

Problem Solving Environment for Flood Forecasting¹

Ladislav HLUCHY, Ondrej HABALA, Branislav SIMO,

Jan ASTALOS, Viet D. TRAN, Miroslav DOBRUCKY

Institute of Informatics, Slovak Academy of Sciences
Dubravská cesta 9, 842 37 Bratislava, Slovakia

ABSTRACT

Flood forecasting using numerical weather models is a computationally-intensive and complex task requiring cooperation of many experts in different areas. To enable this cooperation in a comfortable way, a part of the CrossGrid project is aimed towards developing a flood-forecasting, Grid-based environment. This Problem Solving Environment (PSE) consists mainly of a set of coupled simulation models, a storage system for configuration files, model codes and datasets, a WWW portal with collaboration tools and a powerful Computational Grid, provided by the partners in the

CrossGrid project.

This paper describes prototype of this PSE, targeting mainly the coupled simulation models (simulation cascade), the portal and the Virtual Organization supposed to use the system.

Keywords: Problem Solving Environment, Flood Forecasting, Grid Computing, Collaboration, Portal.

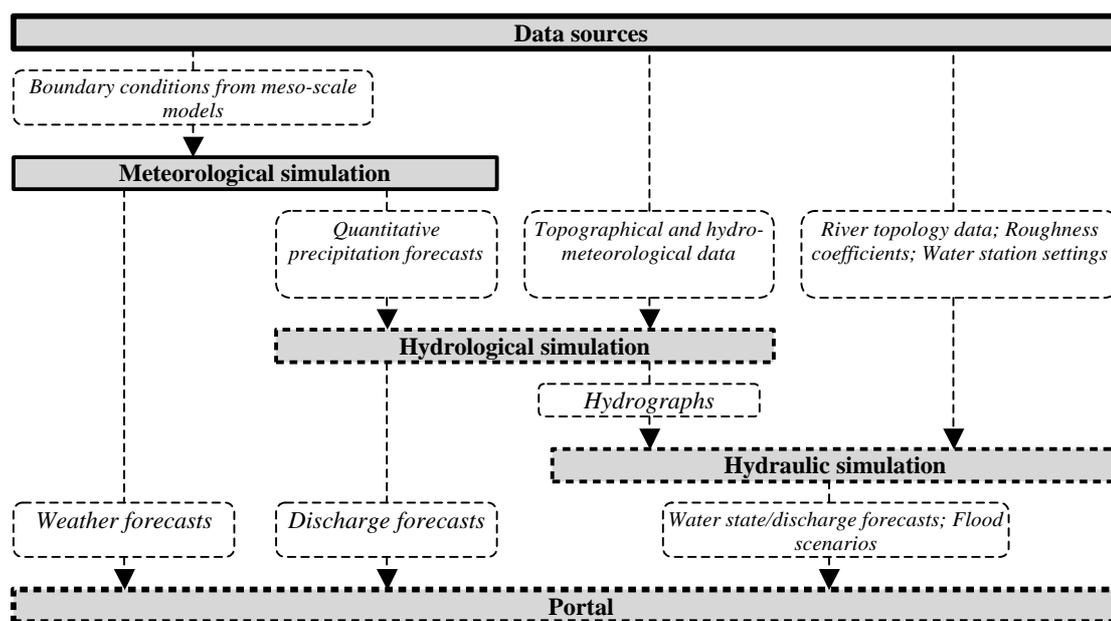


Fig. 1 Cascaded simulation scheme

¹ This work is supported by EU 5FP CROSSGRID IST-2001-32243 RTD project and the Slovak Scientific Grant Agency within Research Project No. 2/7186/20

1. INTRODUCTION

In this paper we present a prototype of a collaborative problem solving environment [1] intended to support a virtual organization for flood forecasting. Over the past few years, floods have caused severe damages throughout the world. Most of the continents were heavily threatened. Therefore, modeling and simulation of flood forecasting in order to predict and to make necessary prevention is very important.

