On-demand post-processing in modeling systems

G. Latini, G. Passerini & S. Tascini
Dipartimento di Energetica, Università Politecnica delle Marche, Italy

Abstract

We present the first results of a project aimed at developing and implementing an on-demand post-processing service based on a nonstop modeling system with the purpose of enhancing and optimizing the usability of simulation results.

The on-demand post-processing implementation is part of a general project we called AUTOMET. The project foresees the development of an autonomous system able to run nonstop simulations of pollution dispersion over a certain area by covering all the involved aspects from data collection to internet publication.

The on-demand approach has given encouraging results. It allows the user to access the widest range of data processed by a modeling chain from meteorological to diffusive parameters. Whoever accesses the service has the chance to visualize any datum or data set without any knowledge about its implementation or programming language.

The interface adopted is user friendly and gives the opportunity, even to non-meteorologists, to query and interpret graphics. Users who may be interested in this information-rendering approach are environmental policy managers, technicians for quick observations, and general public.

At present the AUTOMET project is limited to meteorological modeling. The model employed for this purpose is RAMS (Regional Atmospheric Modeling System) but a MM5 implementation for modeling comparisons is planned.

The RAMS model has been implemented on a parallel Linux platform following an open-source orientation of the whole project farther then Linux configuration capabilities. A GrADS is employed due to its diffusion and its batch programming features. Other technologies considered by this project are PHP and Javascript.

1   Introduction

"*Meteorology is a science among the most ancient. Notwithstanding, it is today one among the least diffused. And yet current and future weather are the sheer masters of human activity*". With these words, the late Edmondo Bernacca, one of the most popular Italian meteorologists, opens a famous meteorology handbook. We should add that analogue considerations are unfortunately true also for air pollution assessment and forecast.

The Province Authority of Ancona has supported a project carried on by the air pollution group of the Marche Politechnic University aiming to develop a management tool that presents information and divulgation opportunities along with a regional policy management tool. The project, discussed in more detail in the next section, foresees the development of an autonomous system able to continuously run simulations of pollution dispersion over a certain area covering all the involved aspects: from data collection to internet publication.

In this scenario, the post-processing and rendering aspects of model simulations represent an important issue. We recognized the rigid structure of data representation designed for popular application and recognized the opportunity for developing a tool useful for decision support as well as scientific monitoring. We had to face, once again [1], the interface issue. Actually, we believe that model interface development requires a solution similar to model development since reduce technical requirement in model use allowing a wider investigation of simulated phenomena.

The driving idea is to allow the user to elaborate dynamically information starting from given data on local meteorology and pollutant concentrations. We commonly refer to this general operation with the term of post-processing. The dynamic aspect is given by a system that asks the user which parameter to elaborate and how to represent them. No representation is cached (except in the user memory) and each one is re-elaborated in real-time

This experiment revealed interesting opportunities and several different applications might be foreseen.

2   The *Automet* project

The main purpose of *Automet* is to have an available and reliable set of meteorological data always ready to elaborate dispersion simulations or weather predictions.

The selected modeling system is RAMS (Regional Atmospheric Modeling System) by ATMET [9]. RAMS is a GNU licensed software representing the state of the art in mesoscale simulation. The system architecture is set up allowing multiple simulations and a MM5 simulation process is forthcoming, the only limit being that of available computational resources.

The computational aspect is approached employing a parallel execution of the mentioned models running on a Linux PC cluster. The system was run on a small six-nodes single-CPU (P4) cluster

The simulation area covers the Ancona Province, in the middle of Italy on the east coast. Simulations are performed over three nested grids with 80x80, 42x42 and 42x42 cells meshes for the duration of 24 hours (plus 3 "boot-strap" hours). The widths of the cells are respectively 25 km, 5 km and 1km. An extensive optimization of all processes, the typical computation time is about 1 hour and 20 minutes.

