Article

Grid computing technology for hydrological applications

LECCA, G., *et al*.

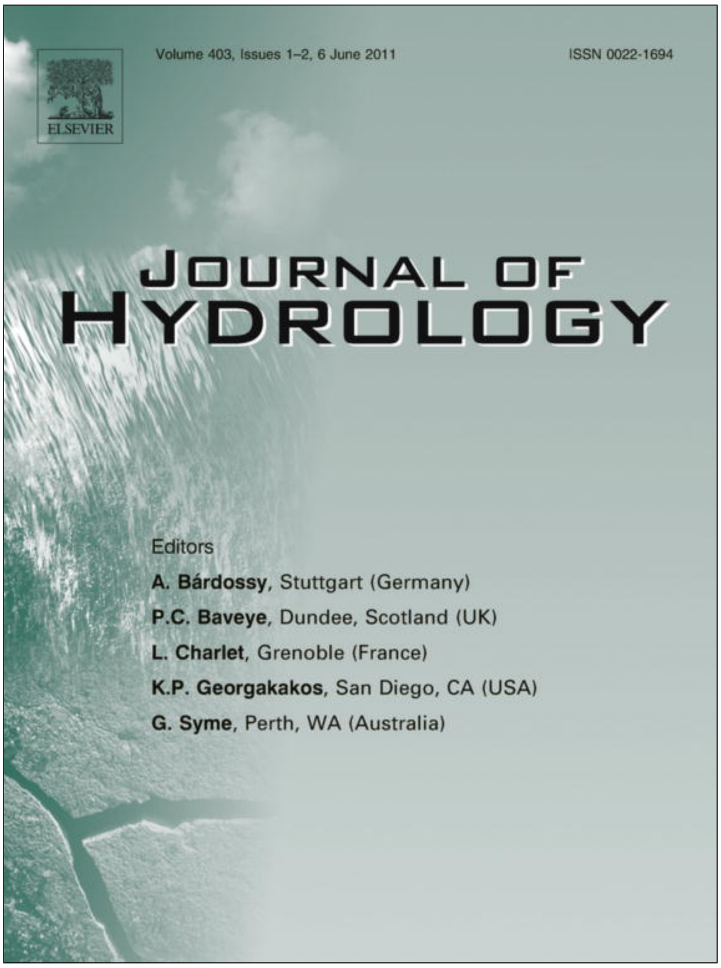
Abstract

Advances in e-Infrastructure promise to revolutionize sensing systems and the way in which data are col- lected and assimilated, and complex water systems are simulated and visualized. According to the EU Infrastructure 2010 work-programme, data and compute infrastructures and their underlying technolo- gies, either oriented to tackle scientific challenges or complex problem solving in engineering, are expected to converge together into the so-called knowledge infrastructures, leading to a more effective research, education and innovation in the next decade and beyond. Grid technology is recognized as a fundamental component of e-Infrastructures. Nevertheless, this emerging paradigm highlights several topics, including data management, algorithm optimization, security, performance (speed, throughput, bandwidth, etc.), and scientific cooperation and collaboration issues that require further examination to fully exploit it and to better inform future research policies. The paper illustrates the results of six dif- ferent surface and subsurface hydrology applications that have been deployed on the Grid. All the appli- [...]

Reference

LECCA, G., *et al*. Grid computing technology for hydrological applications. *Journal of Hydrology*, 2011, vol. 403, no. 1-2, p. 186-199

DOI : 10.1016/j.jhydrol.2011.04.003


Volume 403, Issues 1–2, 6 June 2011
ISSN 0022-1694
ELSEVIER
JOURNAL OF HYDROLOGY
Editors
A. Bárdossy, Stuttgart (Germany)
P.C. Baveye, Dundee, Scotland (UK)
L. Charlet, Grenoble (France)
K.P. Georgakakos, San Diego, CA (USA)
G. Syme, Perth, WA (Australia)

Journal of Hydrology 403 (2011) 186–199

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Review Paper

Grid computing technology for hydrological applications

G. Lecca [a,*], M. Petitdidier [b], L. Hluchy [c], M. Ivanovic [d], N. Kussul [e], N. Ray [f,g], V. Thieron [h]

[a] Center for Advanced Studies, Research and Development in Sardinia (CRS4), Edificio 1, Parco Tecnologico "Polaris", 09010 Pula (Cagliari), Italy
[b] Université Versailles St-Quentin, CNRS/INSU, LATMOS-IPSL, 10-12 Avenue de l'Europe, 78140 Velizy, France
[c] Institute of Informatics of the Slovak Akademia of Science (IISAS), Bratislava, Slovakia
[d] Faculty of Science of the University of Kragujevac, Radoja Domanovica 12, 34000 Kragujevac, Serbia
[e] Space Research Institute of the National Academy of Sciences of Ukraine and National Space Agency of Ukraine (NASU-NSAU), Glushkov Prospekt 40, Build. 4/1, Kyiv 03680, Ukraine
[f] EnviroSPACE Laboratory, Climate Change and Climate Impacts, Institute for Environmental Sciences, University of Geneva, Battelle - Building D, 7 route de Drize, CH-1227 Carouge, Switzerland
[g] Global Resource Information Database – Europe, Division of Early Warning and Assessment, United Nations Environment Programme, International Environment House, 11 Chemin des Anémones, CH-1219 Châtelaine, Switzerland
[h] Ecole des Mines d'Alès (LGEI) – 6, Avenue de Clavières, 30319 Alès Cedex, France

ARTICLE INFO

Article history:
Received 17 February 2010
Received in revised form 5 February 2011
Accepted 3 April 2011
Available online 17 April 2011
This manuscript was handled by P. Baveye, Editor-in-Chief

Keywords:
Computational hydrology
Environmental data
Grid computing
e-Infrastructure
Water resources management

SUMMARY

Advances in e-Infrastructure promise to revolutionize sensing systems and the way in which data are collected and assimilated, and complex water systems are simulated and visualized. According to the EU Infrastructure 2010 work-programme, data and compute infrastructures and their underlying technologies, either oriented to tackle scientific challenges or complex problem solving in engineering, are expected to converge together into the so-called *knowledge infrastructures*, leading to a more effective research, education and innovation in the next decade and beyond. Grid technology is recognized as a fundamental component of e-Infrastructures. Nevertheless, this emerging paradigm highlights several topics, including data management, algorithm optimization, security, performance (speed, throughput, bandwidth, etc.), and scientific cooperation and collaboration issues that require further examination to fully exploit it and to better inform future research policies. The paper illustrates the results of six different surface and subsurface hydrology applications that have been deployed on the Grid. All the applications aim to answer to strong requirements from the Civil Society at large, relatively to natural and anthropogenic risks. Grid technology has been successfully tested to improve flood prediction, groundwater resources management and Black Sea hydrological survey, by providing large computing resources. It is also shown that Grid technology facilitates e-cooperation among partners by means of services for authentication and authorization, seamless access to distributed data sources, data protection and access right, and standardization.

Contents

* Corresponding author. Tel.: +39 070 9250 272; fax: +39 070 9250 216.
E-mail addresses: giuditta.lecca@crs4.it (G. Lecca), monique.petitdidier@latmos.ipsl.fr (M. Petitdidier), hluchy.ui@savba.sk (L. Hluchy), mivanovic@kg.ac.rs (M. Ivanovic), inform@ikd.kiev.ua (N. Kussul), nicolas.ray@unige.ch (N. Ray), vincent.thieron@mines-ales.fr (V. Thieron).

0022-1694/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhydrol.2011.04.003

1. Introduction

Water is critical to well-being of humans, but its environment is changing at unprecedented rates. Growing populations, increasing urbanization, and climate change are shifting the balance between water supply and demand, and impacting the quality of water resources. Appropriated characterization, prediction and response to threats to water resources is limited by our understanding of the dynamics and spatial variability of environmental processes; and how perturbations alter these processes over time scales of decades or longer, and spatial scales that range from local to global.

Surface water and groundwater interaction, field scale behavior of spatially variable processes, multi-scale processes for river flood prediction, interplay between geological, hydrological and biochemical processes, optimal solutions to management of highly complex systems and socio-economic implications of human interventions – just to cite few examples – are challenging research objectives governed by complex mathematical models based on highly-demanding simulation and data analysis tools.

Hydrology science is largely based on the systematic integration of data originating from several disciplines: weather and ocean observations, soils, geomorphology and geology, social and demographic datasets, etc. In addition, water data are often stored in widely dispersed databases, managed by regional and national-government agencies in charge of water protection and management. The data, originating from a large variety of ground-based as well as satellite sensors, are sometimes neither quality assured, analyzed, or interpreted adequately, nor available in a sufficiently timely manner to inform management actions.

The amount of data to exploit and the modeling requirements need more and more computing resources (in terms of CPUs and storage). But despite their rapid growth over the last two decades, a gap still exists and limits the ability of researchers, professionals and decision makers to identify, assemble and analyze all relevant data for a region or problem of interest (Cossu et al., 2010).

In many Earth Science (ES) fields, integrated networks of observing systems and facilities to store, retrieve and analyze the resulting data, are today foreseen as the key building block to provide the foundation on which the understanding of environmental processes can go forward and critical decisions to be made. In hydrology, initiatives are taken in Europe such as the Distributed Research Infrastructure for Hydro-Meteorology Study[1] (DRIHMS) and in the USA such as the WATERS network.[2] Many ES data centers have developed service tools based on Web services for basic research activities like searching, discovering, browsing and downloading of datasets. However applications need a direct access not only to various heterogeneous and geographically-distributed data sets, but also to the required computing resources.

Grid technology as defined by Foster and Kesselman (1999) and Foster (2002) meets these requirements as a distributed resources system allowing (1) the integration and coordination of resources and users living within different control domains and (2) the scaling up of computing power and storage capacity in a way that is impossible for a single institution to do. Grid could be also considered as an e-collaboration platform allowing scientific cooperation within a virtual team, i.e. sharing information, knowledge, applications, data and resources via (dedicated) Web gateways (Bourras et al., 2009).

This paper aims at describing, from the application perspective, various aspects of this emerging paradigm in science and engineering through the description of six different Grid-enabled hydrological applications, ported independently on the Grid.