The system - whose prototype is described here - employs the Grid technology [2] to seamlessly interconnect the experts, data and computing resources needed for quick and correct flood management decisions. Main component of the system is a cascade of simulation models used to predict weather, hydrological river status and hydraulic events in flooded areas. The prototype also includes some basic communication tools, enabling its users to cooperate. In recent years, since the 1991 workshop on Problem Solving Environments [PSE], held in Washington, a lot of research in the area of PSEs has been done. PSEs are no more used exclusively to solve partial differential equations. They are used as expert systems for a wide scale of science applications, from environmental sciences [3] through biology (even DNA mapping [4]) to military problems [5]. Of course, the main stream is oriented towards mathematical tools, computer-aided design and high-energy physics simulations. Current trends present to us a new way of cooperative problem solving - Collaborative PSE, using the Internet technology to connect geographically separated researchers. These CPSEs employ virtual forums, message boards, chat rooms, audio- and videoconferencing tools to better (and in a more comfortable way) connect their users, thus enabling them to solve complex problems more effectively.

Our PSE, as described here, is used to support a virtual organization of experts, developers and hydrometeorological information users from various organizations connected to create, distribute and evaluate the information needed to prevent floods. Because of the complexity and computational demands of flood forecasting, the Grid [2] environment is used to provide the necessary resources for computation as well as for data storage.

2. SIMULATION CASCADE

The PSE is based on a cascade of three types of simulations: meteorological, hydrological and hydraulic.

Its structure, basic data flows and types of the data are shown in Fig. 1.

Meteorological modeling

Forecasting of flood events requires quantitative precipitation forecasts as well as forecasting of temperature (to determine snow accumulation/melting). The simulation of atmospheric processes with the objective to predict future developments is the objective of Numerical Weather Prediction (NWP). The meteorological models are generally computationally intensive and they are usually running on the supercomputer class systems. The output of mesoscale models is used by meteorologists as a boundary condition for regional models. This nesting requires efficient management and transfers of large (tens to hundreds of megabytes) datasets. The prediction of flash floods will require employment of high resolution storm-scale models.

In the first prototype our system will use the ALADIN/SLOVAKIA [16] model currently operated by SHMI. ALADIN is a LAM (Limited Area Model) developed jointly by Meteo France and cooperating countries. It can be viewed as an extension to ARPEGE, a global NWP model operated in Meteo France. Main purpose is to provide more detailed short range forecasts inside limited domain of interest. Currently the model ALADIN is operated in 13 Euro-Mediterranean countries.

The ALADIN model consists of more than 1 million lines of (mainly) Fortran 90 code. Parallel version uses MPI communication library and was successfully ported to our target platform. First tests showed that the model code has high demands on communication network and thus is not very suitable for distributed running in the Grid. The main benefit of using Grid environment could be in performing of parameter studies.

Hydrological modelling

Several hydrological simulation models will be used, depending on conditions and needs which model will be applied for which situation and territory; they can be also used in combination. The following models we plan to integrate in our GCE:

- **HEC-1** model - is **single-event-based**, designed to simulate the surface runoff response of a river

basin to precipitation by representing the basin as an interconnected system of hydrologic and hydraulic components [14]. The result of the modeling process is the computation of streamflow hydrographs at desired locations in the river basin.

- **HSPF** - Program for **continuous** simulation of watershed hydrology and water quality [13]. The HSPF model uses information such as the time history of rainfall, temperature and solar radiation; land surface characteristics such as land use patterns; and land management practices to simulate the processes that occur in a watershed. The result of this simulation is a time history of the quantity and quality of runoff from an urban or agricultural watershed.
- **ERM** (Empirical - Regressive Model) – for **continuous** simulation; it is based on the principle of autoregressive model. The simplest regressive model is relational with the forecasted discharge being function of previous discharge and previous rainfalls. The function parameters are determined by regressive analysis of time series of discharges and rainfalls in selected time interval. In case of snow also temperature time series are taken into account and snow stage is computed.

