

Implementation of the WRF-ARW prognostic model on the Grid

D. Davidović, K. Skala

Centre for Informatics and Computing

Ruder Bošković Institute

Bijenička Street 54, Zagreb, 10000, Croatia

Phone: (+385) 1-456 1091 E-mail: ddavid@irb.hr, skala@irb.hr

Abstract – The numerical weather prediction models demand large execution time and resource allocation; therefore we need large parallel computation facilities, like the grid systems, in order to execute them. In this paper the implementation of the Advanced Weather Research and Forecasting (WRF-ARW) model on the SEEGRID-SCI grid infrastructure is presented. The goal of porting the model to the grid is to get more accurate and detailed forecast with operational speed-up on high-resolution model grids. The results and the application of the WRF-ARW model on the Grid will be shown in the example of the bora winds simulations over the Adriatic coast. We have achieved significant progress in accelerating the model execution and simplicity of starting jobs on the grid. In our researches we have found that the benefits are not only in the reduced execution time (up to 30% for large enough number of CPUs) but also in the simultaneous runs of the same model. This is especially important in the research activities when many similar jobs can be executed simultaneously. We have achieved a great success in porting the WRF-ARW model to the grid and the improvements gained through deployment of the model increase the support and feasibility of new projects in the area of the SEE-GRID-SCI infrastructure and Earth Science community. The implementation of the WRF-ARW model on the grid infrastructure increases scientific research in regional scale particularly in the field of ecology.

I. INTRODUCTION

Atmospheric science of today is advancing towards very complex phenomena at ever smaller temporal and spatial scales. These phenomena have large impacts on both scientific research and operational weather forecast. One of the principal tools utilized in these efforts are numerical weather prediction models.

A numerical weather prediction model is a computer program that produces meteorological information for future times at given regions. The horizontal domain of a model is either global, covering the entire Earth, or regional, covering only part of the Earth. Regional models are also known as limited-area models. Some of the widely used limited-area models are BOLAM, MM5, WRF-ARW, WRF-NMM, ALADIN, RAMS and ARPS. These numerical models usually demand large execution times and resource allocations; therefore, we need large parallel systems in order to execute them. There are many differences between these models. Some models are better predictors for wind simulations, other for cloud simulation, and some are faster than the others, while the WRF-ARW is the best for simulations over the complex terrain, just as the terrain in Croatia and the South-East Europe is.

A comparison of the WRF performance with the performance of the MM5 model is shown in “Fig. 1” and described in [8].

In terms of time-per-timestep, the WRF simulation is considerably more costly, but one can see in “Fig. 1” that the time-to-solution performance for WRF is slightly better than for MM5 and should improve with tuning and optimization [8].

The aim of our research aims is reproduction/forecasting of the airflow over complex terrain (e.g. mountain ranges, coastal area) in Croatia. We are faced with the problem of the lack of computing resources due to the high-resolution model grids, necessary for more accurate and detailed results. Using high model resolution results in two natural consequences: an enlargement of the number of computational grid points, so to keep the total domain size sufficiently large; and reduction of the integration time step in order to avoid numerical instabilities. Both effects result in larger CPU and memory demands and a big storage space that cannot be provided by the ordinary resources available to the user community.

Till today, not much effort has been done in adopting WRF-ARW model to the grid infrastructure. The model has mostly been adopted, with relatively good success, for the execution on local clusters and, with great success, on supercomputers like in NCEP (National Centre for Environmental Predictions) or NCAR (National Centre for Atmospheric Research). Due to the lack of financial resources and user access to such large supercomputers we have searched for other options and found that the grid can be a suitable choice. The grid infrastructure has already been utilized for other purposes and is ready to be exploited by numerous different science applications from different user communities.

The WRF-ARW model porting to the grid is financed by the EU FP7 SEE-GRID-SCI project. The SEE-GRID-SCI (SEE-GRID eInfrastructure for regional eScience: <http://www.see-grid-sci.eu/>) is the project which aims to enlarge the regional eInfrastructure to cater for demands of the communities by increasing the computing and storage resources and involving new partner countries in the region of the South-east Europe with an emphasis on strategic research groups in seismology, meteorology and environmental protection.