The body of the paper is organized into four sections. After this introduction, Section 2 addresses the description of the Grid concept along with other interconnected technological frameworks, relevant to Earth and Environmental Sciences, such as the spatial data infrastructures (SDIs) and the Web developments that are at the very core of the research e-Infrastructures. Section 3 illustrates the Grid-enabled hydrological applications related to flood prediction and monitoring, groundwater resources management and comprehensive hydrological and environmental survey of the catchment area of the Black Sea. For each of them we illustrate the hydrological environment, difficulties and limitations of previous relevant studies, model mechanisms, value added solutions by means of data management and software implementation issues, and future improvements. Section 4 draws some conclusions from the lessons learned, anticipates future developments and highlights research policy issues.

2. What is the Grid?

In this section, an overview of the main Grid technology characteristics is presented without discussing detailed implementations and solutions. For such details the reader is referred to Foster and Kesselman (1999) and Wang et al. (2009). The term *Grid computing* originated in the 1990s as a vision for making computer power as easy to access as electric power (Foster and Kesselman, 1999). A Grid can be defined as a layer on top of network services that typically allows users single sign-on access to distributed collection of resources not centrally controlled. These resources can among others be in the form of computational capabilities, data repositories, software services and applications. According to the resources shared and domain involved the term *Grid* is used for a wide variety of different objectives. A commonly used way is to divide them into Computational or Data Grids according to their focus on provision of computing cycles or data solutions and, into Science Grids (e.g. HealthGrid[3]) or Business Grids according to the category of users and applications.

Basic concepts lying behind the word "Grid" are: medium-term and dynamic collaboration, user and provider communities and security. Associated to the Grid, there are always users belonging to different administrative organizations but to the same community or project, called *Virtual Organization* (VO), and resources from different organizations that constitute the Grid infrastructure. This user community wants to share geographically distributed resources in a secure way. Users as well as resources must be authenticated by a certification authority before acceptance in the VO for users or in the Grid infrastructure for resources. A valid certificate

1 http://www.drihms.org.
2 http://www.watersnet.org.
3 http://www.healthgrid.org/.

Table 1
Main characteristics of the six Grid-enabled hydrology applications.

Ref.	Domain	Model	Data	Parallel system	GIS/OGC	Web interface	Goal
3.1	Flood	Multi-model system	Local and Global (meteorological data)	Distributed Job (MPI)	–	Grid portal	Demonstration
3.2	Flood	–	Global (Satellite data)	Distributed Job (OpenMP)	GIS and OCG	Portal + Web-GIS	Real time
3.3	Flash flood	ALTHAIR	Local and Global (meteorological data)	Independent Jobs (scenario analysis)	OGC	Portal + Web-GIS	Real time
3.4	Subsurface hydrology	CODESA-3D	Local	Independent Jobs (Monte Carlo)	GIS	Portal + Web-GIS	Water Resources Management
3.5	Subsurface hydrology	Lizza-PAKP	Local	Distributed Job (MPI) + parametric analysis	–	–	Water Resources Management
3.6	Hydrology Catchment	SWAT	Local and Global	Independent Jobs (Monte Carlo)	GIS	Portal	Survey and Water Resources Management

authorize users to access the resources based on the policies of the VO. Security and confidentiality are of great importance. In hydrology, as well as in other ES fields, complex data policies govern the access to data. As an example, local or regional data concerning the water management may be very sensitive. Therefore, in their current form, most Grids use encryption and advanced authentication mechanisms, such as public key certificates and different user roles in the VO, to protect data confidentiality.

2.1. The Grid middleware

As a Grid often spans multiple administrative domains, hardware and software installations vary greatly. The Grid middleware is responsible for the operation of the Grid and allows a uniform access to the often highly heterogeneous resources. The Grid middleware usually hides this disparity from the user perspective, so the Grid is seen as a seamless information processing system. Until today, several, often architecturally different, Grid-middleware have been developed or are still being actively developed. Commonly used middleware in Europe are for example gLite,[4] UNiform Interface to COmputing REsources[5] (UNICORE), Globus Toolkit,[6] and Advanced Resource Connector[7] (ARC). The most spread all over the world is the Globus toolkit that has the largest range of higher level services and allows users to easily build their own services, in particular interfaced with Web services. It is used currently by thousands of sites in business and academia but mostly not inter-connected. gLite is the middleware of the largest Grid deployment today, Enabling Grids for E-Science[8] (EGEE), that is designed for the analysis of the petabytes (i.e. 10^{15} bytes) of data produced by the European Organization for Nuclear Research's (CERN) Large Hadron Collider (LHC) experiment in Geneva. However, access to EGEE is not restricted to high energy physics and is currently used by other scientific communities mainly in public research including bioinformatics, astronomy and earth sciences (Renard et al., 2009). In October 2009, EGEE was deployed at around 260 sites, providing around 150,000 CPUs, more than 20 petabytes of storage, and running up to 330,000 jobs/day. A new era has started in May 2010 with the European Grid Infrastructure[9] (EGI) to support the transition from a project-based system such as EGEE to a sustainable pan-European e-Infrastructure to enable access to computing resources for European researchers from all fields of science.

The interoperability of different Grid middleware and the definition of standard components are discussed within the Open Grid Forum[10] (OGF). The Global Grid Forum (GGF) published the Open Grid Services Architecture (OGSA) that represents an evolution towards a Grid system architecture based on Web services technology, assuring interoperability on heterogeneous systems.

2.2. Geospatial data components and Grid

In the last years many European and international projects and initiatives (e.g. INSPIRE,[11] GEOSS,[12] GMES[13]) aimed to define an architectural framework for the realization of the so-called *Spatial Data Infrastructure* (SDI). The Open Geospatial Consortium[14] (OGC) Inc.®, a non-profit, international, voluntary consensus standards organization, is leading the development of standards for geospatial and location based services. It has defined specifications for many different geospatial Web-based services: the Open Geospatial Web Services (OWS). As an example, a Web Processing Service (WPS) is a computer code to publish and perform geospatial processes (e.g. a simple geometric calculation or a complex simulation model) over the Web with a standardized and open interface.

Linking Grid computing to OGC Web services is well suited to accomplish high processing performance and storage capacity along with improved service availability and usability. In the hydrology domain, a specific Spatial Data Infrastructure built upon the EGEE platform has been designed and implemented for the flash flood application presented in Section 3.3 (Mazzetti et al., 2009). Sensor Web services have also been implemented on the Grid in the application of flood monitoring using satellite data, in situ sensors and simulations (Kussul et al., 2009), presented in Section 3.2. OGC components are also implemented with Globus (Lanig and Zipf, 2009).

3. Grid-enabled hydrological applications

The section describes six Grid-enabled applications addressing hydrological problems related to surface and subsurface waters. They are characterized by different computational and technological solutions, summarized in Table 1. Three applications concern

[4] http://glite.web.cern.ch/.
[5] http://www.unicore.eu/.
[6] http://www.globus.org/.
[7] http://www.nordugrid.org/arc/.
[8] http://www.eu-egee.org/.
[9] http://www.egi.eu/.

[10] http://www.ogf.org/.
[11] http://www.inspire-geoportal.eu – Infrastructure for Spatial Information in the European Community.
[12] http://www.earthobservations.org – The Global Earth Observation System of Systems.
[13] http://gmes.info – Global Monitoring for Environment and Security.
[14] http://www.opengeospatial.org/.

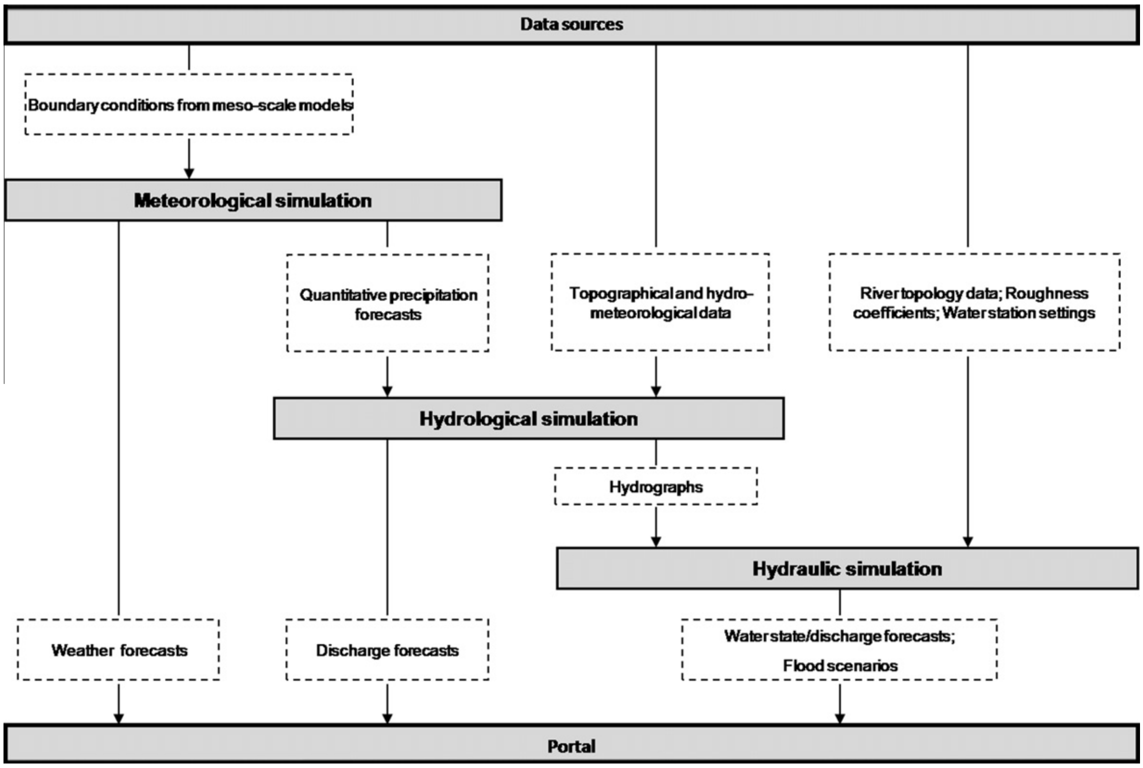


Fig. 1. Workflow of the cascade of simulations for flood forecasting.

river floods (3.1–3.3). The first multi-model multi-scale application (3.1) has started to be developed in the framework of the EU FP6 CROSSGRID[15] project (2002–2005) and demonstrates the capability to implement on the Grid complex workflows using a cascade of nested models operating at different spatial and temporal scales. The second application (3.2) is based on the information extracted from satellite data and is part of the United Nations Environment Programme (UNEP) on flood prediction. The third application (3.3) concerns a prototypal system for flash flood prediction developed in close cooperation with the forecasting service in France. Two applications are related to groundwater resources management (3.4 and 3.5). The first one (3.4) is a suite of modeling services and already deployed case-studies in the Mediterranean area, targeted for contamination problems including seawater intrusion in coastal aquifers and geochemical data analysis; the other (3.5) is a general model for groundwater flow. The last application (3.6) is devoted to a survey of the hydrology and biodiversity of the catchment area of the Black Sea.