With reference to figure 1, it is clear that the analysis phase needs a data processing step [2]. The data collection is not always a straightforward process, of course. Since we have no local sounding system we refer to upper air data (provided by NCEP either as *objective avn data* or *re-analysed avn data* [3]4]) and surface monitoring station data. For the upper air data there are no problems for batch retrieval, since they are freely accessible through an anonymous FTP connection to a public server. Moreover, they are provided for the whole globe in standard *GRIB* format and, unsurprisingly, several data manipulation codes are available on the web to be employed at each step.

Local data, on the other hand, are more difficult to handle since their collection system was not designed for data publication. We can neither access direct monitoring stations nor the global database. Therefore, a remote procedure had to be installed in the Authority server in order to organize data in proper sets. We are still working on security aspects of Internet transactions.
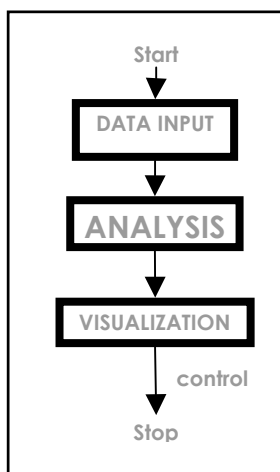

Figure 1:   The Automet modeling system process flow-chart.

On the other side of our elaboration chain, there is information representation. We opted for GRADS as default viewer [10]. The GRADS software has the fundamental feature of being programmable with a Fortran-like scripting language. The capability of writing, storing and recalling complex procedures (batch programming) is central in an automation process.

A final technical note regards information publication and access for which we opted to query pages written in standard HTML and PHP languages. Some more details are given the following sections.

Throughout the whole design phase, we always tried to meet existing standards and we always opted for open-source licensing, whenever available.

3 The on-demand post-processing approach design

3.1 Project specifics

The main purpose of AutoMet is to make information rendering more efficient. In the end, the AutoMet kernel processes a huge amount of numerical data and generates loads of numbers. To make loads larger, we run models over three different grids (related to Italy, Marche region and Ancona province respectively) and we must keep available at least some months of data to allow follow-up.

We can hardly imagine the requirements in terms of mass storage. Moreover, it is evident the rigid character of the stored information. As you would expect, images and numerical data cover different purposes and both are needed. Therefore, the fundamental idea is to maintain just the essential data files and afterwards to derive, on-demand, all necessary information.

Figure 2:   Display-parameter input-page.

A third aspect is the required user-friendly accessibility along with full rendering features. In order to comply with this aspect an intuitive web-based interface has been developed.

3.2   Features of the service

As a first step, the user shall choose the reference simulation that, as previously outlined, shows forecasted parameter of the following 24 hours. The system allows the user, having just a common web browser, to select the simulation grid by a drop-down box (see figure 1).

The last selection regards the parameters to be displayed. At the moment, it is possible to select two sets of 3D fields (such as wind, temperature, pressure, etc) and 2D data (such as mixing height, surface latent heat flux, Monin-Obukhov length, etc). A further confirmation is required in order to submit the query. As long as the fundamental options are confirmed, a dynamic configuration script is generated for the RAMS post-processor called REVU (RAMS/Hypact Evaluation and Visualization Utilities [5]). Then the post-processor will run according to the query and a GRIB format file will be produced [6].
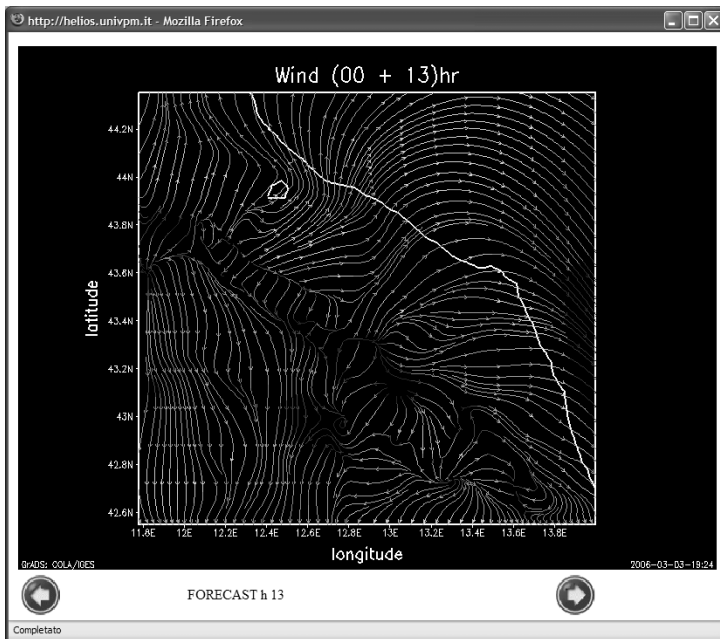

