

Procesamiento paralelo de los pronósticos meteorológicos del modelo WRF mediante NCL

Guerrero-Higueras, A.M.¹, García-Ortega, E.², Matellán-Olivera, V.³, Sánchez, J.L.⁴

Resumen— La predicción meteorológica es un problema clásico de la programación paralela. Muchos de los modelos matemáticos que se utilizan en meteorología para pronosticar el comportamiento de la atmósfera tienen implementaciones preparadas para ejecutarse en entornos de cálculo paralelo.

La información obtenida después de ejecutar un modelo matemático de predicción meteorológica es necesario procesarla para poder visualizar los resultados. Sin embargo, no es fácil encontrar herramientas que permitan el procesado en paralelo de estas salidas.

El presente trabajo plantea la posibilidad de dar un paso más y mejorar el rendimiento, utilizando el paradigma paralelo para procesar las salidas de un modelo de predicción numérica, en concreto, el modelo WRF (Weather Research and Forecasting).

Palabras clave— Meteorología, predicción numérica, WRF, MM5, netCDF, NCL.

I. INTRODUCCIÓN

La predicción meteorológica consiste en la determinación anticipada de los valores de una serie de variables meteorológicas, como temperatura, presión, humedad, nubosidad, precipitación, etc., que afectan a una determinada región. En la actualidad las predicciones se realizan utilizando modelos de predicción numérica que emplean las ecuaciones de las leyes físicas que describen el comportamiento de la atmósfera. La resolución de las mismas es ciertamente complicada puesto que se trata de ecuaciones para las que no existe una solución matemática exacta que permita determinar los valores futuros de las variables. Por esta razón, es necesario recurrir a las llamadas técnicas de modelización numérica.

Formalmente, la utilización de los modelos de predicción del tiempo se basa en resolver de forma numérica, por tanto mediante aproximaciones, un conjunto de ecuaciones diferenciales referidas a la conservación de la masa, la energía y el momento lineal en la atmósfera. Cada modelo hace uso de técnicas peculiares que le confieren ciertas características. Partiendo de unas condiciones iniciales establecidas mediante un proceso de asimilación de datos, a partir de un modelo de circulación general y de un conjunto de observaciones, en un determinado instante de tiempo (t), el modelo, o más

bien sus ecuaciones, obtienen los valores de las variables pronosticadas en un tiempo posterior ($t + \Delta t$). Para ello es preciso hacer uso de las ecuaciones en derivadas parciales, que traducen las leyes generales de la física que rigen la dinámica de la atmósfera terrestre.

La atmósfera es un fluido en movimiento sometido a procesos de intercambio de energía. En consecuencia, se necesita 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.

El movimiento de una partícula de fluido atmosférico viene dada por la ecuación 1.

$$\frac{d\vec{V}}{dt} = \sum_i \frac{\vec{F}_i}{m} = -\frac{1}{\rho} \vec{\nabla} p + \vec{g} - 2\vec{\Omega} \times \vec{V} + \vec{F}_r \quad (1)$$

El término de la izquierda representa la fuerza por unidad de masa que actúa sobre una burbuja de aire, medida desde un sistema de referencia no inercial (la Tierra). Los términos de la derecha señalan las fuerzas por unidad de masa que actúan sobre dicha burbuja: la fuerza debida al gradiente de presión, resultante de la fuerza gravitatoria y centrífuga, la fuerza de Coriolis, que junto a la anterior, son los dos términos ficticios de la ecuación, y el término asociado a las fuerzas de rozamiento. Para resolver esta ecuación es necesario introducir una serie de simplificaciones en función de las escalas espacio-temporales utilizadas en cada caso.

Desde el punto de vista termodinámico hay que tener en cuenta los intercambios de calor asociados a los cambios de fase del agua (vapor, líquido y estado sólido) que afectan a la formación y desarrollo de las nubes y a los intercambios de energía con la superficie de la Tierra, los océanos y la troposfera.

Otro factor a tener en cuenta es el hecho de que la atmósfera es un fluido gaseoso que sufre variaciones

¹Grupo de Física de la Atmósfera, Univ. de León, e-mail: am.guerrero@unileon.es.

²Grupo de Física de la Atmósfera, Univ. de León, e-mail: eduardo.garcia@unileon.es.

³Dpto. de Ingenierías Mecánica, Informática y Aeroespacial, Univ. de León, e-mail: vicente.matellan@unileon.es.

⁴Grupo de Física de la atmósfera, Univ. de León, e-mail: jl.sanchez@unileon.es.

de densidad como consecuencia de la dinámica a la que se encuentra sometida.