3.1. Multi-model multi-scale flood workflow

Flood is a common problem in Slovakia due to the Danube River and its tributaries. Therefore, building a system for flood prediction, warning and prevention is topical. However, the problem is quite difficult and implies multi-disciplinary cooperation of many organizations, with a request for a huge amount of compute and storage resources. Flood prediction is a complex multi-model multi-scale application that is suitable to evaluate the capacity of Grid to implement a complex workflow. Flood forecasting requires quantitative precipitation forecasts based on meteorological simulations of different resolutions from meso- to storm-scale. From these quantitative precipitation forecasts, hydrological models are used to determine the discharge from the affected area. In turn, hydraulic models, based on this information, simulate flow through various river structures to predict the impact of the flood. Fig. 1 illustrates this cascade of models.

The meteorological models, data and expertise are provided by the Slovak Hydro-Meteorological Institute (SHMI). One of the meteorological models is ALADIN/SLOVAKIA model, which is currently operated by SHMI in daily operation. ALADIN[16] is a model with primeval equations, based on spectrum technique including hydrostatic, eulerian or semi lagranean options, digital filter initiating, optimum interpolation analysis and lately variational assimilation of 3D data. ALADIN is license-restricted so another mesoscale meteorological model, MM5,[17] is used. The PSU/NCAR mesoscale model (known as MM5) is a limited-area, nonhydrostatic, terrain-following sigma-coordinate model designed to simulate or predict mesoscale atmospheric circulation. Conversion of input data for MM5 from ALADIN/LACE is provided by SHMI. The WRF[18] (Weather Research and Forecasting) model, a next-generation mesoscale numerical weather prediction system designed to serve both operational forecasting and atmospheric research needs, is also under test.

Hydrological and hydraulic models, data and expertise are provided by the Water Research Institute (WRI) and the Slovak Water Enterprise, which manages all river authorities in Slovakia. The organizations have rich historical hydrological data from the network of water stations distributed over all important localities in Slovakia. However, the quality of terrain maps is very unequal. For some locations, high-resolution digital maps (1×1 m) have been obtained using LIDAR (Light Detection and Ranging) technology; for some other places, the maps even are completely missing.

[15] http://www.eu-crossgrid.org – Development of Grid Environment for Interactive Applications.

[16] http://www.cnrm.meteo.fr/aladin.
[17] http://www.mmm.ucar.edu/mm5/mm5-home.html.
[18] http://www.wrf-model.org/index.php.

Therefore, different hydrological models are applied for different river tributaries in Slovakia. When accurate data are available, physical models based on terrain maps like HEC-1[19] and HSPF[20] are applied. These models simulate the surface-runoff response of a river basin to precipitation by representing the basin as an interconnected system of hydrologic and hydraulic components. Models use continuous rainfall and other meteorologic records and the result of the modeling process is the computation of streamflow hydrographs and pollutographs at desired locations in the river basin. In the other cases, ERM (Empirical – Regressive Model) a purely empirical model based on historical hydrological data is used. Similarly, several hydraulic models to simulate surface-water flows are used, including FESWMS-2DH[21] and DaveF (Hluchy et al., 2002). Although both models have nearly identical input data (terrain maps, inflows, coverage) and produce similar output (water levels and velocities at flooded areas), FESWMS-2DH is based on the finite element method and is suitable for steady-state and semi-steady-state flow (plain flood), while DaveF is based on the finite volume method and is suitable for critical flow (flash flood). Both models are supported by a graphical user interface that can be used for pre- and post-processing.

The workflow (Habala et al., 2007) of the flood simulation is shown in Fig. 1. Each of the three simulation modules is considered as a composite block that can be further decomposed into several sub-blocks. Each simulation takes the input data set from the storage, preprocesses the data and starts the computational task. Then the post-processing of the output dataset is performed. Pre- and post-processing filter blocks (dashed boxes in the same figure) extract the data of interest from the dataset and format them into a structure suitable for the following processing or storage. The filter blocks may or may not appear in the actual simulation block depending on specific data structure requirements of individual components. Output filtering before storing the resulting datasets may be important in case of large datasets (like meteorological output) where only a subset is of interest. Simulation results are imported into GIS (Geographic Information System) software, such as ArcView[22] and GRASS,[23] to integrate them into maps for impact assessment.

Most of the models mentioned above, especially meteorological and hydraulic models, require a lot of computing resources. The use of Grid technology to provide enough computing power will be vital for an operational use of this workflow for flood prediction. Another point in favor of the use of Grid technologies is that it provides solutions for cooperation among the different organizations, e.g. authentication and authorization, distributed data sources, data protection and access right, and mainly the automatic execution of complex workflows with many components.

The workflow has been tested on a pilot site in Slovakia provided with topological LIDAR data. Before doing service composition, the meteorological data of the pilot area have to be extracted from global data and the hydrological and hydraulic models have to be calibrated for actual terrain conditions. The River Vah Authority was the customer of this computational service to simulate flood conditions for a small airport. Currently in Slovakia models for all the rivers are not calibrated, mainly due to missing topological data. Therefore, the workflow cannot be widely used in Slovakia but is planned to be used in other countries.

There are still some obstacles in running the application in fully operational mode on generic Grid infrastructures like EGEE. In operational mode, the application should be executed in real-time, assimilating data and boundary conditions from larger scale models and displaying the current state of the flood to forecasters. This implies the availability of the required computing resources when needed; since in EGEE the reservation of resources is not currently possible, there is no guarantee to have all the resources at the given time. However the Grid infrastructure is able to provide enough computation power and tools for cooperation and research investigation after the critical event.

3.2. Flood mapping from satellite Earth Observation data

Efficient monitoring and prediction of floods and risk management for large river is quasi-impossible without the use of Earth Observation (EO) data from space. As a matter of fact one of the most important problems associated with flood monitoring is the difficulty to determine the extent of the flood area as even a dense network of observations cannot provide such information. The flood extent information is used for damage assessment and risk management, and benefits to rescuers during flooding; it is also very important for calibration and validation of hydraulic models to reconstruct what happened during the flood and determine what caused the water to go where it did (Horritt, 2006).

The EO domain, in general, is characterized by large volumes of data that should be processed, catalogued, and archived (Fusco et al., 2007; Shelestov et al., 2006). The processing of satellite data can be viewed as a complex workflow that is composed of many tasks (Kussul et al., 2008). For flood mapping from satellite data, the workflow consists of the following stages: (1) data transfer upon user request, through the SAR (Synthetic-Aperture Radar) portal, from the ESA (European Space Agency) rolling archives to the local or Grid resources; local storage of the concerned SAR metadata in a catalog in order to enable efficient retrieval of SAR images; (2) geometric (ortho-rectification) and radiometric correction of the SAR image using digital elevation model (DEM); regions with shadows and layover are identified at this stage as well; (3) processing of SAR image using a neural network classifier. The neural network assigns to each pixel of the output image a binary value corresponding to one of two classes ("Water" and "No water") taking as input a moving window of image pixel intensities; (4) removal of those regions that were detected as shadow/layover at the stage No. 2, since the pixel values in these regions are not related to physical conditions of the environment; (5) geocoding, i.e. transformation to geographical projection using ground-control points (GCPs) incorporated in the SAR image; (6) visualization of the results using ESRI shape files, KML (Keyhole Markup Language) files, and OGC WMS. In order to provide a more comprehensive analysis of a flood situation, other data are used when available: optical satellite data (e.g. Landsat-5), land cover/land use maps (Corine Land Cover, ESA GLOBCOVER), and vector data (assets, cities/villages/settlements, population density/quantity, road networks). The produced flood maps are delivered to civil protection and emergency agencies to calibrate and validate hydrological models and to analyze different scenarios of disaster.

Dealing with EO data, we have also to consider the security issues regarding satellite data policy with the need of near-real-time processing in order to provide a fast response within international programs and initiatives, in particular the International Charter "Space and Major Disasters" and the International Federation of Red Cross. A considerable need therefore exists for an appropriate infrastructure that will enable the integrated and operational use of multi-source data for different application domains. From a technological point of view, Grids can provide solutions to the above-mentioned problems (Foster, 2002; Fusco et al., 2007; She-

[19] http://www.ground-water-models.com/products/wms_hec-1/wms_hec-1.html.
[20] http://www.epa.gov/ceampubl/swater/hspf.
[21] http://water.usgs.gov/software/FESWMS-2DH – Finite Element Surface-Water Modeling System.
[22] http://www.esri.com/software/arcview/index.html.
[23] http://grass.itc.it.

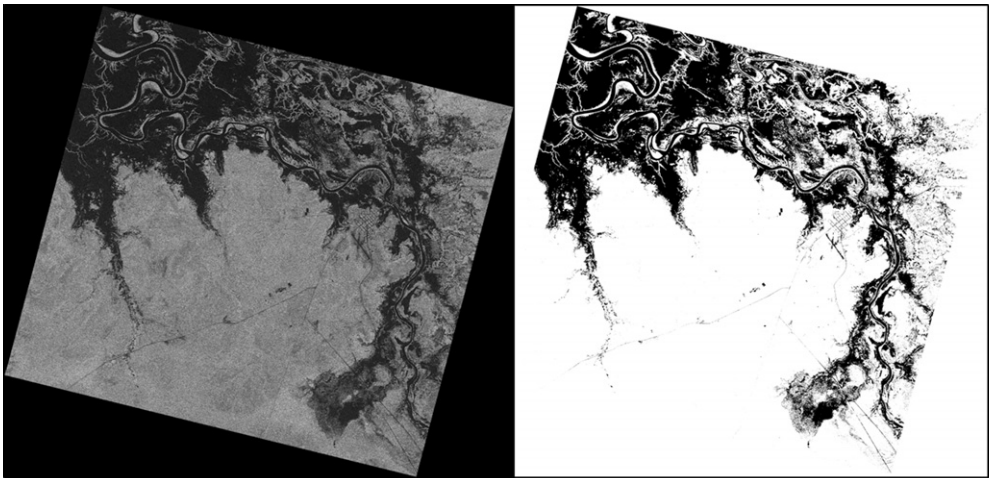


Fig. 2. SAR image acquired from RADARSAT-2 satellite (14.02.2009) during the flooding on the river Norman, Australia (left) and derived flood extent map © CSA 2009 (right).

lestov et al., 2006). In this case, a Grid environment can be considered not only for providing high-performance computations, but also to facilitate interactions between different actors by providing a standard infrastructure and a collaborative framework to share data, algorithms, storage resources, and processing capabilities (Fusco et al., 2007).