We started to use the HSPF model and other models will be added later. These models need not be parallelized as they are not computationally intensive. However they are data-intensive, and a numerous parameters need to be specified by model calibration via parametric run in a Grid environment. Model tuning needs more effort to obtain land characteristic data and developers can profit from utilization of our first GCE version running in our Grid testbed.

Measured historical data are collected in tables (plain text, Excel table) and transformed into WDM file [13] using converter scripts and USGS tools (annie, iowdm, hspf) [15]. Also HSPF itself can manage most of the input data and transform it to/from WDM file.

New forecasted data (precipitation and temperature for 2 days) can be read by HSPF, inserted to WDM file and immediately used for simulation. When measured data are ready in following days, HSPF can overwrite the forecasted ones. HSPF run is steered by UCI file in which starting/ending data and other parameters (tuning parameters) can be modified via portal. Output data are stored back into the WDM file and are also written into small output file that is passed to the next step in the simulation cascade.

Hydraulic modeling

One of the hydraulic models in the PSE is FESWMS (Finite Element Surface-Water Modeling System) Flo2DH [11] which is a 2D hydrodynamic, depth averaged, free surface, finite element model. Flo2DH computes water surface elevations and flow velocities for both super- and sub-critical flow at nodal points in a finite element mesh representing a body of water (such as a river, harbor, or estuary). Effects of bed friction, wind, turbulence, and the Earth's rotation can be taken into account. In addition, dynamic flow conditions caused by inflow hydrographs, tidal cycles, and storm surges can be accurately modeled. Since Flo2DH was initially developed to analyze water flow at highway crossings, it can model flow through bridges, culverts, gated openings, and drop inlet spillways, as well as over dams, weirs, and highway embankments. Flow through bridges and culverts, and over highway embankments can be modeled as either 1D or 2D flow.

Simulation of floods is very computation-expensive. Several days of CPU-time may be needed to simulate

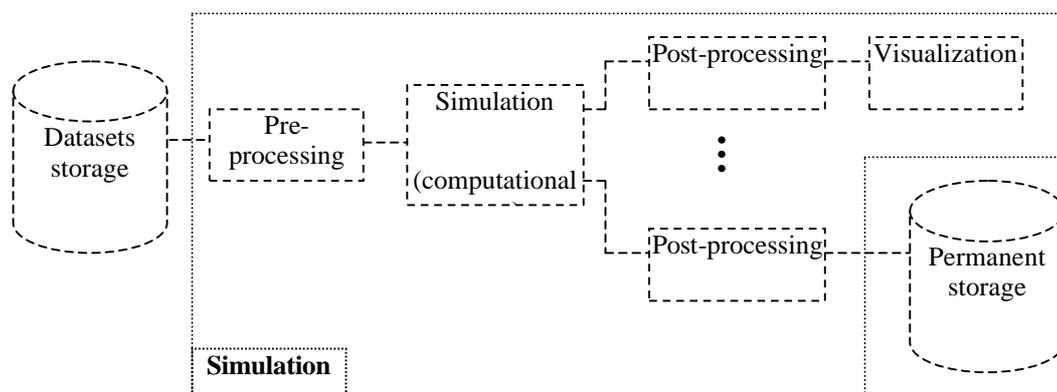


Fig. 2: Internal structure of a simulation block

large areas. For critical situations, e.g. when a coming flood is simulated in order to predict which areas will be threatened, and to make necessary prevention, long computation times are unacceptable. Therefore, FESWMS Flo2DH was parallelized in order to achieve better performance. In the parallel version, iterative linear solvers based on Krylov subspace methods [12] are used instead of the direct solver. These iterative solvers consist of matrix and vector operations only; thus, they offer large potential parallelism. In comparison with the direct solvers, iterative solvers are often faster and require less additional memory. The only drawback of iterative solvers is that they do not guarantee convergence to a solution. Therefore, preconditioning algorithms [13] are used to improve the convergence of iterative algorithms.