Las ecuaciones matemáticas que describen todos estos procesos no son lineales y, por tanto, las soluciones no se pueden obtener de forma analítica. Para resolverlas es preciso recurrir al cálculo numérico, que proporciona una solución aproximada.

Para obtener estas soluciones aproximadas, el primer paso consiste en discretizar el área geográfica objeto de estudio. Para ello se define una rejilla (*grid*) tridimensional que la abarque. La distancia entre dos puntos de la rejilla se conoce como resolución del modelo.

El siguiente paso consiste en resolver las ecuaciones, utilizando cálculo numérico, en todos los puntos de la rejilla. Es en este punto donde el paradigma de la programación paralela entra en juego. Casi todos los modelos de predicción numérica tienen implementaciones que pueden ejecutarse en entornos de cálculo paralelo, aprovechando, en la medida de lo posible, la relativa independencia entre cada punto de la rejilla.

II. MODELOS MESOESCALARES

Se entiende por modelo mesoescalar aquel que permite estudiar y pronosticar fenómenos meteorológicos, incluyendo los mesoescalares, es decir, aquellos que ocurren a una escala espacial relativamente pequeña. Históricamente, los primeros modelos eran hidrostáticos pues consideraban despreciables los movimientos en la escala vertical frente a los movimientos en la horizontal. 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 hipótesis iniciales de partida. Sin embargo, cuando la escala de la rejilla del modelo se reduce, la resolución espacial aumenta¹ y es posible incorporar movimientos en la vertical, y por tanto, la hipótesis hidrostática deja de verificarse. Esto es importante pues permite incorporar algunos factores meteorológicos como la convección, ondas, etc. que son de tipo mesoescalar.

Por tanto hay dos grandes tipos de modelos numéricos: hidrostáticos, como HIRLAM², y no hidrostáticos, como MASS, MM5, o la versión avanzada de este denominada WRF.

En la actualidad, 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 de precipitaciones intensas y severas. En España se utilizan, mayoritariamente, dos modelos no hidrostáticos: MASS³ y MM5,

¹La resolución espacial aumenta a medida que la distancia entre puntos de la rejilla disminuye.

²Utilizado por la Agencia Estatal de Meteorología.

³Empleado de forma operativa por el Servei Meteorologic de Catalunya.

aunque recientemente se ha comenzado a trabajar con WRF. Estos últimos se emplean principalmente por grupos científicos y algunos servicios regionales de meteorología. Experimentalmente también se utiliza ocasionalmente Meso NH, desarrollado en Francia.

El Grupo de Física de la Atmósfera (GFA) de la Universidad de León trabaja desde hace diez años con modelos mesoescalares [1] [2] [3]. Inicialmente se utilizó Meso NH, más tarde MM5 y WRF. Este último se ejecuta de manera operativa en el centro de supercomputación de Castilla y León⁴. Las salidas pueden consultarse en <http://gfa.unileon.es>.

Todos los modelos de simulación mesoescalares 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.

III. EL MODELO MESOESCALAR WRF

The Weather Research and Forecasting (WRF) 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 [4] [5]. Se trata de un modelo a disposición de la comunidad científica cuyas principales características son las siguientes:

- Capacidad de anidamiento múltiple.
- Dinámicas no hidrostáticas, lo que permite al modelo ser usado en una escala de unos pocos kilómetros.
- Una gran capacidad de asimilación de información.
- Múltiples opciones de parametrizaciones físicas.

WRF ha sido desarrollado en colaboración entre la National Oceanic and Atmospheric Administration (NOAA), el National Center for Atmospheric Research (NCAR) y otras organizaciones.

La implementación del modelo tiene capacidad multitarea para memoria compartida y distribuida. Está preparado para trabajar en entornos de cálculo paralelo con memoria compartida, mediante OpenMP, y de memoria distribuida, mediante MPI. Además, el modelo tiene la capacidad de combinar ambas tecnologías.

A. Módulos

El modelo de predicción numérica WRF se compone de varios módulos que se reparten distintas funcionalidades [6]:

- GEOGRID: Permite configurar el área geográfica objeto de estudio.
- UNGRID: Prepara los datos de inicialización del modelo y las condiciones de contorno.
- METGRID: Interpola horizontalmente los diferentes campos meteorológicos extraídos de los datos de inicialización del modelo.

⁴Más información sobre el centro de supercomputación de Castilla y León en <http://www.fcsc.es>.