Figure 3:   Streamline representation in a restricted view of the middle grid.

At this point the user is requested to choose between downloading the requested data or to visualize them. If they opt for the second option, a utility will run converting GRIB data into a .ctl format file (a proprietary format of GRADS viewer). It might be argued that it is possible to produce a GRADS format file directly but the employment of a standard intermediate format derives from an optimization policy for future uses. Furthermore, the development of a function for ASCII-format file writing is required to have a ready-to-use product to be easily imported in spreadsheets and many other software applications.

Once the on-screen visualization has been selected, a page specialized for the requested parameter(s), is loaded. The page allows the user to input basic display parameters such as latitude and longitude span, initial and final altitude, initial and final forecast hour, type of representation (isolines, fields, vectors, etc). Figure 3 shows an example of visualization.

Interesting and intuitive controls have been introduced through the spatial parameters input. In fact, introducing identical values both for initial and final parameter the actual parameter is "fixed" and the correspondent slice is selected in the 3D field. If two parameters are fixed, the associated profile is displayed (e.g. in Figure 1 – profile representation with fixed latitude and altitude). Finally, fixing all three coordinates a single value is given.

As soon as all options are defined, the post processing standard output is reported at the end of the page and two action-buttons become visible; they give the possibility to show the newly generated images separately or as animation. In both cases, a pop-up window is loaded.
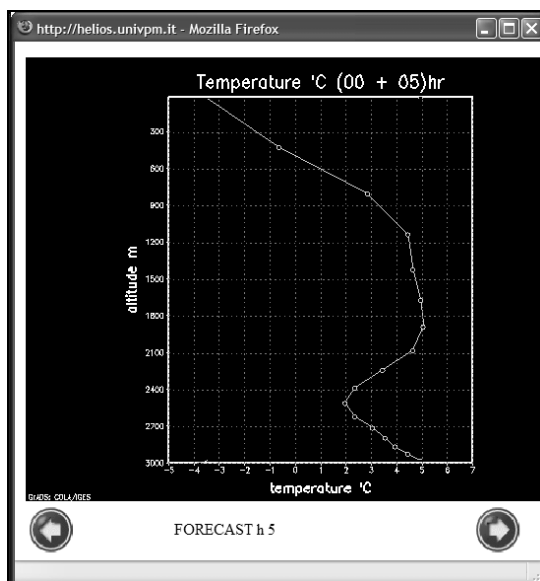

Figure 4:   Profile representation with fixed latitude and altitude.

3.3   System architecture

Beyond the simple using of the developed interface there is a simple but effective software architecture based on HTML, PHP, and JAVA-script languages.

As soon as a set of information have been input and a submit action is performed, the related information are passed by the HTML client to the PHP server, which provides a new proper configuration file and runs the proper

application. This operation, for a simple visualization of meteorological parameters, is performed twice.

The main page consists of two frames structured by HTML code. Each frame calls a PHP code that dynamically generates the frame content. More precisely, the code in first form loads the second PHP script containing the display parameters to be input. Here, a special point is the need of *code parameterization* i.e. the capability of passing variable contents through the code. A sample excerpt follows

```
echo"<center>
<a href='$parameter?date=$date&map=$map&simulation=$ simulation'
target='main'>SUBMIT</a></center>";
```

The simulation dates reported in the drop down box are collected by the PHP script searching through the target directory the available analysis-files. The script also reads the simulation start time and stop time. The name of selected simulation is written in the proper variable.