Cascaded Simulating

Each of the three simulations depicted in Fig. 1 can be considered as a block that can be further decomposed. Each simulation block consists of several sub-blocks as shown in Fig. 2.

Each simulation takes the input data set from the storage, preprocesses the data and starts the computational part. After its finish the post-processing of the output dataset is performed.

Pre- and post-processing filter blocks (dashed boxes in the picture) extract the data of interest from the dataset and format it to a structure suitable for the following processing or storage. The filter blocks may or may not appear in actual simulation block depending on specific data structure requirements of individual components. Output filtering before storing the resultant datasets may be important in case of large datasets (like meteorological

output) where only a part of it is interesting.

3. VIRTUAL ORGANIZATION FOR FLOOD FORECASTING

The scheme of the virtual organization as proposed in [8] was a general one. To actually create a real product, we have had to specify real users, data providers, storage providers, and cycle providers. In Fig. 3 we can see actual components of the prototype (grey color).

- Data providers – currently the only data provider is the Slovak Hydrometeorological Institute (SHMI). It is responsible for supplying initial boundary conditions for the meteorological simulation (these come from an external source) and for radar measurements.
- Cycle providers – the system will be tested in the CrossGrid international testbed environment which will consist of 11 institutions distributed across 8 European countries. The testbed will provide applications with distributed computing and storage resources connected by high-performance network (GEANT). One of the main objectives of the CrossGrid testbed is to ensure interoperability with other grid testbeds.
- Storage providers – the storage space for computation output is provided by the Institute of Informatics of the Slovak Academy of Sciences (II SAS). Temporary storage space needed to locally store input and immediate output data for running simulations will be provided by testbed contributors, whose

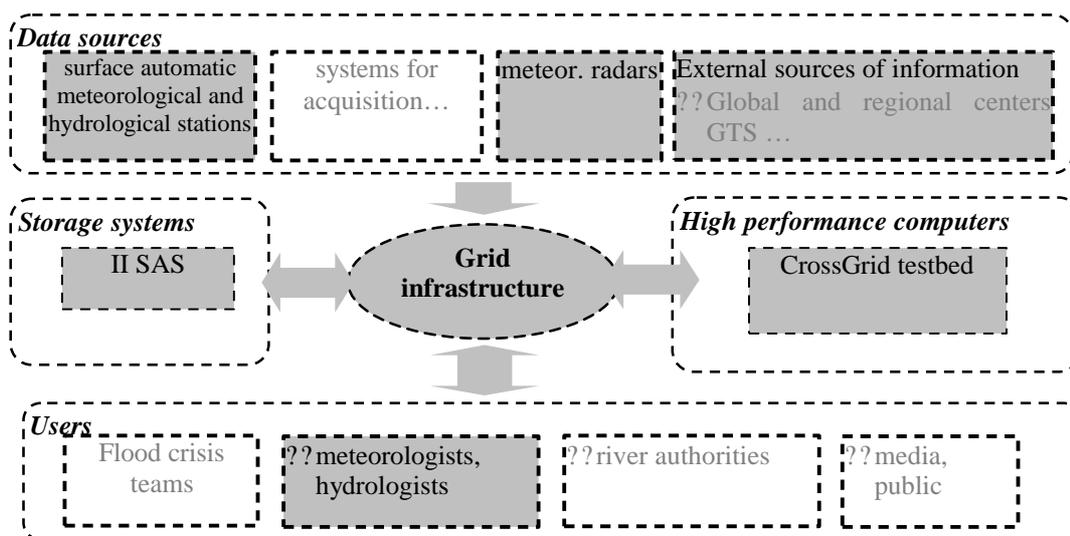


Fig. 3 Virtual Organization for Flood Forecasting

processing facilities will be used.