- REAL: Realiza la interpolación vertical de los datos.
- WRF: Contiene las ecuaciones físicas de pronóstico y diagnóstico que permiten realizar una predicción con un horizonte temporal prefijado.

A.1 GEOGRID

Geogrid 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 de dicho dominio.

El anidamiento es necesario para realizar pronósticos en áreas del orden de mesoescala. La resolución espacial de las condiciones iniciales fijadas desde un modelo de circulación general obliga a definir dominios que permitan ejecutar el modelo, imponiendo las condiciones de contorno necesarias para asegurar la estabilidad de las salidas.

A.2 UNGRIB

En este módulo se preparan los datos de inicialización del modelo y las condiciones de contorno, fijados por la salida de un modelo de mayor cobertura espacial, como GFS⁵ o ECMWF⁶, adaptándose a las características de los dominios definidos en GEOGRID.

A.3 METGRID

En este módulo los valores de los campos meteorológicos correspondientes a las condiciones iniciales, son interpolados horizontalmente a la rejilla definida.

A.4 REAL

REAL realiza la interpolación vertical desde los niveles de presión hasta el sistema de coordenadas normalizadas sigma de WRF (ver figura 1). 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 predicción.

A.5 WRF

El módulo WRF contiene las ecuaciones físicas de pronóstico y diagnóstico que permiten realizar una predicción con un horizonte temporal prefijado.

En este módulo se incluyen también las parametrizaciones que permiten describir los fenómenos físicos cuya escala espacial/temporal es inferior a la resolución fijada en las simulaciones.

B. Las salidas de WRF. El formato netCDF

Las salidas que proporciona el modelo WRF contienen información pronosticada, para un intervalo de tiempo determinado, de diferentes variables meteorológicas. Los ficheros que componen estas salidas

⁵Global Forecast Model, ejecutado por NOAA (National Oceanic and Atmospheric Administration).

⁶European Centre for Medium-Range Weather Forecasts

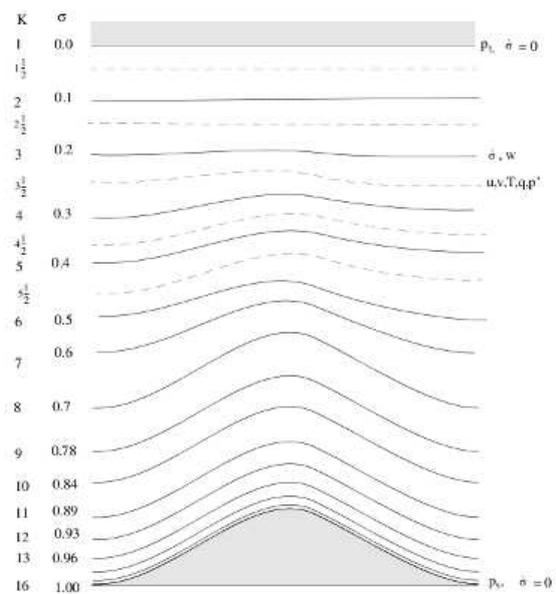


Fig. 1. Coordenadas verticales sigma normalizadas.

cumplen con las especificaciones de un formato conocido como netCDF [7] [8].

NetCDF (network Common Data Form) es un conjunto de interfaces de acceso para datos en forma de *array* y una colección de librerías, libremente distribuidas, de acceso para datos implementadas para C, Fortran, C++, Java y otros lenguajes. Las librerías proporcionan un formato, multiplataforma, para representar información científica. Juntos, interfaces, librerías y formato permiten la creación, el acceso y el intercambio de datos científicos.

Por tanto, la información en formato netCDF es:

- *Autodescrita.* Un fichero netCDF incluye información sobre los datos que contiene.
- *Portable.* Un fichero netCDF puede ser accedido por ordenadores con diferentes formatos para números enteros, caracteres y números en coma flotante.
- *Escalable.* Un pequeño subconjunto de datos que forman parte de un conjunto más grande de datos es accesible de una manera eficiente.
- *Agregable.* Es posible añadir datos a un fichero netCDF bien construido sin redefinir su estructura.
- *Intercambiable.* Un escritor y múltiples lectores pueden acceder de manera simultánea al mismo fichero netCDF.
- *Archivable.* Las versiones futuras de netCDF serán compatibles con las anteriores.

El proyecto netCDF está enmarcado dentro del programa Unidata de la UCAR (University Corporation for Atmospheric Research).

B.1 Estructura de un fichero netCDF

Los ficheros netCDF, habitualmente con extensión *.nc*, son de tipo binario. Pero para poder analizarlos y editarlos se pueden representar en modo texto.