The available parameters can be selected among all the possible choices. The selection follows no special criterion but our common use or interest. The selected parameter is associated to a specific parameter-code that is written in the proper variable.

When the *submit* button is clicked all these information (along with a *flag* for ASCII output or GRIB output) are introduced in the post–processor configuration file.

Another parameterization step is performed after the second form has been filled; it is followed by a proper variable-writing analogue to the previous one. All the parameters and the variables are then passed to the GRADS script. A general statement is reported below.

```
$command="gradsc -lbc \"run batchscript.gs $date $simulation $map
$levelmin $levelmax $latmin $latmax $lonmin $lonmax $resolution
$timeinit $timefin\"";

passthru($command, $return);

echo $return;
```

The *passthru()* PHP function allows the execution of a shell command at the operating system prompt level and to catch the execution status code. This operation is performed in a "transparent" way for the browser.

The GRADS script receives arguments in the following way

```
function batchscript0 (args)
…
date=subwrd(args,1)
simulation=subwrd(args,2)
map=subwrd(args,3)
…
```

Finally, the Java Scripts have been embedded in the PHP code of the main form by calling a pop-up window that is able to visualize the generated images

(see Figure 2). A special feature is the easy animation of simulations as image sequence cycling.
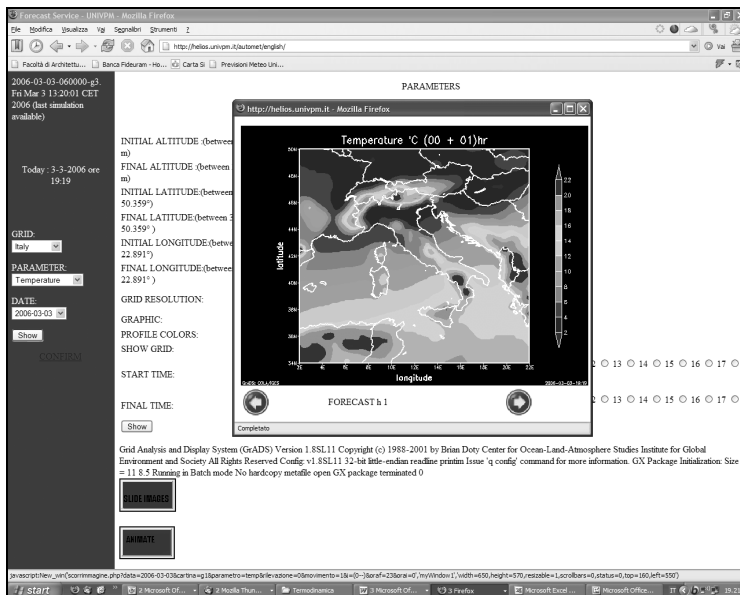

Figure 5:   Colored areas representation for temperature in the main grid.

3.4   Resources

3.4.1 Meteorological data

Meteorological Data are provided by the National Center for Environmental Prediction (NCEP). We download the AVN objective and the AVN reanalysis files produced every synoptic hour (up to 72 hour forecast each hour).

We adopted a *wget* utility (a freeware available on the web) that manages batch FTP processes rather than directly downloading the files. As an example, some lines of the related script are reported

```
wget --progress=dot:mega
ftp://anonymous:automet@ftpprd.ncep.noaa.gov/
   pub/data/nccf/com/avn/prod/avn.$YEAR$MONTH$DAY/
   $REANALYSIS
```

3.4.2 Data decompression and selection

We easily performed this task through the *fdgrib* utility (a Fortran procedure kindly distributed by ATMET). Once configured and compiled, this utility is able to "degrib" a dataset and to write it according to the RALPH2 format, the ASCII format required by RAMS [8]. By properly modifying the extraction routine, it is possible to restrict the output data (e.g. a 60x60 degree area centered on Southern Europe).