To benefit from Grid computing capabilities, a parallel version of the application of flood mapping from SAR imagery was developed on the Grid infrastructure. Parallelization of the image processing is performed by splitting the SAR image into uniform parts that are processed on different nodes using the OpenMP[24] Application Program Interface. The use of the Grid reduced the time required to process a single SAR image from more than 10 min on a single workstation to less than 1 min on the Grid infrastructure. Currently, the Grid infrastructure integrates resources of several geographically distributed organizations:

– Space Research Institute NASU-NSAU (Ukraine) with deployed computational and storage nodes based on Globus Toolkit 4, NorduGrid[25] and gLite 3 middleware, access to geospatial data and a Grid portal.
– Institute of Cybernetics of NASU (Ukraine) with deployed computational and storage nodes (SCIT-1/2/3 clusters) based on Globus Toolkit 4 middleware and access to computing resources (approximately 500 processors).
– The Center for Earth Observation and Digital Earth of Chinese Academy Science-CEODE-CAS (China) with deployed computational nodes based on gLite 3 middleware and access to geospatial data (approximately 16 processors).

Other Grid resources may be accessible if needed.

In all cases, the Grid Resource Allocation and Management (GRAM) service is used to execute jobs on the Grid resources (Feller et al., 2007). Access to the resources of the Grid environment is organized via a high-level Grid portal that has been deployed using GridSphere framework.[26] Through the portal, users can access the required satellite data and submit jobs to the computing resources of the Grid in order to process satellite imagery. The workflow of the data processing steps (transformation, calibration, orthorectification, classification, etc.) in the Grid is controlled by a Karajan engine.[27] The open-source OpenLayers framework[28] and UMN Mapserver v5[29] are used to visualize the results of data processing in the Grid environment. Having created WMS services for the EO derived products, we use them in the OpenLayers framework and in Google Earth by generating corresponding KML files.

The approach to determine flood extent from SAR imageries acquired by different satellite instruments (ERS-2/SAR, ENVISAT/ASAR, RADARSAT-1/2) has been applied to a number of case-study areas: Ukraine and Hungary (2001), China (2007), Mozambique (2008), Zambia (2009), Australia (2009, Fig. 2). Classification rates for independent testing data sets were 85.40%, 98.52%, 95.99% for ERS-2, ENVISAT and RADARSAT-1 satellite imagery, respectively. In case of emergency in any region, the UN-SPIDER[30] knowledge portal provides flood maps based on satellite data along with continuous monitoring of areas prone to seasonal floods. These areas include Ukraine (Carpathian region), Namibia (Caprivi region) and Australia (Norman River). The derived flood maps in conjunction with rainfall estimates, water level and water flow data can be used by statistical models to predict the flood extent.

In conclusion, the use of Grids allows us to reduce the overall computing time required for satellite image processing, to effectively manage complex workflows, and to make also possible the fast response within international programs and initiatives concerned with emergencies.

3.3. Flash flood and crisis management on a OGC platform

In South-East France, flash flood phenomena represent the most important natural threat for population and infrastructures. During the flash flood that occurred on September 2002 in the Gard region, 22 persons died and the economic damages were estimated at 1.2 billion euro (Delrieu et al., 2005; Sauvagnargues-Lesage and Ayral, 2007). After such dramatic floods, the proposed actions concerned the enhancement of the existing operational functioning and of the flash flood hydrological modeling (Anquetin et al., 2010; Braud et al., 2010; Estupina-Borell et al., 2004; Gaume et al., 2004).

In France, Grand Delta flood forecasting service (SPC-GD) handles the monitoring and the forecasting of flash floods, based on the real-time analysis of rain gauges and water level stations. Additionally, the forecasters use the ALHTAÏR model to simulate runoff from all the supervised watersheds with the rainfall intensity deduced from data of the ground-based radar system that covers this

[24] http://www.openmp.org.
[25] http://www.nordugrid.org.
[26] http://www.gridsphere.org.
[27] http://www.gridworkflow.org/snips/gridworkflow/space/Karajan.
[28] http://www.openlayers.org.
[29] http://www.mapserver.org.
[30] http://www.un-spider.org/ – United Nations Platform for Space-based Information for Disaster Management and Emergency Response.

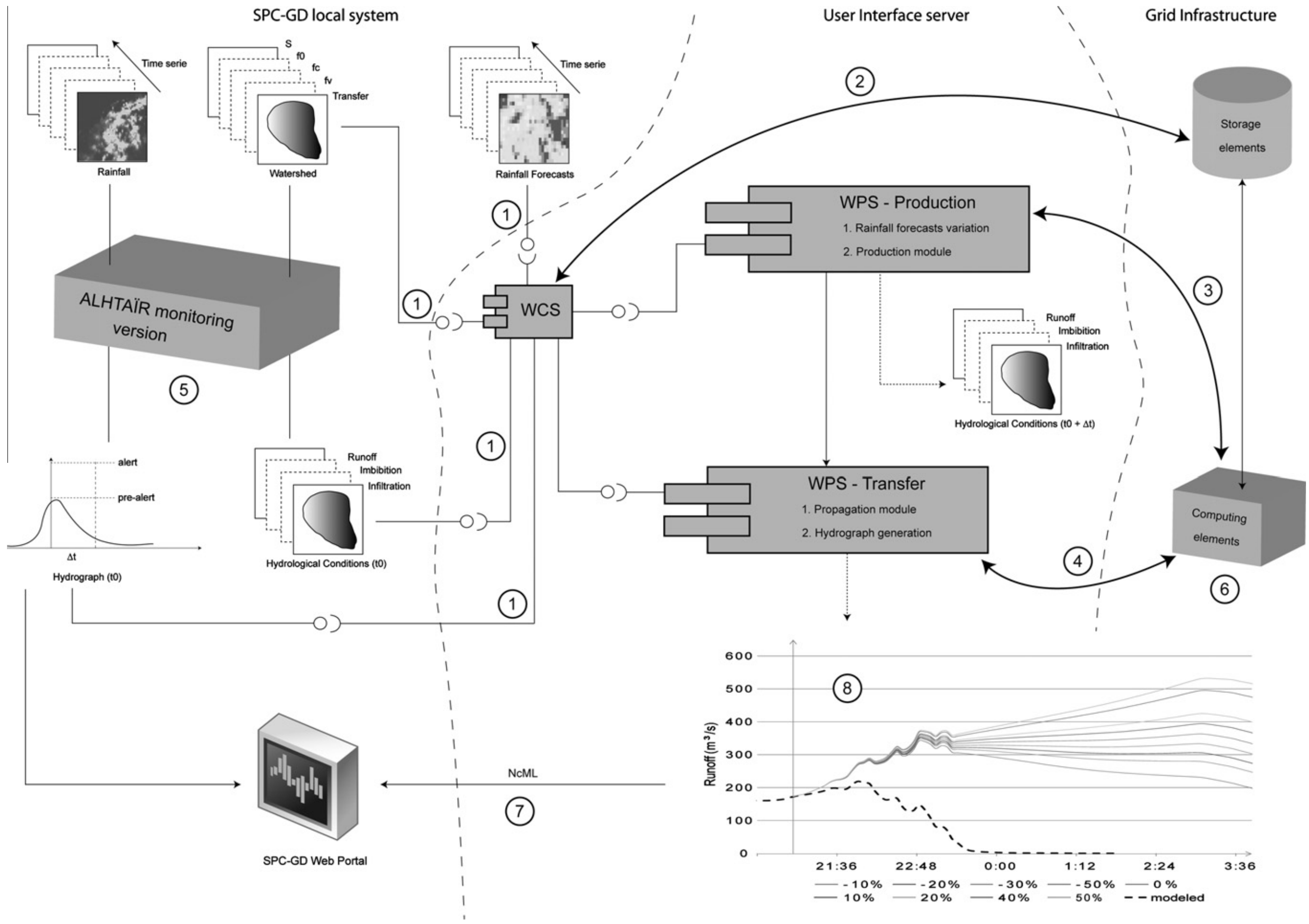


Fig. 3. G-ALTHAIR platform is based on three layers: existing client–server SPC-GD local system, OWS services implemented on the User Interface server and the Grid Infrastructure providing computational and storage resources.

region. This operational model has already enabled realistic hydrological monitoring; however the lack of a high-end computing infrastructure limits its use for short-term operational forecasting. The main scope of this research has been to design a prototype of the hydrological modeling platform G-ALHTAÏR able to operate ALHTAÏR model both in real-time and after the event, under different scenarios (Duband, 2000; Doswell et al., 1996).

The ALHTAÏR is a distributed rainfall-runoff model derived from the Horton principle (Ayral et al., 2007; Horton, 1933). Based on an event-driven functioning, it includes two independent modules:

– The "production module" is dedicated to convert raw rainfall provided by the radars in effective rainfall over the given watersheds. For each cell of the watershed, the local surface runoff is calculated according to the infiltration capacity. New hydrological conditions are simulated and then used as input of the next time step.
– The "propagation module" propagates the surface runoff of each watershed's cell to its outlet (i.e. geo-morphological transfer function).
– The main output of ALHTAÏR is a real-time hydrograph with a forecast horizon equal to the time of concentration of the watershed.