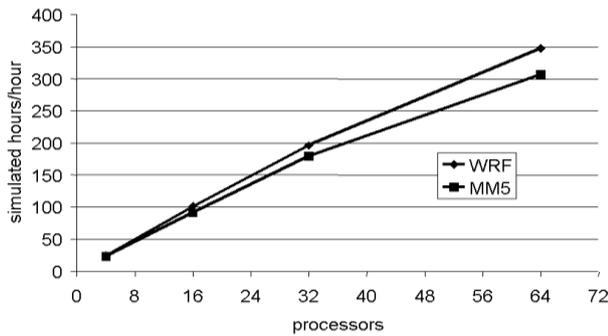


Fig. 1. WRF performance compared with MM5

II. THE WRF-ARW PROGNOSTIC MODEL

The WRF-ARW model (Weather Research and Forecasting model – Advanced Research WRF) is a next-generation numerical weather prediction system designed to serve both operational forecasting and atmospheric research needs. It features a software architecture allowing for computational parallelism and system extensibility. WRF-ARW is suitable for a broad spectrum of applications across scales ranging from meters to thousands of kilometres.

The equation set for ARW is fully compressible, Eulerian and non-hydrostatic with a run-time hydrostatic option. It is conservative for scalar variables. The model uses terrain-following, hydrostatic-pressure vertical coordinate with the top of the model being a constant pressure surface. The horizontal grid is the Arakawa-C grid. The time integration scheme in the model uses the third-order Runge-Kutta scheme, and the spatial discretization employs 2nd to 6th order schemes. The model supports both idealized and real-data applications with various lateral boundary condition options. The model also supports one-way, two-way and moving nest options. It model supports parallel execution using MPI (Message Parsing Interface) and/or OpenMP (Open Multi-Processing).

“Fig. 2” depicts the component programs of the WRF Modelling System. The WRF model can be run with either idealized initialization or real-data initialization. The WRF model supports the Eulerian mass solver, referred to as the

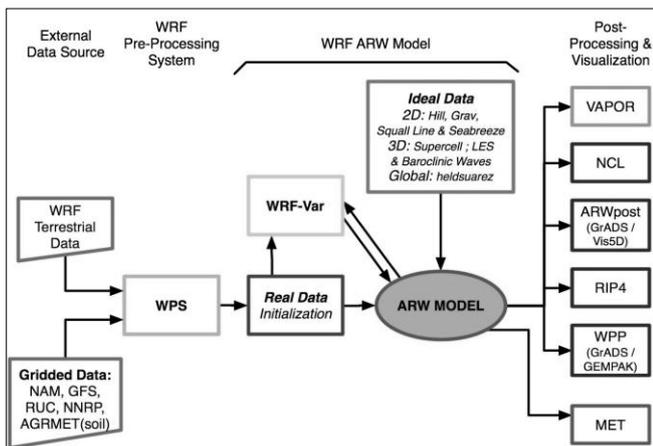


Fig. 2. WRF-ARW modelling system flow chart
Picture taken from *WRF model users page*,
<http://www.mmm.ucar.edu/wrf/users/model.html>

advanced research WRF (ARW) dynamical solver. The purpose of the ideal.F (pink) and real_em.F (blue) programs is to generate input and (if necessary) boundary files for the WRF model. This involves a hydrostatic balance adjustment in addition to setting up the initial 3d and 2d fields of the WRF variables. The WRF model supports full physics, two-way, one-way and two-way moving nests, analysis and observation nudging.

The function of the WRF Pre-processing System (WPS) is to define WRF grid, generate map, elevation and land information for WRF, take real-data analyses/forecasts from another model, and interpolate the data to the WRF grid. The time-dependent (analysis) fields consist of 3d wind, potential temperature, and water vapour, and a number of 2d fields.

WRF-Var (which supports 3D-VAR capability in the current release) can be used to assimilate observations into model initial conditions (we are not using this module in our research).

The standard output from WPS, real, and WRF model is in netCDF format [9] [10] (one of WRF model input/output format) that can be displayed by the standard graphic tools like NCAR Graphics NCL [11], GrADS [12], or RIP4 [13].