3.4.3 Scheduling

The scheduling has been committed to the efficient *crond* demon running under Linux (currently a RedHat distribution is installed). Every 12 hours, at 6am and 6pm, the scheduled scripts perform the sequence of operations required for a simulation, mainly data downloading, data degribbing, and model running. Along with these, other control operations are carried out such as the check of available disk quotas dedicated to simulations and eventually a disk clean up. Further checks are performed upon uncompleted or unsuccessful operations e.g. the unavailability of upper-air files at a certain hour is to be foreseen. Once the files are available, the system checks the possibility of performing a simulation without blocking the next one. Further controls are present for unexpectedly interrupted processes, which are reprised when possible. Finally, a specific event logging is performed and logs are dispatched to service administrators.

4   Discussion and conclusions

The on-demand post-processing service has after few months of tests given encouraging results. In fact, a brief survey over web forecast services outlined that they generally offer a restricted range of information and higher maintenance requirements. In AutoMet, whoever accesses the service has a fully operative viewer available, without any further knowledge on its functioning or any knowledge of programming languages.

The interface adopted is user friendly and gives the opportunity, even to non-meteorologists, to query and interpret graphics. This is the main feature required by the Province Authority both for internal use and for the dissemination of information. Users who may be interested in this information rendering approach are environmental policy manager, citizens, researchers in related fields, and Authorities for quick response to episodes. It is also interesting its implicit capability to be linked to a wider database.

The service benefits from all characteristics of web-based applications. It gives the opportunity of intrinsic multi-user connections each with his own dynamically-created pages.

There are several promising developments such as the doubling of meteorological data forecasting introducing the implementation of parallel MM5 simulations.

The second phase of the project will be implemented by introducing dispersion models whose results will be available on-demand as well. The first model to be implemented will be *HYPACT*, a hybrid Lagrangian/Gaussian model from ATMET, which has been designed to be directly hooked to the RAMS processor. The second model yet planned is the well known *CalPuff*.

Acknowledgement

References

[1]   G. Latini, G. Passerini, S. Tascini*, "Interfacing environmental models through World Wide Web".* VII International Conference on Development and Application of Computer Techniques to Environmental Studies, Bilbao, Spain, 28-30 June 2000.

[2]   T.N. Krishnamurti, L. Bounoua, 1996: *"Numerical Weather Prediction techniques (An introduction to)"* CRC Press, Inc.

[3]   Huag, F., 1995: *"Description of Selected Products from the National Oceanic and Atmospheric Administration (NOAA), National Meteorological Center (NMC) Gridded Data".*

[4]   Jensenius S., J.P. Dallavalle and S.A. Gilbert: *"The AVN based statistical guidance message".* Techniques Development Laboratory, National Weather Service

[5]   M.J. Bell, R.L. Walko, C.J. Tremback: *"REVU: RAMS/HYPACT Evaluation and Visualization Utilities, version 2.3.1 user's guide".* ASTER/MRC Division, Fort Collins, CO, 80522.

[6]   Clifford H. Dey, 1998: *"Office Note 388, GRIB (edition 1), The WMO format for the storage of weather product information and the exchange of weather product message in gridded binary form as used by NCEP central operations".* NCEP/NOAA/NWS.

[7]   Robert L. Walko, Craig J. Tremback, 2000: *"RAMS: The Regional Atmospheric Modeling System version 4.3/4.4 Model input namelist parameters".* ASTER/MRC Division, Fort Collins, CO, 80522.

[8]   AA.VV. 2000: "RALPH datasets format, version 2- available at http://www.atmet.com/html/docs/data/ralph2.pdf

[9]   http://www.atmet.com - *official RAMS site*

[10]   http://grads.iges.org – *Official* GrADS *site*

[11]   http://crati.it – *Italian research center site. There is a section dedicated to RAMS and MM5 meteorological forecast on Calabria region*

[12]   http://www.ba.infn.it/~regano/nowcasting/prototipo4_php - *An Italian now-casting experimental center – Puglia Region*

[13]   http://www.ecwmf.int - *European Centre for Medium Range Forecast. Official site*