Although from the ground-based radar system one-hour rainfall forecast is provided, the modeled runoff is not always in good agreement with observed values due to uncertainties on rainfalls. In order to take into account the rainfall uncertainties, the simulations are carried out with different rainfall scenarios at different forecast horizons (i.e. time period at which the forecast is delivered). Within an operational lead time (i.e. expected duration to get the simulation output), the forecasters should execute simulations with different hydro-meteorological conditions on each supervised watershed (bunch of hydrographs) and select the most probable upcoming runoffs. The needed compute and storage resources are not available locally at SPC-GD but on the European Grid infrastructure. Hence, G-ALHTAÏR is dedicated to handle these new hydrometeorological capabilities, exploiting the resources provided by the EGI.

Within the framework of the EU FP6 Cyber Infrastructure for Civil protection Operative ProcedureS-CYCLOPS[31] project (2006–2008) such a platform has been developed taking advantage of Grid technology capabilities combined with OGC Web services (Mazzetti et al., 2009). The flash flood application demonstrates the capability of the platform to handle low level Grid tasks for hydrological processing by inexperienced end-users, through simple HTTP/XML request. The prototype required (1) to adapt input/output data (radar, hydrological conditions, watersheds, and hydrographs) to comply with Web services requirements and (2) to re-code ALHTAÏR algorithms to enable their independent execution on the Grid middleware (gLite). A new geospatial Web services layer has been designed on top of the Grid infrastructure in order to exchange data with the existing SPC-GD information system (Mazzetti et al., 2009).

[31] http://www.cyclops-project.eu/.

Fig. 3 shows the G-ALTHAIR platform that is a simplified version of the platform developed in the framework of CYCLOPS. Numbers on the figure are used to describe the different tasks of the workflow. The two Web services, implemented on the User Interface server, are the Web Coverage Service (WCS) to transform initial data to model input data (re-sampling and sub-setting) [1] and to make them accessible to other grid services [2], and the Web Processing Service (WPS) to wrap the ALHTAÏR production [3] and propagation [4] modules. WPS provides an easy customizable processing on the Grid. Existing ALTHAÏR model holds up efficiently real-time hydrological monitoring on the SPC-GD local system [5], while G-ALHTAÏR is mainly dedicated to support on-demand forecasting scenarios on the Grid Infrastructure [6]. WPS automatically manages: WCS for local data retrieving (last hydrological conditions and most recent forecasting rainfall and chosen watershed) [1], production and transfer module execution on worker nodes [6], and output transfer on local system [7]. The output corresponds to a bunch of hydrographs representing every variation of the rainfall intensity required by the forecaster [8]. Forecaster can also provide additional customized meteorological scenarios to G-ALHTAÏR platform.

Through G-ALTHAIR end-users easily submit hydrological simulations in a more intensive way than usual, given the large computational capabilities of the Grid infrastructure. In-progress realistic tests show that SPC-GD, in adopting such system, can potentially benefit from the results of more hydrological modeling instances (Thierion et al., 2011). Taking into account that each forecasting scenario requires two sequences of four elementary grids jobs (three WCS instances and one WPS instance), the tests pointed out that G-ALTHAÏR can potentially perform approx. three hydrological forecasting scenarios for each of the 30 supervised watersheds of SPC-GD, i.e. around 300 elementary grid jobs in the operational lead time (Thierion et al., 2011). However, the EGEE Grid is generally characterized by shared access without any resource reservation. Therefore during the critical period results may not be provided on time due to some Grid-component outages or algorithm software compatibility issues on some sites.

The experimental platform G-ALTHAÏR will easily allow to add other hydrological modules such as kinematic waves transfer function, SCS production function (Gaume et al., 2004), real-time runoff assimilation (Beven, 2001; Collier, 2007; Estupina-Borell et al., 2006; Rabuffetti, 2006). Hence, in a operational situation, forecasters could have, at their disposal, a suite of hydrological modules to rapidly build different processing workflows adapted to the variability of hydrometeorological processes occurring in mountainous watersheds (Beven, 2003; Doswell et al., 1996; Rabuffetti and Barbero, 2005).

Further research is planned to develop in the G-ALHTAÏR platform a discharge assimilation process and hydraulic models for flood area forecast in crisis context. G-ALHTAÏR will be tested under operational service at SPC-GD during the autumn of 2011.

3.4. Computational services for optimal decision making in groundwater protection (AQUAGRID)

Groundwater models are becoming increasingly important in the decision making process as they provide systematic and consistent information on water availability, impacts of climate and land use changes, and analyses of non-point source pollution. The main environmental problem addressed by AQUAGRID is the seawater intrusion in regional coastal aquifers, under explicit consideration of uncertainty. The uncertainty assessment, performed by means of Monte Carlo analysis on the scenario outputs, considers as potential uncertainty sources either hydraulic parameters (Lecca and Cau, 2009) or model stresses (e.g. pumping rates, Kerrou et al., 2010). Running such complex and regional seawater intrusion models on a single PC for a sufficient number of stochastic realizations is often just not feasible. That difficulty can be overcome by using a cluster of computers on which each stochastic realization is executed independently. This is a common technology today, but it still remains rather expensive and not all laboratories have access to such an infrastructure, particularly where hydrogeologists are located in countries with low incomes. This is an issue because also these countries frequently face the problem of seawater intrusion with the most dramatic consequences for their economy. Therefore, a Grid computing infrastructure could be used to run complex and large simulations remotely, and possibly directly from the countries where the problems occur (Kerrou et al., 2010).

Groundwater simulation is at the heart of complex data workflows, composed by nested and linked models of varying degrees of complexity, applied to the same region or problem. The computational engine is the CODESA-3D hydro-geological model (Gambolati et al., 1999) to simulate complex and large coupled density-dependent and variably-saturated 3D groundwater flow and salt transport problems. CODESA-3D is a three-dimensional finite element simulator for groundwater flow and solute transport in variably-saturated porous media on unstructured domains. The flow and solute transport processes are coupled through the variable density of the filtrating mixture made of water and dissolved matter (salt, pollutants). The flow module simulates the water movement in the porous medium, taking into account different forcing inputs: infiltration/evaporation, recharge/discharge, withdrawal/injection, etc., while the transport module computes the migration of the salty plume due to advection and diffusion processes. Model parameters and system excitations are assumed variable in space and/or time. Among the open source third-party software integrated into AQUAGRID workflows are: HYDRO_GEN[32] a spatially distributed random field generator for correlated hydrogeological fields to be used in Monte Carlo analysis, PHREEQC[33] to perform speciation, batch-reaction, one-dimensional transport, and inverse geochemical calculations, PHAST[34] to simulate groundwater flow, solute transport, and multi-component geochemical reactions, PEST[35] to calibrate models by nonlinear optimization of parameters against field measurements, and the Fortran Genetic Algorithm Driver[36] to perform water resources management optimization using a search technique based on natural selection.

AQUAGRID is the subsurface hydrology service of the GRIDA3[37] computing platform, centered on groundwater modeling and geochemical data analysis. It makes available to users a Web framework integrating data sets, modeling, analysis and visualization tools (Lecca et al., 2009). The underlying computing infrastructure, located at CRS4, provides a number of environmental science and engineering software applications on data collections shared among project partners (Murgia et al., 2009). Remote users access and control computing resources via the EnginFrame Grid portal framework[38] and distributed data and metadata via the iRODS middleware.[39] The principal goal achieved using a data-Grid in conjunction with a high-end computing platform is to federate and share know-how, data and computational services across a community of multi-disciplinary experts, via a dedicated Web gateway. Major application fields of AQUAGRID entail the scientific and technical support to the planning and management of optimal aquifer development schemes and monitoring networks, and the selection of the most

[32] http://www.ing.unitn.it/~bellin/frames/hydrogen.php.
[33] http://wwwbrr.cr.usgs.gov/projects/GWC_coupled/phreeqc/.
[34] http://wwwbrr.cr.usgs.gov/projects/GWC_coupled/phast/.
[35] http://www.parameter-estimation.com/.
[36] http://www.cuaerospace.com/carroll/ga.html.
[37] http://grida3.crs4.it.
[38] http://www.enginframe.com.
[39] https://www.irods.org/.