Es posible guardar esta representación en formato

de texto plano en un fichero, habitualmente con extensión `.cdl`, y generar un netCDF binario a partir de este y viceversa, gracias a los comandos `ncdump`, `ncview` y `ncgen`⁷:

- `ncdump FILENAME`: Muestra el contenido completo del fichero `FILENAME`.
- `ncdump -h FILENAME`: Muestra solamente las cabeceras del fichero `FILENAME`, sin incluir los datos, lo cual es muy útil para analizar el fichero.
- `ncdump -v VARIABLE FILENAME`: Muestra el contenido de la variable `VARIABLE` contenida en el fichero `FILENAME`.
- `ncview FILENAME`: Permite visualizar contenidos de `FILENAME`.
- `ncgen -o NC_FILENAME CDL_FILENAME`: Genera un fichero binario llamado `NC_FILENAME` a partir de una representación en texto plano en el fichero `CDL_FILENAME`.
- `ncdump NC_FILENAME > CDL_FILENAME`: Genera una representación en texto plano en el fichero `CDL_FILENAME` a partir del fichero binario `NC_FILENAME`.

Los datos en formato netCDF se almacenan en forma vectorial. Cada vector almacena los datos de una variable concreta y tiene unas dimensiones y atributos acordes a los datos. La estructura básica del fichero netCDF, en su representación plana, tiene 4 partes, como se puede ver en el ejemplo de la figura 2.

- Dimensiones: Especificación de las diferentes dimensiones que pueden tener las variables contenidas en él. Como por ejemplo latitud, longitud, altura, tiempo, etc.
- Variables: Definición de las diferentes variables incluídas en el fichero, especificando sus dimensiones y atributos.
- Atributos globales: Lista de atributos aplicables a todas las variables del fichero.
- Datos: Los datos propiamente dichos.

C. Herramientas de postproceso

WRF proporciona salidas en formato netCDF. Sin embargo, esta información es necesario procesarla y generar una serie de gráficas que permitan analizarla y entenderla de una manera visual.

Hay un buen número de herramientas disponibles para visualizar la información de modelo WRF. Esta información puede ser visualizada desde cualquier herramienta capaz de reconocer el formato netCDF. Entre las más conocidas destacan GrADS, RIP4, Vis5D, VAPOR y NCL.

El GFA utiliza, históricamente, GrADS [9] para procesar las salidas de WRF, como antes hizo con las salidas de MM5. Sin embargo en la actualidad se está trabajando con NCL (NCAR Command Language) [10] debido a su mayor versatilidad y a que dispone de funciones específicas para trabajar con las salidas de WRF.

NCL es un lenguaje de script específicamente diseñado para el proceso y la visualización de información científica [11] [12]. Puede leer ficheros netCDF y genera gráficas de alta resolución altamente configurables.

Las predicciones operativas del GFA generan gráficas para las siguientes variables meteorológicas:

- Temperatura y geopotencial a 850, 700, 500 y 300 hPa.
- Temperatura y viento en superficie.
- Humedad relativa y viento a 850, 700, 500 y 300 hPa.
- Presión en superficie.
- Precipitación y precipitación acumulada.
- Espesor de nieve.
- Agua en forma de nieve.

IV. PROCESAMIENTO SECUENCIAL

La mayoría de herramientas que permiten visualizar las salidas del modelo WRF no están diseñados para el procesamiento en paralelo. Esto, a priori, no es un problema ya que el coste de ejecutar un script para visualizar el contenido de una variable meteorológica no es elevado. El problema aparece cuando se necesita visualizar el valor de múltiples variables, en múltiples instantes de tiempo.

Para hacer esto, un script de postprocesado, tiene que recorrer todo el fichero netCDF. Aunque el acceso a este tipo de ficheros está muy optimizado, la cantidad de información que contienen es muy grande. Por ejemplo, las salidas de una predicción del GFA ocupan en torno a los 7,0 GB.

La figura 3 muestra un ejemplo de script secuencial simplificado en GrADS⁸ utilizado por el GFA para generar gráficas de temperatura y viento en superficie. Como puede observarse en el script, se recorre todo el fichero netCDF iterando entre los diferentes frames de tiempo. En 48 horas de predicción a intervalos de 6 horas se obtienen 8 *frames* de tiempo. En cada uno de estos *frames*, se genera una gráfica temperatura y viento en superficie similar a la de la figura 4.

Todo esto puede complicarse mucho si se necesita, como suele ser habitual, intervalos de tiempo más pequeños y más de una variable meteorológica.

