintos grupos de investigación que componen la "Red Ibérica para la investigación y desarrollo de aplicaciones en base al modelo atmosférico MM5" <http://redibericamm5.uib.es/> integran separadamente este modelo en distintas regiones peninsulares).

La integración eficiente de un modelo atmosférico aprovechando el entorno grid involucra el análisis de la escalabilidad de la implementación paralela dependiendo de los recursos computacionales y de la latencia de la red (recursos que, a diferencia de la ejecución en un cluster local, varían de una ejecución a otra). Por tanto, una aplicación piloto de gran interés científico sería estudiar el rendimiento de distintas paralelizaciones de **MM5 en el entorno GRID**.

3) Otra aplicación piloto de gran interés para la comunidad científica es la realización de proyectos de **reanálisis de alta resolución**. En este caso, se tratará de integrar el mismo modelo atmosférico durante un largo período de tiempo (representativo de la climatología del problema que se quiera abordar). En este caso, se podría utilizar el modelo MM5 como aplicación paramétrica que sería ejecutada con distintas fechas (parámetros) en el entorno GRID aumentando la productividad. Esto permitiría disponer de reanálisis regionales en áreas de interés en un tiempo razonable. El modelo puede inicializarse a partir de condiciones de reanálisis del NCEP o del ECMWF.

4) Una vez que en el GRID haya sido almacenada información de reanálisis (global y regional), podrían implementarse **técnicas inteligentes de acceso a la información (minería de datos)**. En este caso se trata de obtener patrones de información (Empirical Orthogonal Functions EOFs, clasificaciones, etc.), en lugar de la cantidad ingente de datos crudos. Por ejemplo, en muchas ocasiones sólo es necesario acceder a las componentes principales de ciertos campos, a las correlaciones canónicas de campos y observaciones, o incluso a

los campos análogos a un campo dado (los vecinos); en estos casos, sería de gran utilidad disponer de servicios de acceso a la información que proporcionasen directamente la información solicitada. Este tipo de aplicaciones se encuadra en el marco más general de la Minería de Datos, necesaria para gestionar de forma apropiada complejas bases de datos con ingente información.

## 6. Grupos Nacionales Interesados en la Propuesta

La propuesta de colaboración en el desarrollo de una infraestructura GRID en el ámbito de la Meteorología cuenta con el interés de distintos grupos de investigación, que se beneficiarían tanto de la nueva tecnología desarrollada, como de la colaboración con otros grupos en el desarrollo de problemas complejos de interés común. Una primera lista de grupos interesados es la siguiente:

- Grupo de Inteligencia Artificial en Meteorología <http://grupos.unican.es/ai/meteo>  
**Universidad de Cantabria / CSIC / Instituto Nacional de Meteorología**  
*Contacto:* José Manuel Gutiérrez ([gutierjm@unican.es](mailto:gutierjm@unican.es))
- Grupo de Física No Lineal <http://chaos.usc.es>  
Unidade de Observacion y Prediccion Meteorologica <http://meteo.usc.es>  
**Universidad de Santiago de Compostela**  
*Contacto:* Vicente Pérez Muñuzuri ([uscfmvpm@cesga.es](mailto:uscfmvpm@cesga.es))
- Grupo de Meteorología  
**Universidad de las Islas Baleares**  
*Contacto:* Juan M. Torres ([joanm.torres@uib.es](mailto:joanm.torres@uib.es))
- Grupo de Predicción y caracterización del recurso eólico  
**CIEMAT**  
*Contacto:* Jorge Navarro Montesinos ([jorge.navarro@ciemat.es](mailto:jorge.navarro@ciemat.es))
- Meteorology and Climate Applications and Modelling (MCAM)  
**Universidad Complutense - Universidad de Murcia**  
*Contacto:* Juan Pedro Montavez ([jpmontav@ucmail.ucm.es](mailto:jpmontav@ucmail.ucm.es))
- Grupo de Meteorología  
**Universidad del País Vasco**  
*Contacto:* Jon Saenz ([jsaenz@wm.lc.ehu.es](mailto:jsaenz@wm.lc.ehu.es))