- Users – there are several groups of users in the VO – meteorological and hydrological experts, developers and end users.
- Experts – the system will be used solely by meteorological and hydrological experts from the SHMI and by trained staff from II SAS. These experts will provide configuration parameters for the simulations and they will decide which simulations need to be executed.
- Developers – all development is made by II SAS. They are preparing the portal, storage facilities and scripts for the simulations. They will also modify, test and deploy simulation model codes.
- End users – will be, once again, experts from SHMI, who will evaluate outputs of the system. In this stage no other users will be involved.

4. PORTAL

Collaboration tools

The prototype application will contain only basic communication services. We will employ a discussion forum, where users can post messages. These messages will be public to all other users of the PSE and they will be able to reply to them, thus creating a discussion.

The first version of the portal will be based on the NPACI GridPort Toolkit [10], a set of Perl scripts running on top of the Globus toolkit. GridPort supports basic services such as login and logout to/from the grid,

job submission, manipulation of data files, and monitoring of resources.

Visualization of results will be carried out by either simulation tool itself (as is the case for meteorological model ALADIN) or external tools such as OpenDX, gnuplot etc.

Simulation preparation, run and visualization

The user can choose, via portal, a model, input data and simulation parameters to be used. The portal then checks whether it is possible to use the given model and input data together, pre-processes input files, possibly generating definitions files for models, and finally generates scripts for job submission.

Using these files and scripts the simulation will be submitted to the grid via standard grid submission tool.

After the simulation has finished, the user can specify parameters of visualization tool(s) (sizes of images, locations, zoom levels, time length, etc.) and relevant pictures are presented through the portal.

Dataset naming and meta informations

Datasets are registered in a data catalogue, which stores information about their physical location, time and date the prediction is made for, time and date of creation, simulation model used, type of dataset (if not fully determined by simulation model), ...

5. STORAGE FACILITY

The storage space for simulation outputs and direct measurements used by the application is provided by II

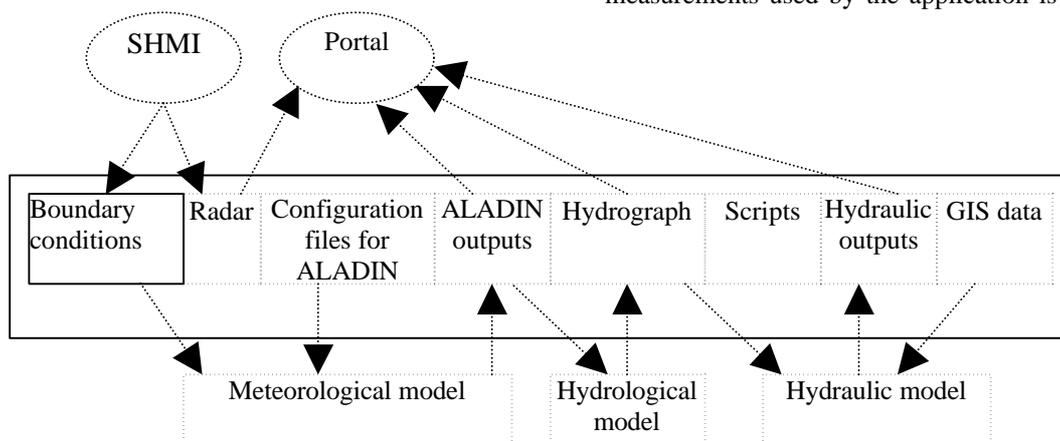


Fig. 4 Architecture and usage of the storage system

SAS. Hourly outputs of meteorological simulation, hydrographs provided by the hydrological part of the cascade and selected hydraulic outputs will be stored. The storage will also hold configuration files for the simulations and some other resources, needed to operate the application.

In Fig. 4, from the entities involved in the simulating process, only those essential for the storage facility design and operation are shown. As we can see in the Fig. 4, initial input data (ALADIN boundary conditions) will be supplied by SHMI. The same organization will also provide postprocessed radar pictures with rainfall density and other meteorological data measured in ground stations. The boundary conditions together with configuration files will be loaded into the ALADIN model code and a weather simulation (ALADIN outputs in Fig. 4) will be computed. This weather simulation will be used to produce a hydrograph with a selected hydrological model. The produced hydrograph may be used to create a hydraulic simulation of water flow. Outputs of the three simulation stages, together with radar pictures, will be displayed on the portal.