El tiempo que puede llevar la generación de gráficas que permitan visualizar toda la información del modelo puede ser considerablemente grande respecto al tiempo que lleva la ejecución paralela del modelo. La tabla I muestra un listado de tiempos de ejecución de algunas predicciones operativas que ejecuta el GFA, separando por un lado la ejecución paralela del modelo, y por el otro el procesado secuencial de las salidas producidas por el mismo. Como puede observarse, el tiempo que lleva procesar las salidas es considerablemente grande respecto al tiempo de ejecución del modelo. Sobre todo si se tiene en cuenta que para ejecutar el modelo, se utilizan 128 nodos

⁸La versión completa del script se puede encontrar en <http://193.146.101.113/toolbox/postproc.gs>.

⁷Incluidos en las librerías netCDF.

```

netcdf wrfout_d01_2011-12-22_00:00:00 {
  dimensions:
    Time = UNLIMITED ; // (17 currently)
    west_east = 122 ;
    south_north = 111 ;
    bottom_top = 34 ;
    ...
  variables:
    float T(Time, bottom_top, south_north, west_east) ;
      T:description = "perturbation potential temperature (theta-t0)" ;
      T:units = "K" ;
    float RAINC(Time, south_north, west_east) ;
      RAINC:description = "ACCUMULATED TOTAL CUMULUS PRECIPITATION" ;
      RAINC:units = "mm" ;
    ...
  // global attributes:
    :TITLE = " OUTPUT FROM WRF V3.1.1 MODEL" ;
    :START_DATE = "2011-12-22_00:00:00" ;
    ...
  data:
    T =      -7.835083, -7.864655, -7.891296, -7.904358, -7.911102, -7.931854,
            -7.968079, -8.019257, -8.089844, -8.18219,  -8.287872, -8.430267, ... ;
    RAINC = 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
            0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ... ;
    ...
}

```

Fig. 2. Representación en modo texto de un fichero netCDF.

```

inputPath = '/home/gfa_1/gfa_1_2/output/'
outputPath = '/home/gfa_1/gfa_1_2/output/'

filename = inputPath%'outputd01.ctl'
'open '%filename
i = 1
while (i <= 8)
  t2mw10m(1, i, outputPath%'dom1_snh')
  i = i + 1
endwhile
'close 1'
'quit'

function t2mw10m(domain, iteration, outputFile)
  domainData(domain)
  'set mpt 1 1'
  'set mpt 2 1'
  'set mpt 3 1'
  'clear'
  'set t '%iteration + 1
  vel()
  saveFile(domain, iteration, outputFile,
    'Viento a 10m y Temperatura a 2m (°C)', 0)
return
...

```

Fig. 3. Script GrADS de postprocesado de las salidas de WRF.

del *cluster* de cálculo paralelo del centro de supercomputación de Castilla y León.

V. PROCESAMIENTO EN PARALELO

A la vista de lo comentado anteriormente, existe un punto de mejora importante dentro de un pronóstico operativo en el procesado de las salidas. El modelo está preparado para su ejecución en entornos de memoria compartida y distribuida, pero no hay implementaciones que permitan añadir algún

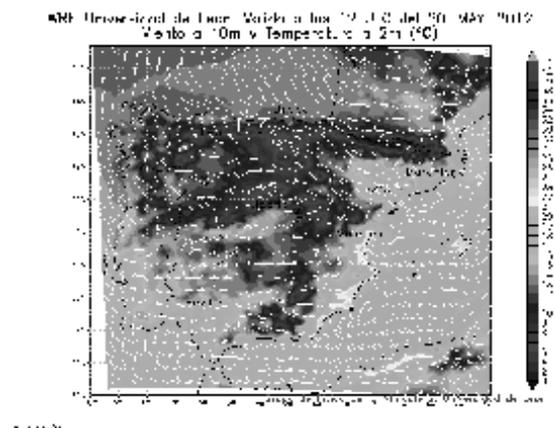


Fig. 4. Gráfica de temperatura y viento en superficie.

grado de paralelismo al procesado operativo⁹ de las salidas.

Descomponiendo el problema, *procesar la información de múltiples variables meteorológicas en múltiples instantes de tiempo*, en múltiples problemas individuales, donde solamente se procesa la información de una única variable, en un determinado instante de tiempo, se obtiene una primera aproximación paralela a la solución.