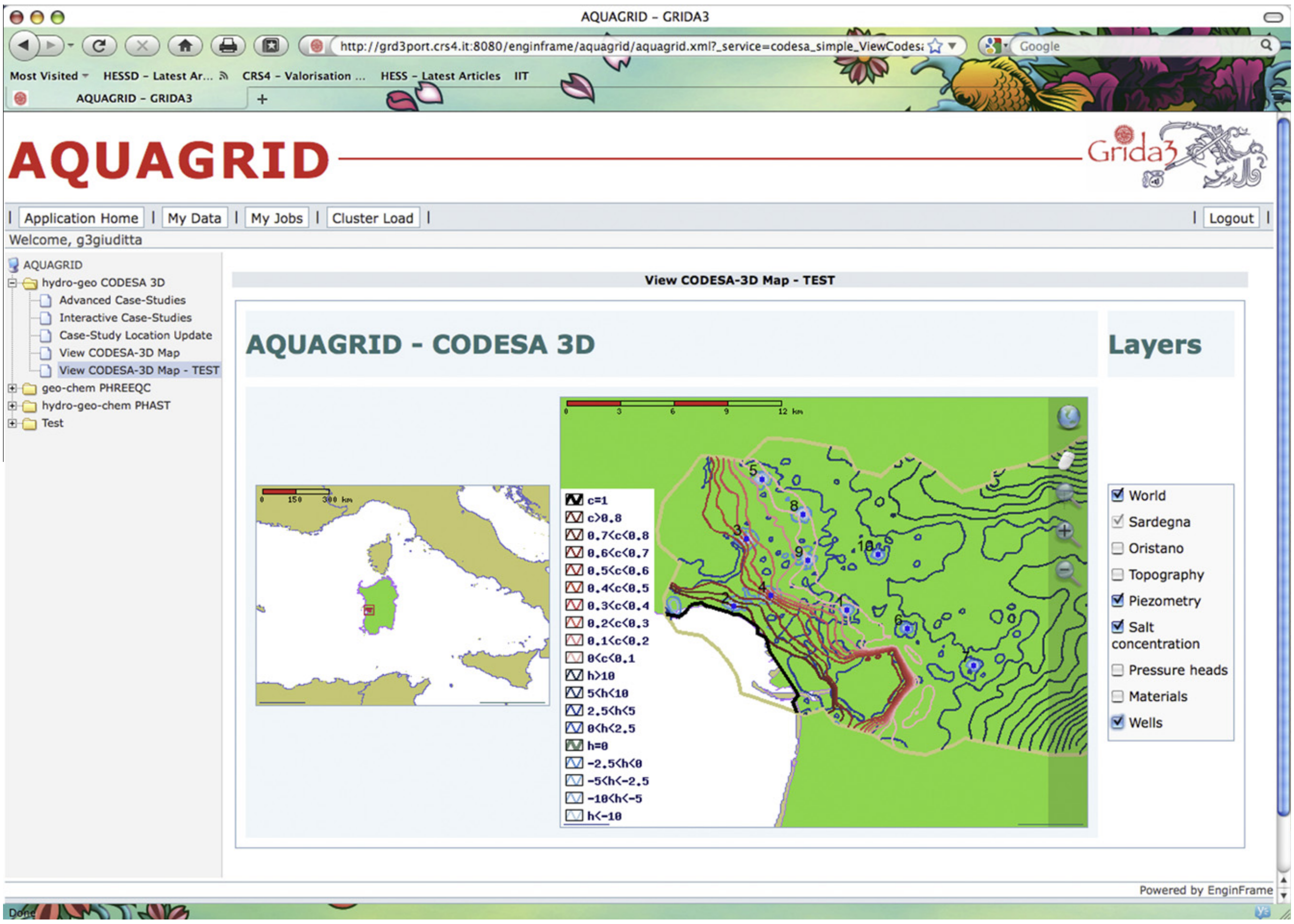


Fig. 4. Piezometric and concentration maps of the seawater intrusion simulation of the Oristano coastal aquifer (Sardinia, Italy) are calculated and displayed on the AQUAGRID computational portal.

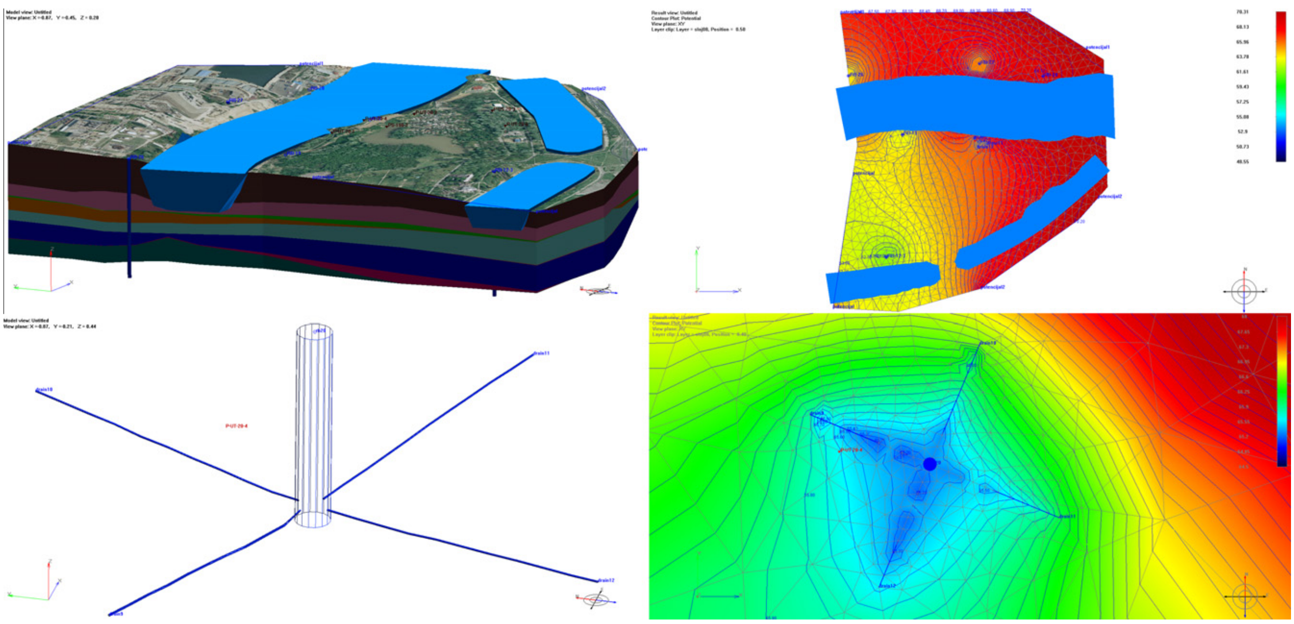


Fig. 5. Topology view of the Belgrade Water Supply Center (top left), single Ranney well with four drains (down left), regional model results – potentials with streamlines (top right) and local model results around the well – potential plot (down right).

effective remediation scenarios. The Web portal currently provides authorized users with three different Grid-enabled interactive modeling applications: (i) 3D coupled groundwater flow and solute transport on four case-studies: Oristano-Italy (Lecca and Cau, 2009), Korba-Tunisia (Kerrou et al., 2010), Gaza-Palestine (Alnahhal et al., 2010) and Oued Laou-Morocco coastal aquifers; (ii) geochemical speciation of aqueous solutions on 900 samples grouped into 20 online Sardinian mine-districts campaigns (Biddau et al., 2008); and (iii) 3D multi-component reactive transport simulations to design and assess the performance of remediation systems as open limestone channels and permeable reactive barriers.

As an example, we consider the 127 km^2 aquifer system of the Oristano coastal plain (Sardinia, Italy) consisting of two main productive units interbedded by a thin, possibly vanishing, clayey aquitard. Using a Monte Carlo technique, a range of aquifer system configurations has been explicitly simulated to study the effect of aquitard spatial discontinuities on the seawater intrusion mechanism (Lecca and Cau, 2009). Fig. 4 shows the groundwater head and salt concentration maps resulting from one of the 100 stochastic transient simulations (all submitted as independent jobs), displayed directly on the infrastructure via the UMN Mapserver[40] Web-GIS interface.

AQUAGRID is an application shared by a small but growing community. The CODESA-3D modeling service has been also deployed in the framework of the EU FP6 Sustainable Water Management in Mediterranean Coastal Aquifers-SWIMED project (2003–2006, Benavente et al., 2004) and of the large European Grid projects (2006–2008) EGEE and EUMedGRID.[41]

3.5. Lizza PAKP Grid-enabled groundwater flow simulation

Groundwater flow dynamic parameters are of the highest importance in the business field of drinking water supply, management and planning. *Lizza–PAKP* is a groundwater flow simulation system that integrates a solver (PAKP) based on finite element (FE) method and a user interface (Lizza). PAKP module includes a solver based on Darcy's law which describes the flow of a fluid through a porous medium (Wang and Anderson, 1982). It also provides fully 3D modeling capabilities, stationary and non-stationary analysis, water saturated and non-saturated regime calculations, as well as mass and heat transport handling. The application was developed in cooperation between Institute for Water Resources ''Jaroslav Cerni'', Belgrade and the University of Kragujevac, Serbia.

Lizza–PAKP introduces innovative solutions to model Ranney wells (Kojić et al., 2007), compared to similar software, including regional and local models together with user-friendliness. A Ranney well is a kind of well that has a center caisson with horizontal perforated pipes extending radially into an aquifer (Fig. 5, down left). It is particularly applicable to the development of thin aquifers at shallow depths. The coupling of 1D elements corresponding to radial pipes of Ranney well and 3D elements corresponding to the surrounding ground material is one of the most important application achievements.

Using the desktop environment *Lizza*, one forms arbitrary shaped 3D model by specifying data in the usual way in hydrological practice, i.e., terrain contours, bottom surface contours and layer material and river-bed characteristics. Boundary conditions are prescribed as specified levels, flow rates and type of wells (tube and Ranney). Input data is specified in the CSV (Comma Separated Values) format, while background maps and textures are allowed to be provided in a number of standard image formats.

While all pre-processing tasks are chosen to be desktop-oriented, the FE analysis and output storage have been moved into the Grid environment. As soon as the Grid-enabled FE analysis finishes, *PAKP* module transfers the huge output files in the Universal File Format to the Grid storage elements. Despite the wide output portability, *Lizza* itself seems to be the most appropriate post-processing and visualization tool according to user preferences. Since this requires the transfer of large output files from remote Grid storage nodes to the local desktop, different solutions are under study to address the issue.

When Grid porting began, Lizza–PAKP was already a mature application with two years of active development behind. At that time, the application was capable to perform heavy computations employing a parallel multi-frontal solver using the MPI (Message Passing Interface) library on a local cluster. The main objectives of the Grid porting were: (1) to speed-up application using additional processing resources, (2) to provide support to simulate much more large and complex models thanks to Grid processing/memory resources, (3) to considerably improve parametric analysis by means of parametric and collection jobs that allow to launch a bulk of jobs using a single submission and to retrieve all the output at once, (4) to employ Grid storage resources in order to store and archive large output files, and (5) to simplify end-user experience.

In order to achieve the above objectives, the Grid porting employed various Grid services provided by the EGEE middleware, gLite, most notably WMS (Workload Management System) and LFC (Logical File Catalog). The application also provides a *Work Binder* interface,[42] one of the EU FP6 SEE-GRID[43] application services, which helps in hiding the Grid complexity from the end-user viewpoint and allows a easier handling of various parametric analyses. Its main purpose is to quickly allocate new jobs for Grid users, together with presenting a more interactive behavior to them (Marovic et al., 2009). It makes also possible to follow calculation progress, then to transfer, present and analyze partial results, without waiting for the Grid job to finish.

The application was employed to successfully model a system of Ranney wells near the Sava River (Belgrade Water Supply Center). The global model includes porous medium on both sides of Sava River, with dimensions approximately 2100 m × 2100 m in the horizontal plane and a depth of 40 m. The total number of finite element nodes in the model discretization was about one million. Boundary conditions consist of impermeable bottom plane, impermeable vertical surface bounding the model, and river modeled by prescribed potential at the river–soil boundary. Fig. 5 shows a topology view of the Belgrade Water Supply Center (top left panel) and a sketch of a single Ranney well with four drains (down left panel), along with regional and local model results (potentials with streamlines, top and down right panels).

Without Grid resource support, it would not be possible to analyze such a huge model and a few dozens of its variants. Single model variant is still suitable to be analysed in a cluster environment, but Grid is a key term in an attempt to analyze multiple model variants concurrently.