III. GRID IMPLEMENTATION

A. Implementation idea

The grid system is a very complex system and most of the users, especially the new ones, need a lot of effort and knowledge to learn how to use the system and prepare their jobs for execution. Therefore, our primary idea was to develop an easy to use implementation of the WRF-ARW model that will hide the grid complexity from the end users.

In order to simplify the execution on the grid, the WRF-ARW is divided into three different stages: pre-processing, core model execution and post-processing.

The first stage incorporates downloading real-data analyses/forecast from another (global or local) model and preparing the model input data as WRF grid, map generation, elevation and land information, and interpolating the data to the WRF grid, that are the terrestrial data and the initial boundary conditions data. The second one enables running of the model on a grid Computing Element i.e. grid cluster component and the third one operate with the model output (produced) data.

The basic scheme is to store all the WRF model data, including binaries and input and output model files, on the grid storage elements, while the execution scripts are installed in the user’s home folder on the User Interface node. With the submission process, the execution scripts are sent to the grid CE, the necessary model input files and binaries are downloaded and the model execution is started. After the model execution is finished all the results are stored on the grid storage elements from where users can download them directly to the UI node or local computer.

B. Storage structure

Storage scheme is, together with the model grid operating scripts, the most important part of the model grid implementation, so we will describe it in more details.

Due to the project requirements and different users' communities using the model, the users are divided into regions of interest by regional affiliation. Now we have 5 different regions: Armenia, BiH (Bosnia and Herzegovina), Croatia, Georgia, and SEEurope as the default region.

In order to be able to use model on the grid, the users first have to be members of the meteorology VO established in the SEEGRID-SCI project. As a part of the meteorology VO, WRF is using tools for data management provided by the VO. The tool for the data management and replica file system is the LFC (LCG File Catalogue) [14]. The root folder for the WRF-ARW model in the LFC catalogue is:

```
/grid/meteo.see-grid-sci.eu/WRF-ARW
```

Hereinafter the "root" will mark the root folder for WRF-ARW in the LFC file catalogue. The root folder contains three sub-folders containing the model binaries, model input data and the folder in which are stored the model output data: "bin", "input_data" and "output_data", respectively. The whole model storage scheme is shown in "Fig. 3".

The following subsections describe 3 main subdirectories.

1) Bin folder

The "bin" folder contains all the binaries and auxiliary files for pre-processing ("WPS" folder), the core application binaries ("WRF" folder), and the post-processing binaries and visualization tools ("postproc" folder). Also, the binaries for different model versions are stored in this folder. The users can fetch and store binaries by using the LFC [14] tools for files manipulation.

```
/grid/meteo.see-grid-sci.eu/WRF-ARW
  /bin
    /WPS
      /WPSV<version_number>_<arch>
    /WRF
      /WRFV<version_number>_<arch>
    /postproc
  /input_data
    /boundary
      /<REGION>
      /raw
      /precompiled
    /terrain
  /output_data
    /<REGION>
      /sci
      /oper
```

Fig. 3. The WRF-ARW model LFC storage structure

2) Input data folder

In the "input_data" folder the data for WRF pre-processor (WPS) are stored. There are the static terrestrial data that are not changing between different runs, stored in the "terrain" folder, and the time-dependent initial and boundary conditions in the "boundary" folder (which are the input for the `ungrib.exe` application in the WPS pre-processing system). The initial and boundary conditions are divided between regions of interest because these data are region-specific and to prevent unnecessary mixing of data between different users and groups.

3) Output data folder

The "output_data" folder contains all the outputs after the model has finished its execution on the grid. These are the data produce by the model in NetCDF file format and the standard model input, output and error files. Data are separated depending on the region of interest and depending on weather the model was run in one of the two possible modes: scientific or operational, with the "sci" and "oper" folders respectively.

C. Model workflow

The model workflow is described on the "Fig. 4". The main entrance gate to the grid is UI node where the model execution scripts are installed.

The concept in a nutshell is that the user, working on UI node, submits model (i.e. start WRF bash scripts) to the grid WMS (Workload Management System). WMS allocates the grid CE (Computing Element) i.e. grid cluster to our job, and before the model start to execute on the allocated CE, the input terrestrial and initial/boundary conditions data, as well as the WRF binaries, are downloaded from the grid SE (Storage Element). When all input data and binaries are downloaded to CE, the model starts with the execution using the MPI (Message Passing Interface) [15] runtime environment.