A. Primeros pasos

El hecho de que las salidas de WRF cumplan con las especificaciones netCDF es de gran utilidad para el enfoque paralelo planteado. Los ficheros netCDF están diseñados para que varios procesos puedan leer de ellos de manera simultánea. Dicho esto, el primer

⁹Entendiendo *procesado operativo* como el procesado regular de múltiples variables meteorológicas en múltiples instantes de tiempo. Para casos de estudio donde solo se analice una variable puede no ser necesario.

TABLA I

TIEMPO DE EJECUCIÓN DEL MODELO Y DEL POSTPROCESADO SECUENCIAL DE LAS SALIDAS.

Simulación operativa	Modelo	Postproceso
14/05/2012 12:00 UTC	0:51:01	0:17:10
15/05/2012 0:00 UTC	0:51:06	0:19:07
15/05/2012 12:00 UTC	0:50:24	0:18:11
16/05/2012 0:00 UTC	0:49:38	0:17:20
16/05/2012 12:00 UTC	0:48:20	0:18:29
17/05/2012 0:00 UTC7	0:49:05	0:16:59
17/05/2012 12:00 UTC	0:50:33	0:18:34
18/05/2012 0:00 UTC	0:49:53	0:18:26
18/05/2012 12:00 UTC	0:49:16	0:17:41
19/05/2012 0:00 UTC	0:50:08	0:19:24

punto que es preciso abordar es la adaptación de los scripts de postproceso secuencial para poder utilizarlos en la aproximación paralela.

Como se ha visto, con un enfoque secuencial, se lee el fichero completo y van generando las gráficas una a una. Ahora es necesario acceder a unos datos concretos dentro del fichero y generar la gráfica que corresponda. A priori parece una simplificación, pero lo cierto es que es preciso modificar los scripts de tal manera, que acepten una serie de parámetros que permitan acceder a una determinada sección dentro de una variable concreta. Es recomendable utilizar un lenguaje que permita esto de una manera sencilla, como NCL.

Los parámetros que es necesario pasar al script incluyen:

- La ruta del fichero netCDF que contiene la información.
- La variable meteorológica que se desea leer.
- El intervalo de tiempo (*frame*) para el que se desea conocer el valor de la variable.
- El nivel de presión cuando se trate de variables que tengan valores en la vertical.
- El nombre del fichero de salida que contendrá la gráfica y formato del mismo, normalmente **png**.

La figura 5 muestra un ejemplo sencillo¹⁰ de como pasar parámetros a un script NCL y utilizarlos para generar una gráfica de presión en superficie.

Todos los scripts NCL que utiliza el GFA para obtener los valores de las variables meteorológicas objeto de estudio pueden consultarse en <http://193.146.101.113/toolbox>.

- Temperatura y geopotencial a 700, 500 y 350 hPa: **tc_geopt.ncl**.
- Temperatura y viento en superficie: **t2_w10.ncl**.
- Humedad relativa y viento a 700, 500 y 350 hPa: **rh_wind.ncl**.
- Presión en superficie: **slp.ncl**.
- Precipitación y precipitación acumulada: **rain.ncl**.

¹⁰El script completo puede descargarse en <http://193.146.101.113/toolbox/slp.ncl>.

```

begin
  if (.not. isvar("inputFile")) then
    print("Input file not defined")
    exit
  end if
  if (.not. isvar("outputFileType")) then
    print("output file type not defined")
    exit
  end if
  if (.not. isvar("outputFileName")) then
    print("output file name not defined")
    exit
  end if
  if (.not. isvar("time")) then
    time = 0 ; Valor de tiempo por defecto
  end if

  ...

  ; Open the WRF ARW input file
  a = addfile(inputFile, "r")
  ; Open workstation
  wks = gsn_open_wks(outputFileType,
                    outputFileName)

  ...

  ; Get the variables from the WRF input file
  slp = wrf_user_getvar(a, "slp", time)

  ...

  ; Create graph
  contour_slp = wrf_contour(a, wks, slp, opts)
  ; Plot graph
  plot = wrf_map_overlays(a, wks, contour_slp,
                        pltres, mpres)
end

```

Fig. 5. Script NCL para calcular la presión en superficie en un instante determinado.

- Espesor de nieve: **snowh.ncl**.
- Agua en forma de nieve: **snow.ncl**.

Una vez adaptados los scripts para poder generar gráficas, de una única variable, en un instante de tiempo determinado, es necesario que construir un *array* de trabajos individuales para ejecutar en un *cluster* de cálculo paralelo.