The stored files are accessible through standard Grid tools used in the CrossGrid testbed [6]. We will also have to establish a common description scheme for these files and to store the metadata in a Grid-aware database system. We are going to use the Datagrid Spitfire [9] engine, together with MySQL RDBMS [7].

The metadata structure (to be developed) will have to include detailed information about origin of the file, time of its creation, the person who actually created it, etc. In case the file is the output of a simulation, the metadata will also contain names of the input files, model executable and configuration files.

6. CONCLUSION

In this paper we have presented a prototype of a PSE intended to support a team of hydrometeorological experts in flood forecasting. This PSE employs a set of coupled simulation models, a web-based interface, a data storage system and Computational Grid as its native environment. In the future we plan to extend the PSE with additional simulation model codes and with an enhanced interface; the data management will be also deployed, using appropriate Grid-aware tools.

7. REFERENCES

[1] Gallopoulos, S., Houstis, E., Rice, J.: Computer as Thinker/Doer: Problem-Solving

Environments for Computational Science. **IEEE Computational Science and Engineering Magazine**, 1994, Vol. 2, pp. 11-23.

[2] Foster, I., Kesselman, C.: **The Grid: Blueprint for a New Computing Infrastructure**. Morgan Kaufmann Publishers, Inc., 1999.

[3] Fine, S. S., Ambrosiano, J.: The Environmental Decision Support System: Overview and air quality application. **American Meteorological Society: Symposium on Environmental Applications**, pp. 152-157. 1996, Atlanta, USA.

[4] Ziébelin, D., Parmentier, T.: Distributed Problem Solving Environment Dedicated to DNA Sequence Annotation. **Knowledge Acquisition, Modelling and Management**, pp. 243-258, 1999.

[5] Fox, G., Furmanski, W.: Problem Solving Environments from Simulation, Medicine and Defense using the Web. CRPS Annual Meeting, May 1996. <http://www.npac.sur.edu/users/gcf/crpcsemay96/fullhtml.html>

[6] Marco, R.: Detailed Planning for Testbed Setup. The CrossGrid Project, 2002. <http://grid.ifca.unican.es/crossgrid/wp4/deliverables/CG-4-D4.1-001-PLAN.pdf>

[7] Widenius, M., Axmark, D.: **MySQL Reference Manual**. O'Reilly & Associates, Inc., 2002.

[8] Hluchy, L., Astalos, J., Habala, O., Tran, V. D., Simo, B.: Concept of a Problem Solving Environment for Flood Forecasting. Recent Advances in Parallel Virtual Machine and Message Passing Interface. **LNCS 2474**, Springer, 2002.

[9] Bell, W. H., Diana Bosio, D., Hoschek, W., Kunszt, P., McCance1, G., Silander, M.: Project Spitfire - Towards Grid Web Service Databases. Global Grid Forum 5, Edinburgh, 2002.

[10] NPACI GridPort Toolkit, <https://gridport.npaci.edu>

[11] FESWMS - Finite Element Surface Water Modeling System. <http://www.bossintl.com/html/feswms.html>

[12] Y. Saad, H. Vorst: Iterative Solution of Linear Systems in the 20-th Century. **J. Comp. Appl. Math**, Vol. 123, pp. 1-33, 2000.

[13] K. Ajmani, W.F. Ng, M.S. Liou: Preconditioned conjugate gradient methods for the Navier-Stokes equations. **J. Computational Physics**, Vol. 110, pp. 68-81, 1994.

[14] HEC-1 <http://www.bossintl.com/html/hec-1.html>

[15] Hydrological Simulation Program-Fortran <http://water.usgs.gov/software/hspf.html>

[16] The International ALADIN Project. <http://www.cnrm.meteo.fr/aladin/>