3.6. EnviroGRIDS: building capacity for a Black Sea Catchment observation and assessment system

The EU FP7 EnviroGRIDS[44] project (2009–2013) aims at building capacities in the Black Sea region on new international standards to gather, store, distribute, analyze, visualize and disseminate crucial information on past, present and future states of this region in order to assess its sustainability and vulnerability (Lehmann et al., 2009). The project is essentially focused on the Black Sea hydrological Catchment (BSC), whose surface amounts to about 2 M km^2 with a population of 160 M inhabitants over 24 countries. This catchment is subject to numerous environmental pressures and threats that generate several direct consequences such as pollution of surface/

[40] http://mapserver.org/.
[41] http://www.eumedgrid.org.
[42] http://wiki.egee-see.org/index.php/Work_Binder_Application_Service.
[43] http://www.see-grid.org/.
[44] http://www.envirogrids.net/.

groundwater, eutrophication, degradation of biodiversity, and accelerated runoff/erosion (PCU, 1999; Tavitian et al., 2008).

One of the main scientific objectives of EnviroGRIDS is to assess how the sustainability of water usage in this catchment may evolve in the future under different scenarios of change for climate changes, land use, agriculture and demography. To achieve this objective, the Soil and Water Assessment Tool (SWAT; Arnold et al., 1998) model will be used. SWAT is a widely used basin-scale, continuous-time model that is designed to predict the impact of management on water, sediment, and agricultural chemical yields in ungauged watersheds (Gassman et al. 2007). SWAT was already used to simulate the hydrology of large-scale body masses such as the African continent (Schuol et al., 2008), the entire U.S. with river discharges data at around 6000 gauging stations (Arnold et al., 1999), and of twelve large river catchments in India (Gosain et al., 2006). In the EnviroGRIDS project, SWAT will be used to apply a high-resolution (sub-catchment spatial and daily temporal resolution) water balance model to the entire BSC. The key data sets needed as inputs are Digital Elevation Models, soil and land cover maps. The BSC model will be calibrated and validated using long-time series of daily river discharge data, river water quality data, and crop yield data (Abbaspour et al., 2007). Some of this data comes from remotely sensed datasets including soil moisture and meteorological data (rainfall, wind speed, evapo-transpiration, humidity). The transnational nature of the Black Sea catchment makes it very difficult to get the same quantity and quality of data in all areas of the catchment. Raw environmental monitoring data are often limited to distribution because of their commercial value at the national level. In the first phase of the project (ended in mid-2010), the available data from the Consortium were gathered to construct and calibrate a coarse-resolution SWAT model for the full catchment. Recent results obtained with this model (Abbaspour, pers. comm.) include long-term averages of river discharge, precipitation, actual and potential evapo-transpiration, soil moisture and aquifer recharge over the entire catchment. In the ongoing second phase of the project, data policies or agreements are being sought with a maximum of regional institutions in order to access additional high-resolution temporal and spatial data sets.

The Grid infrastructure will be the keystone in computing a high-resolution water balance of the BSC under various environmental scenarios. The first step has been to Grid-enable (gridification) the SWAT toolset in order to run it on the EGEE infrastructure. The main challenge of this task is to be able to spatially partition the input data by independent hydrological sub-catchments, process them as independent jobs on the Grid, and merge the results to reconstruct the complete hydrological network. Using this technique, recent results from our consortium have confirmed the substantial gain in computation time when running large SWAT models. Classical Monte Carlo simulations for SWAT sensitivity analysis will also greatly benefit from the Grid, with associated computational time requirements potentially scaled down by an order of magnitude.

The second step in the gridification process is the appropriate management of large input/output spatial data sets. To this aim, the use of OGC Web services (registered in the GEOSS) on the Grid to access data sets held in a SDI will also be explored (Maué and Kiehle, 2009), although bridging architectural gaps between Grids and SDIs remains challenging (Padberg and Greve, 2009; Giuliani et al., 2011). GIS are the main applications through which SWAT simulations are launched, notably through the ArcSWAT extension to ArcGIS platform (Winchell et al., 2007). The challenge here is to take this into account and develop the least disruptive workflow using SWAT on the Grid. This will be achieved by means of a dedicated GIS module to allow direct access to the Grid infrastructure from within the GIS interface (Fig. 6). The module will link input spatial data to the Grid-enabled SWAT package, which could dramatically help SWAT users scaling up their computational experiments.

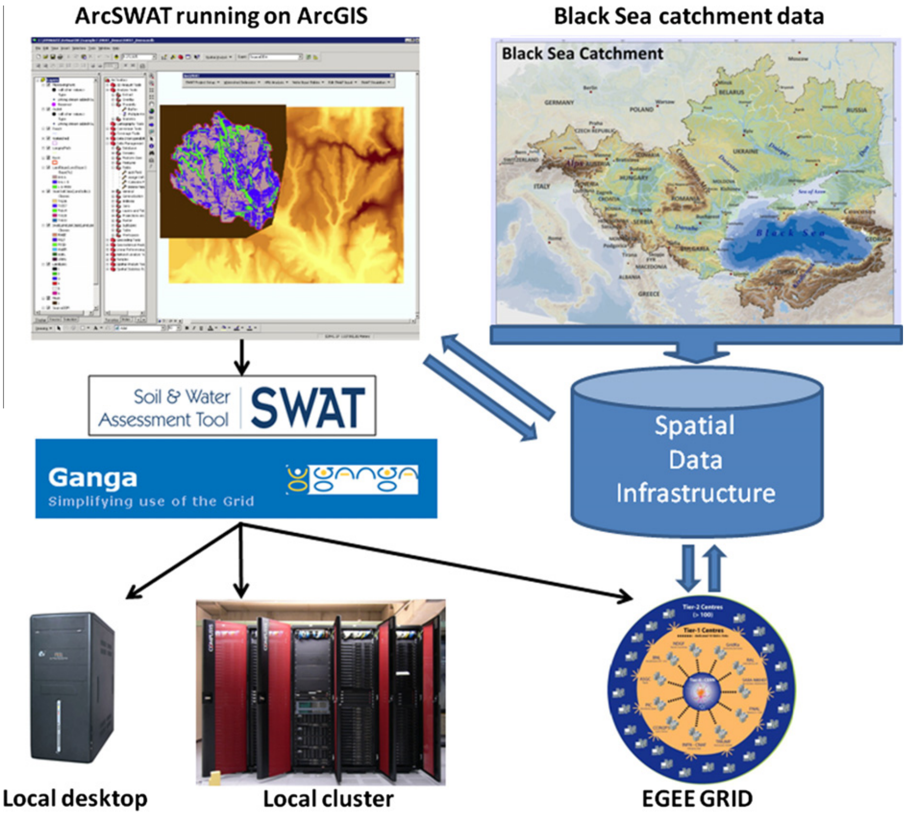


Fig. 6. Schema of the relationship among the various components of the anticipated gridified version of SWAT.

In conclusion, Grid technology is essential for the assessment of the future trends in water balance in the BSC using high-resolution spatial data. In addition to a substantial gain in total simulation time, the use of a Grid infrastructure will certainly be a driving force in the standardization processes of many data sets currently found in various formats in different countries of the region. Moreover, regional organizations, such as the Commission on the Protection of the Black Sea against Pollution and the International Commission for the Protection of the Danube River will be able to take advantage of EnviroGRIDS to analyze large trans-boundary environmental datasets in a harmonized way in order to support the conceptualization and implementation of environmental and relevant sustainable development policies.

4. Concluding remarks and future developments

The paper illustrates a sample of the enormous potential of Grid computing technology in hydrology science and water resources engineering, by means of six Grid-enabled numerical applications. The history and goals of each application are quite different. However their principal motivation is to use the Grid to significantly improve flood prediction, groundwater resources management and comprehensive hydrological survey, all grand challenges for the Civil Society at large.

Results pointed out that the scaling up of computing resources allows to get results on a shorter time interval, to increase the spatial and time resolution of the models and to assess their robustness against uncertainty, that are important issues for decision makers in realistic applications. However, real-time applications still pose some difficulties on the European Grid Infrastructure because a high number of computational nodes is not guaranteed to be available at any given time, without a proper resources reservation mechanism.

In terms of processing capacity, the Grid usually performs considerably better than a local cluster due to the availability of many clusters in the distributed infrastructure instead of a local one. However for the Grid-enabled applications, critical points concern the post-processing and visualization of the results, usually made on a desktop, because of the transfer of large output files stored on remote Grid storage nodes. Different solutions are under study to address these problems.

In addition, as it was shown in most of the applications, Grid technology promotes the use of applications modules by several teams, an efficient cooperation among them, and economies of scale to assemble a critical mass of people and investments. As an example, the use of a Grid infrastructure has been a driving force in the standardization processes of many data sets currently found in various formats in different countries of the Black Sea Catchment area. This standardization process will enable regional organizations to take advantage of EnviroGRIDS to analyze large trans-boundary environmental datasets in a harmonized way to support the conceptualization and implementation of environmental and relevant sustainable development policies.

User-friendly interfaces to the Grid are Web gateways that integrate a community-developed set of tools, applications, and data customized to meet the needs of the targeted community. Some of these science gateways greatly facilitate the use of Grid high-end compute, data, and visualization resources through community-designed interfaces. TeraGrid[45] has developed gateways in various fields, AQUAGRID is an example to a lesser extent, and the EnviroGRIDS and flash flood applications are planning to provide ones open to various communities.

The capacity of Grid technology to federate on a world-wide scale not only computing resources but also experts, is an important issue for research. In Europe several virtual scientific Grid communities have been developed around the Grid. The European-wide Earth Science virtual community is spread over many countries and scientific and technical domains (Cossu et al., 2010). The community building, carried out in the last years, has enabled flow of information, knowledge exchange and common proposals and actions. The next step will be to set-up a collaboration framework in order to define a common strategy and to coordinate new developments.

Acknowledgments

The authors wish to acknowledge colleagues P.A. Ayral (EMA) and V. Tran (IISAS) who participated to the porting of the applications and contributed to the paper with fruitful discussions. The AQUAGRID application is supported by the GRIDA3 project (MIUR FAR D.Lvo 297/99No. 1433/2006) and Sardinian Regional Authorities. The multi-model multi-scale flood workflow application is supported by projects ADMIRE FP7-215024, APVV DO7RP-0006-08, DMM VMSP-P-0048-09, EGEE-III FP7-222667 and VEGA 2/0184/10. The Lizza–PAKP application is supported by the EU FP7 SEE-GRID-SCI initiative (Contract No. 211338). Nicholas Ray is supported by the EU FP7 EnviroGRIDS project (Grant Agreement No. 227640).