Como hacer esto depende del gestor de colas que se utilice en el entorno de ejecución. El centro de supercomputación de Castilla y León utiliza el gestor de colas SGE. La figura 6 muestra una implementación en python¹¹ de un *array* de trabajos para SGE. La idea consiste en construir una lista de elementos, cada uno de los cuales tiene la información necesaria para ejecutar los scripts de postproceso, fichero de entrada, variable meteorológica, *frame* de tiempo, etc. En función del índice que el gestor de colas asigne a un nodo determinado, del total de nodos incluidos en el trabajo, este tomará los parámetros de la lista. De tal manera que si se tiene igual número de nodos, que elementos tenga la lista, cada nodo genera únicamente una gráfica para un instante determinado.

En la tabla II puede consultarse el tiempo de eje-

¹¹La versión completa puede consultarse en <http://193.146.101.113/toolbox/postproc.py>.

```

joblist = []

# Temperatura y viento en superficie (frame 0)
joblist.append({'inputfile' : input_path + '/wrfout_d01.nc' ,
               'outputfile' : output_path + '/dom1_t2m-w10m_00' ,
               'script' : 't2_w10.ncl' ,
               'args' : '\time=0\'}

# Temperatura y geopotencial a 850 hPa (frame 0)
joblist.append({'inputfile' : input_path + '/wrfout_d01.nc' ,
               'outputfile' : output_path + '/dom1_tmp-gpt_850hPa_00' ,
               'script' : 'tc_geopt.ncl' ,
               'args' : '\time=0\ \'levels=(/850./)\'}

# Presión en superficie (frame 0)
joblist.append({'inputfile' : input_path + '/wrfout_d01.nc' ,
               'outputfile' : output_path + '/dom1_pslv_00' ,
               'script' : 'slp.ncl' ,
               'args' : '\time=0\'}

# Humedad relativa y viento a 850 hPa (frame 0)
joblist.append({'inputfile' : input_path + '/wrfout_d01.nc' ,
               'outputfile' : output_path + '/dom1_hum-wind_850_00' ,
               'script' : 'rh_wind.ncl' ,
               'args' : '\time=0\ \'levels=(/850./)\'}

# Espesor de nieve en superficie (frame 0)
joblist.append({'inputfile' : input_path + '/wrfout_d01.nc' ,
               'outputfile' : output_path + '/dom1_snowh_00' ,
               'script' : 'snowh.ncl' ,
               'args' : '\time=0\'}

...

i = int(os.environ['SGE_TASK_ID']) - 1 # Índice dentro del array de trabajos
runNclCommand(joblist[i]['inputfile'], joblist[i]['outputfile'], joblist[i]['script'],
              args = joblist[i]['args'])

```

Fig. 6. Script python para ejecutar un *array* de trabajos en SGE.

cución del procesamiento paralelo de algunas predicciones operativas del GFA ejecutadas en el *cluster* del centro de supercomputación de Castilla y León. Si se comparan estos números con los de la ejecución secuencial de la tabla I se puede observar claramente la mejora.

TABLA II
TIEMPO DE EJECUCIÓN DEL POSTPROCESADO PARALELO.

Simulación operativa	Tiempo medio
0:00 UTC	-
6:00 UTC	-
12:00 UTC	-
18:00 UTC	-

B. Enfoque final

Aún es posible dar un paso más y obtener un resultado mejor. Si bien los ficheros netCDF están diseñados, y optimizados, para que múltiples lectores puedan leer de un único fichero, parece claro que el acceso será más rápido si cada ejecución de los scripts de postproceso accede a un fichero netCDF diferente.

WRF dispone de muchos parámetros que es posible modificar, editando el fichero de configuración `namelist.input`, y que permiten alterar su comportamiento. Entre ellos hay uno, denominado

`frames_per_outfile`, que permite decidir cuantos *frames* de tiempo se añaden en cada fichero de salida. El valor por defecto para este parámetro es 1000. Esto quiere decir que hay 1000 *frames* de tiempo por cada fichero de salida. O lo que a efectos prácticos viene a ser lo mismo, que todos los frames de tiempo de la predicción están en un único fichero de salida.

Si modificamos el valor por defecto e introducimos 1, WRF que solamente guarda un *frame* de tiempo en cada fichero de salida que genera, y de esta forma se tiene un fichero netCDF por cada *frame* de tiempo.

Con este nuevo enfoque no es necesario cambiar los scripts de postproceso, solamente los parámetros que hay que pasarles, y para ello basta con modificar el script que envía el *array* de trabajos al gestor de colas.

La tabla III muestra el tiempo de ejecución de una serie de predicciones operativas con postprocesado paralelo sobre ficheros independientes. Comparando estos números con los que aparecen en la tabla II, es posible comprobar la ganancia de este nuevo enfoque.