The model execution starts with the pre-processing system (WPS), described in the section "The WRF-ARW prognostic model". The WPS consists of three applications: *geogrid*, *ungrib*, and *metgrid*. These three applications are run in serial mode one-by-one starting with *geogrid* and finishing with *metgrid* application. The final product of the WPS system is the NetCDF file with the horizontally interpolated boundary conditions data onto the simulation domain.

When the pre-processing is finished the core application starts first with the real application which sets up vertical model levels for the model input and boundary files and finally the *wrf* application for numerical integration. Both applications are run using MPI on multiple CPUs/cores. The WRF-ARW workflow on a grid CE is illustrated in "Fig. 5".

After successful completion, the output data are stored on a pre-defined grid SE. User can download model output data from SE using the LCG-tools, DM-Web application service developed within SEEGRID project or by using developed scripts for automatic WRF model's data collecting.

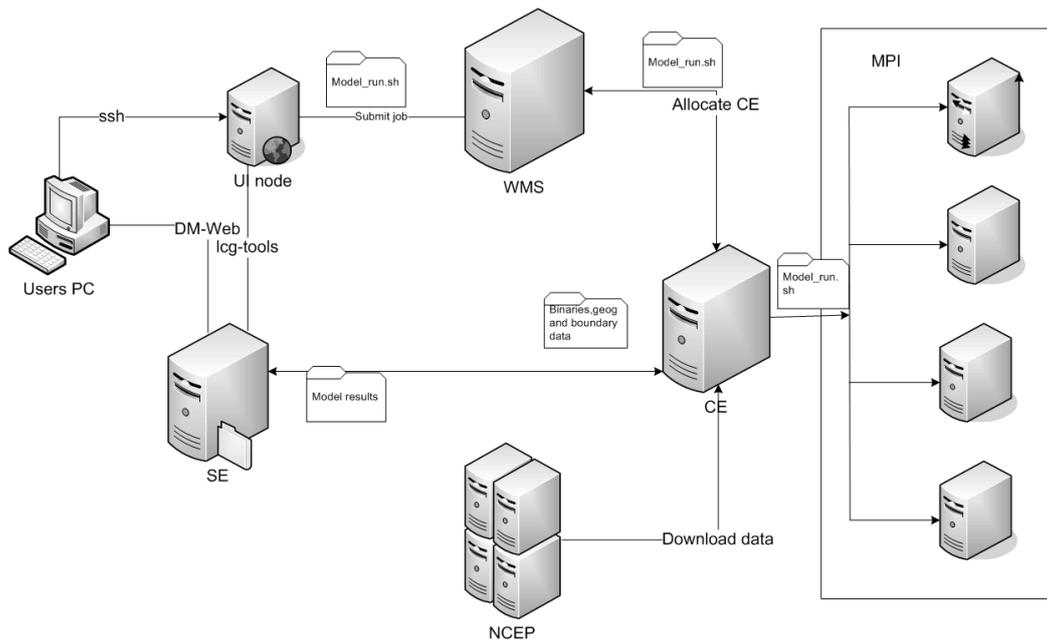


Fig. 4. The WRF-ARW model grid workflow

D. Starting model

There are two different WRF-ARW model modes developed: scientific and operational mode. The scientific mode is intended to be used for research purposes, where the users have full control over all setting and relevant model parameters while the operational mode is intended to be used for daily operational forecast with most of the model settings already pre-defined.

1) Scientific mode

The scientific mode is used for the research purposes when the users do not need current boundary conditions data downloaded from NCEP, ECMWF (European Centre for Medium-Range Weather Forecasts) or any other large-scale forecasting centre. Users are expected to manually upload and save their initial/boundary conditions data at the LFC catalogue in the input data folder for their region (/boundary/REGION/raw/ folder on the “Fig. 3”).

As in a serial, non-grid execution of the model, user has to describe his model using the WRF files, namelist.wps and namelist.input. The user needs to define all relevant variables in these two files so that the model can execute normally.