VI. CONCLUSIONES

El postprocesado operativo en paralelo de las salidas del modelo de predicción numérica WRF es un problema complejo que obliga replantear el proceso completo.

Por un lado, hay que revisar la ejecución del pro-

TABLA III

TIEMPO DE EJECUCIÓN DEL POSTPROCESADO PARALELO SOBRE FICHEROS INDEPENDIENTES.

Simulación operativa	Tiempo medio
0:00 UTC	-
6:00 UTC	-
12:00 UTC	-
18:00 UTC	-

pio modelo, cambiando algunas opciones de configuración.

Por otro, también hay que re-pensar y re-implementar, realizando los cambios necesarios, el procesado de las salidas, para adaptarlo al paradigma de la programación paralela.

La tabla IV muestra los tiempos medios de ejecución del procesado de las salidas del modelo siguiendo los tres métodos utilizados:

- Procesado secuencial de las salidas.
- Procesado en paralelo de las salidas.
- Procesado en paralelo sobre ficheros de salida independientes.

TABLA IV

TIEMPOS MEDIOS DE EJECUCIÓN DEL PROCESADO DE LAS SALIDAS DE WRF.

Ejecución secuencial	Ejecución paralela	Ejecución paralela Ficheros independientes
0:51:01	0:51:01	0:51:01

Para calcular los tiempos medios, en cada uno de los casos, se han utilizado los datos de las tablas I, II y III respectivamente.

Si añadimos a estos tiempos medios, el tiempo medio de ejecución del modelo (0:51:01), calculado en base a los datos de la tabla I, que será el mismo en los tres casos, obtenemos el tiempo global de ejecución que se muestra en la tabla V

TABLA V

TIEMPOS MEDIOS DE EJECUCIÓN GLOBAL.

Ejecución secuencial	Ejecución paralela	Ejecución paralela Ficheros independientes
0:51:01	0:51:01	0:51:01

La reducción del tiempo global de ejecución, que incluye el modelo y el procesado de las salidas, es evidente. A la vista de estos datos, se observa que la hipótesis inicial, donde se afirma que el procesado de las salidas en paralelo mejora el rendimiento global, se cumple.

Por tanto, podemos afirmar que es recomendable acometer la paralelización del procesado de las salidas del modelo de predicción numérica WRF, cuando queremos visualizar los valores de múltiples variables meteorológicas, en múltiples instantes de tiempo.

AGRADECIMIENTOS

Todas las pruebas y ejecuciones operativas que realiza el GFA se realizan en el *cluster* de cálculo paralelo del centro de supercomputación de Castilla y León.

REFERENCIAS

- [1] García-Ortega, E., Fita, L., Romero, R., López, L., Ramis, C., Sánchez, J. L., *Numerical simulation and sensitivity study of a severe hailstorm in northeast Spain*, Atmospheric Research, vol. 83, pp. 225-241, 2007.
- [2] García-Ortega, E., López, L., Sánchez, J. L., *Diagnosis and sensitivity study of two severe storm events in the Southeastern Andes*, Atmospheric Research, vol. 93, pp. 161-178, 2009.
- [3] García-Ortega, E., Merino, A., López, L., Sánchez, J. L., *Role of mesoscale factors at the onset of deep convection on hailstorm days and their relation to the synoptic patterns*, Atmospheric Research, in press.
- [4] W.C. Skamarock, J.B. Klemp, J. Dudhia, D.O. Gill, D.M. Barker, W. Wang, J.G. Powers, *A description of the Advanced Research WRF Version 2*, NCAR Tech., Note NCAR/TN-468+STR, 2005.
- [5] *The Weather Research and Forecasting (WRF) Model website*, <http://www.wrf-model.org/>.
- [6] *ARW Version 3 Modeling System User's Guide*, National Center for Atmospheric Research (NCAR), Mesoscale & Microscale Meteorology Division, 2010.
- [7] *Network Common Data Form (netCDF) website*, <https://www.unidata.ucar.edu/netcdf>.
- [8] Ed Hartnett, *The NetCDF Tutorial*, Unidata Program Center, 2011
- [9] *GrADS website*, <http://www.iges.org/grads/>.
- [10] *CISL's NCAR Command Language (NCL) website*, <http://www.ncl.ucar.edu/>.
- [11] *NCAR Command Language (NCL). Mini-Language Reference Manual*, National Center for Atmospheric Research (NCAR), 2011.
- [12] *NCAR Command Language (NCL). Mini Graphics Manual*, National Center for Atmospheric Research (NCAR), 2011.