The model is executed the by starting the command wrf-submit. An example of how to submit model in the

scientific mode, for the region “CROATIA” and using 16 CPUs is:

```
$wrf-submit -m s -r CROATIA -p 16
```

This command automatically checks for the namelist files in the current folder, sets the environment variables, generates JDL job description files and submits the jobs using glite-wms-job-submit command from the gLite tools. The output of the wrf-submit command is a file that contains user’s job ID required to check the job status and to retrieve the output when the job finishes. After the execution has finished, the model output data are stored using the LCG-tools in the LFC file scheme in the folder:

```
/WRF-ARW/output_data/<REGION>/sci/<file_name>
```

The user can retrieve the output data on the UI node using the lcg-cp tool (which is the standard LFC tool to copy grid files to UI) or the program wrf-get-data that collects the output of the finished jobs and download the model output results from the grid CE (LFC) to UI.

2) Operational mode

The operational mode is designed and adjusted to be as simple as possible and not to require much of user intervention for everyday automatic executing. Users have to adjust all model variables in the namelist files, while the start and end date of the simulation are set automatically (the start date is set to the current date). When executing in the operational mode, the input initial/boundary conditions data are downloaded automatically from the NCEP server on the grid CE and the user’s voms proxy certificate is created automatically. An example of starting the operational forecast, using 8 processors, with the simulation time of 48 hours (2 days weather forecast) and “BIH” region would be:

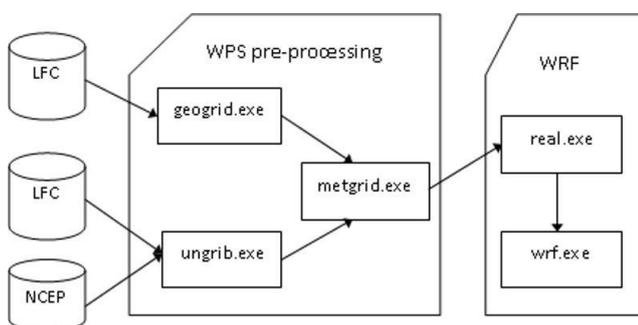


Fig. 5. The WRF-ARW grid CE workflow

The output of the command is a file containing the job identification number. In the operational forecast, most of the parameters are set by default, e.g. simulation time period (48 hours in the above example) or number of processors.

IV. EXAMPLE OF GRID USAGE AND RESULTS OBTAINED

The WRF-ARW model shows very good results for the simulations over the complex terrain. Therefore we have decided to use the WRF-ARW model for the research simulations over the Croatia, especially the coastal area (Istria peninsula, Kvarner Bay, and Dalmatia).

The research simulation are usually very time-consuming and resource expensive. One of such simulations is the wind simulations alongside Adriatic coasts (central and south Dalmatia). The simulations were run concurrently, several times (simultaneously on the different computing elements), allocating 16 CPUs each. The WRF-ARW model version was 3.1.1, 64bit. The results are described in

the next paragraphs.

“Fig. 6” shows the 10-m wind field on 24 April 2006 at 1300 UTC (that corresponds to 1400 h of local time). During the early afternoon, a sea breeze flow reached its maximum intensity. Over the land, the wind is rather irregular due to very complex topography. Wind speeds mostly varied in the range of 3.5 m s⁻¹ to 5 m s⁻¹ along the Adriatic coast, (e. g., Šibenik, Split, Makarska, and Dubrovnik). While at the land measuring sites the sea breeze development was from the southwest direction, island’s measuring sites showed the significant influence of the north-westerly large-scale wind.

V. CONCLUSION

The WRF-ARW model is well-know state-of-the-art numerical weather prediction model, suitable for weather simulations over complex terrain. The aim of our research was to perform simulations on high-resolution model grids, necessary for more accurate and detailed forecast. This approach results in enlargement of the CPU power, memory and storage capacity. We have shown that the grid infrastructure is sufficiently good for our research because

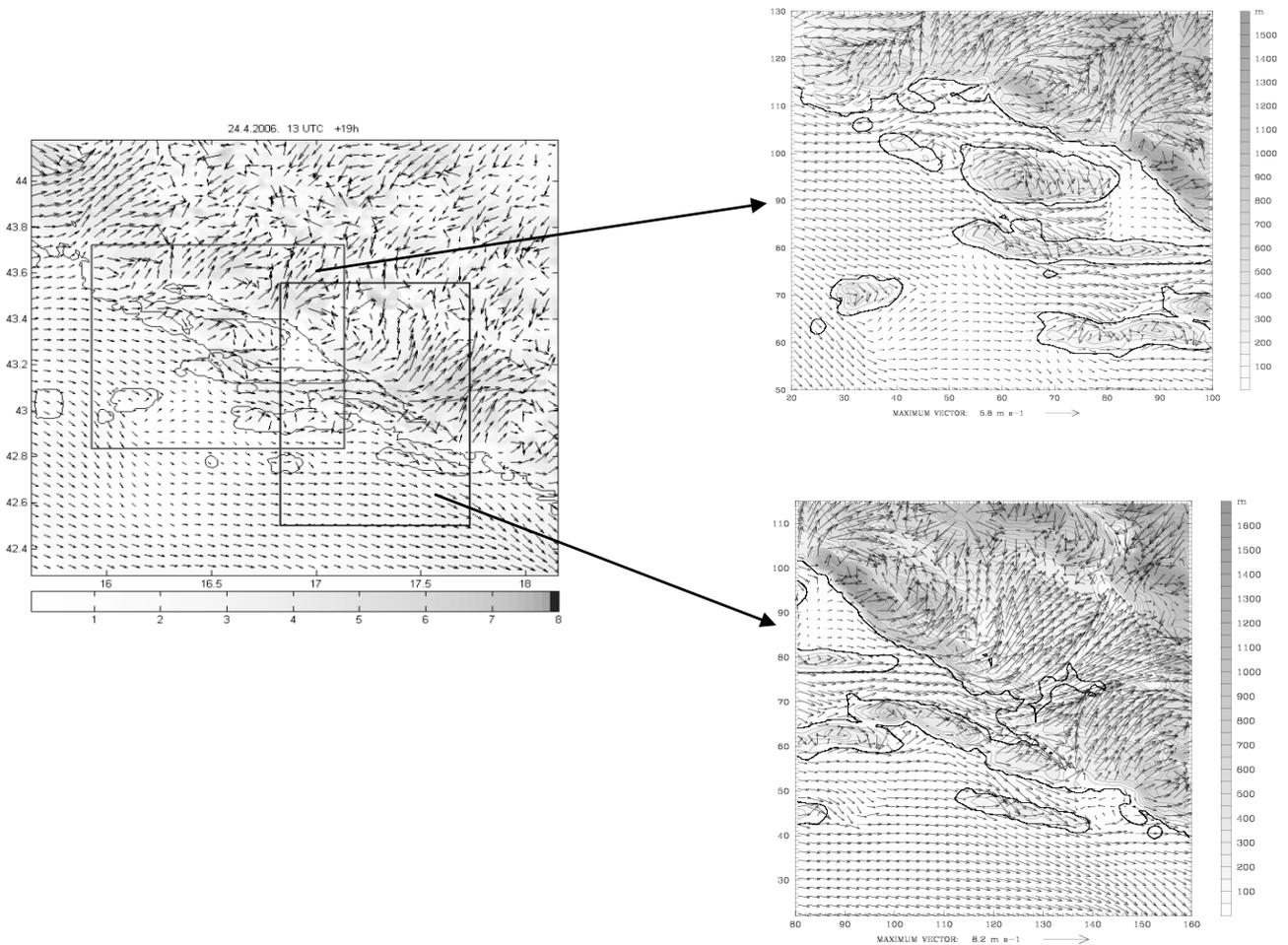


Fig. 6. Figure (upper) Modeled WRF 10-m wind vectors (m s⁻¹) for 24 April 2006 at 1300 UTC. The wind vectors are given at a horizontal resolution of 3 km. The wind speed is depicted by filled areas (shown by legend) with a 1 m s⁻¹ interval. The thick red and blue rectangles denote the horizontal cross-sections shown in the right panels. A close-up of the flow structure: (upper right) above the island of Brač and (left) above the wider Neretva area.

it offers a large number of available computing resources. We have also developed the model storage scheme using the LFC file catalogue for easier file and data manipulation on the distributed grid storage elements. The examples have shown that with large enough problems (e.g. increased number of the model grid points) we can get execution speedup. The conclusion is that the grid infrastructure is not optimal for small tasks with low computing resource demands (i.e. problems that can be easily solved on local computers), but rather for large tasks that require large CPU power and storage.

During our research we have concluded that the speed-up of the model execution is not only in the increase of a number of processors but also in ability to run multiple instances of the model with the different input data or the model description files. This is possible because of the availability of many grid computing elements where on each of them one instance of the WRF-ARW model can be run.

Therefore, we found that the grid infrastructure and our deployment of the WRF-ARW model on the grid is quite useful and shows large execution speed-up for weather research purposes for a wider meteorology user community. The model implementation on the grid is problem-independent and its intent is to attract other users to start running their time and resource consuming WRF-ARW application runs on the grid infrastructure where they can benefit mostly from enormous model execution speed-up and large storage space available for storing their data.

The next step in the process of WRF-ARW gridification will be an improvement of the grid execution for the operational forecast. In that sense, the model submission latency needs to be decreased together with the improvement of fault-recovery because of the grid heterogeneous nature.

Acknowledgments

This work has been funded by the EU FP7 SEEGRID-SCI project under the grant number 211338.

REFERENCES

- [1] The WRF-ARW model Tutorial:
http://www.mmm.ucar.edu/wrf/users/docs/user_guide_V3.1/contents.html
- [2] Michalakes, J., S. Chen, J. Dudhia, L. Hart, J. Klemp, J. Middlecoff, and W. Skamarock (2001): "Development of a Next Generation Regional Weather Research and Forecast Model" in *Developments in Teracomputing: Proceedings of the Ninth ECMWF Workshop on the Use of High Performance Computing in Meteorology*. Eds. Walter Zwiefelhofer and Norbert Kreitz. World Scientific, Singapore. pp. 269-276.
- [3] Michalakes, J. "Performance and Maintainability: Conflicting Concerns in Large Community Weather Models." University of Chicago Computational Institute, January 12, 2001.
- [4] Michalakes, J., R. Loft, A. Bourgeois (2001): "Performance-Portability and the Weather Research

- and Forecast Model" in *on-line proceedings of the HPC Asia 2001 conference*, Gold Coast, Queensland, Australia, September 24-28, 2001.
- [5] W.Wang, C.Bruyere, M. Duda, J. Dudhia, D. Gill, H.-C. Lin, J. Michalakes, S. Rizvi, and X. Zhang, "ARW Version 3 Modeling Systems User's Guide", Mesoscale&Microscale Meteorology Division and National Center for Atmospheric Research, January 2009
- [6] S. Burke, S. Campana, E. Lanciotti, P. Mendez Lorenzo, V. Miccio, C. Nater, R. Santinelli, and A. Sciaba, "gLite 3.1 User Guide", Manuals Series, CERN-LCG-GDEIS-722398, April 2009
- [7] "SEE-GRID Wiki", SEE-GRID Wiki – EGEE-see Wiki, [http://wiki.egee-see.org/index.php/SEE-GRID_Wiki]
- [8] Michalakes, J., J. Dudhia, D. Gill, J. Klemp and W. Skamarock: Design of a next-generation regional weather research and forecast model : Towards Teracomputing, World Scientific, River Edge, New Jersey, 1998, pp. 117-124.
- [9] NetCDF home page,
<http://www.unidata.ucar.edu/software/netcdf/>
- [10] NetCDF User documentation,
<http://www.unidata.ucar.edu/software/netcdf/docs/>
- [11] NCAR Graphics, <http://ngwww.ucar.edu/>
- [12] GrADS, <http://www.iges.org/grads/>
- [13] Rip4 home page,
http://www.mmm.ucar.edu/mm5/WRF_post/RIP4.htm
- [14] LFC file catalog documentation,
http://www.gridpp.ac.uk/wiki/LCG_File_Catalog
- [15] MPICH home page,
<http://www.mcs.anl.gov/research/projects/mpi/mpich1/>
- [16] D. Davidovic, D. Belusic, M. Telisman Prtenjak, 2009, "Grid implementation of the Weather Research and Forecasting model", *Proceedings of the SEEGRID-SCI User Forum 2009, Istanbul, Turkey, 9th – 10th December, 2010*, pp. 128-135