Appendix A

See Table 2.

Table 2
List of acronyms.

ARC	Advanced Resource Connector
ASAR	Advanced Synthetic-Aperture Radar
BSC	Black Sea hydrological Catchment
CERN	European Organisation for Nuclear Research
CEODE-CAS	Center for Earth Observation and Digital Earth of Chinese Academy Science
CSV	Comma Separated Values
CYCLOPS	Cyber Infrastructure for Civil protection Operative ProcedureS
DEM	Digital Elevation Model
DRIHMS	Distributed Research Infrastructure for hydro-meteorology Study
EGEE	Enabling Grids for E-Science
EGI	European Grid Infrastructure
ENVISAT	Earth Observation Satellite
EO	Earth Observation
ERM	Empirical Regressive Model
ERS	European Remote Sensing Satellite system
ES	Earth Science
ESA	European Space Agency
ESR	Earth Science Research – Virtual Organisation
EU	European Union
FE	Finite Element
FESWMS	Finite Element Surface Water Modeling System
FP6/FP7	Framework Programme 6 or 7
GEOSS	Global Earth Observation System of Systems
GGF	Global Grid Forum
GIS	Geographical Information System
GMES	Global Monitoring for Environment and Security
GRAM	Grid Resource Allocation and Management
GRASS	Geographic Resources Analysis Support System
HSPF	Hydrologic Simulation Program Fortran
INSPIRE	Infrastructure for Spatial Information in the European Community
KML	Keyhole Markup Language
LFC	Logical File Catalog
LHC	Large Hadron Collider

(continued on next page)

[45] http://www.teragrid.org.

LIDAR	Light Detection and Ranging
MPI	Message Passing Interface
NASU	National Academy of Sciences of Ukraine
NCAR	National Center for Atmospheric Research (USA)
NSAU	National Space Agency of Ukraine
OGC	Open Geospatial Consortium
OGF	Open Grid Forum
OGSA	Open Grid Service Architecture
OWS	Open Geospatial Web Services
PSU	Penn State University (USA)
SAR	Synthetic-Aperture Radar
SCS	Soil Conservation Service
SDI	Spatial Data Infrastructure
SEE-GRID	South-Eastern European Grid-enabled e-Infrastructure Development
SHMI	Slovak Hydro- Meteorological Institute
SWAT	Soil and Water Assessment Tool
SWS	Sensor Web Service
UNICORE	Uniform Interface to Computing Resources
UNEP	United Nations Environment Programme
UMN	University of Minnesota (USA)
VO	Virtual Organisation
WCS	Web Coverage Service
WMS	Web Mapping Service (3.3)
WMS	Workload Management System (3.6)
WPS	Web Processing Service
WRF	Weather Research and Forecasting model
WRI	Water Research Institute

References

Abbaspour, K.C., Yang, J., Maximov, I., Siber, R., Bogner, K., Mieleitner, J., Zobrist, J., Srinivasan, R., 2007. Spatially-distributed modelling of hydrology and water quality in the pre-alpine/alpine Thur watershed using SWAT. J. Hydrol. 333, 413–430.

Alnahhal, S., Afifi, S., Qahman, K., Dentoni, M., Lecca, G., 2010. A Simulation/Optimization approach to manage groundwater resources in the Gaza aquifer (Palestine). In: J. Carrera (Ed.), Proceedings of the XVIII International Conference on Water Resources CMWR 2010. CIMNE, Barcelona. <http://congress.cimne.com/cmwr2010/Proceedings/docs/p35.pdf>.

Anquetin, S., Braud, I., Vannier, O., Viallet, P., Boudevillain, B., Creutin, J.-D., Manus, C., 2010. Sensitivity of the hydrological response to the variability of rainfall fields and soils for the Gard 2002 flash-flood next term event. J. Hydrol. 394, 134–147.

Arnold, J.G., Srinivasan, R., Muttiah, R.S., Williams, J.R., 1998. Large area hydrologic modeling and assessment. Part I: Model development. J. Am. Water Resources Assoc. 34, 73–89.

Arnold, J.G., Srinivasan, R., Muttiah, R.S., Allen, P.M., 1999. Continental scale simulation of the hydrologic balance. J. Am. Water Resources Assoc. 35, 1037–1051.

Ayral, P.-A., Sauvagnargues-Lesage, S., Bressand, F., Gay, S., 2007. Forecasting flash floods with an operational model: application in the South-East of France (Gard). In: Begum S., Stive M.J.F., Hall J.W. (Eds.), Adv. in Natural and Technol. Hazards Res. Flood risk management in Europe: Innovation in Policy and Practice, vol. 25, pp. 335–352.

Benavente, J., Larabi, A., El Mabrouki, K. (Eds.), 2004. Monitoring, Modeling and Management of Coastal Aquifers. Water Research Institute, University of Granada. ISBN:84-699-7903-5.

Beven, K.J., 2001. Rainfall–Runoff Modelling: The Primer. Wiley.

Beven, K.J., 2003. On environmental models of everywhere on the GRID. Hydrol. Process. 17, 171–174.

Biddau, R., Lecca, G., Fanfani, L., Murgia, F., 2008. AQUAGRID: a grid services for advanced modeling and geochemical data analysis in groundwater resources management. Rendiconti Online Soc. Geol. Ital. 3 (1), 112–113.

Bourras, C., Giannaka, E., Tsiatos, E., 2009. e-Collaboration concepts, systems and applications. In: N. Kock, N. (Ed.), Information Science Reference, E-Collaboration: Concepts, Methodologies, Tools, and Applications, vol. 1 (Section I, Chapter 1.2).

Braud, I., Roux, H., Anquetin, S., Maubourguet, M.M., Manus, C., Viallet, P., Dartus, D., 2010. The use of distributed hydrological models for the Gard 2002 flash flood next term event: Analysis of associated hydrological processes. J. Hydrol. 394, 162–181.

Collier, C.G., 2007. Flash flood forecasting: what are the limits of predictability? Quart. J. Roy. Meteorol. Soc. 133, 3–23.

Cossu, R., Petitdidier, M., Linford, J., Badoux, V., Fusco, L., Gotab, B., Hluchy, L., Lecca, G., Murgia, F., Plevier, C., Renard, P., Schwichtenberg, H., Som de Cerff, W., Tran, V., Vetois, G., 2010. A roadmap for a dedicated Earth Science Grid platform. Earth Sci. Informatics 3 (3), 135–148. doi:10.1007/s12145-010-0045-4.

Delrieu, G., Ducrocq, V., Gaume, E., Nicol, J., Payrastre, O., Yates, E., Kirstetter, P., Andrieu, H., Ayral, P., Bouvier, C., Creutin, J.-D., Livet, M., Anquetin, S., Lang, M., Neppel, L., Obled, C., Parent-du-Châtelet, J., Saulnier, G.-M., Walpersdorf, A., Wobrock, 2005. The catastrophic flash-flood event of 8–9 September 2002 in the Gard region, France. A first case study for the Cévennes-Vivarais Mediterranean Hydrometeorological Observatory. J. Hydrometeorol. 6, 34–52.

Doswell, C.A., Brooks, H., Maddox, R., 1996. Flash flood forecasting: an ingredients-based methodology. Weather Forecast. 11, 560–581.

Duband, D., 2000. Thinking over hydrometeorological forecasting. Houille Blanche 1, 42–50.

Estupina-Borell, V., Llovel, C., Maubourguet, M. M., Chorda, J., Dartus, D., Ababou, R., Alquier, M., 2004. Flash flood modeling for prediction, warning and risk assessment. In: ''Trends der Wasserwirtschaft: reagieren oder agieren''/33. IWASA, Internationales Wasserbau-Symposium, Aachen 2003. Shaker Verlag, pp. 175–223. ISBN:3 8322-3083-1.

Estupina-Borell, V., Dartus, D., Ababou, R., 2006. Flash flood modelling with the MARINE hydrological distributed model. Hydrol. Earth Syst. Sci. Discuss. 3, 3397–3438.

Feller, M., Foster, I., Martin, S., 2007. GT4 GRAM: a functionality and performance study. In: TERAGRID 2007 Conference, Madison, Wisconsin, USA.

Foster, I., 2002. The grid: a new infrastructure for 21st century science. Physics Today 55 (2), 42–47.

Foster, I., Kesselman, C., 1999. The Grid: Blueprint for a New Computing Infrastructure. Morgan Kaufmann.

Fusco, L., Cossu, R., Retscher, C., 2007. Open grid services for Envisat and earth observation applications. In: Plaza, A.J., Chang, C.-I. (Eds.), High Performance Computing in Remote Sensing, first ed. Taylor & Francis Group, New York, pp. 237–280.

Gambolati, G., Putti, M., Paniconi, C. 1999. Three-dimensional dimensional model of coupled density-dependent flow and miscible salt transport. In: Bear, J., et al. (Eds.), Seawater Intrusion in Coastal Aquifers – Concepts Methods and Practices. Kluwer Academic, Dordrecht, The Netherlands, pp. 315–362 (Chapter 10).

Gassman, P.W., Reyes, M.R., Green, C.H., Arnold, J.G., 2007. The soil and water assessment tool: historical development, applications, and future research directions. Trans. ASAE 50 (4), 1211–1250.

Gaume, E., Livet, M., Desbordes, M., Villeneuve, J., 2004. Hydrological analysis of the river Aude, France, flash flood on 12 and 13 November 1999. J. Hydrol. 286, 135–154.

Giuliani, G., Ray, N., Lehmann, A., 2011. Grid-enabled spatial data infrastructure for environmental sciences: challenges and promises. Fut. Generat. Comput. Syst. 27 (3), 292–303.

Gosain, A.K., Rao, S., Basuray, D., 2006. Climate change impact assessment on hydrology of Indian River basins. Curr. Sci. 90, 346–353.

Habala, O., Mamiska, M., Hluchy, L., 2007. Service-based flood forecasting simulation cascade in K-Wf grid. In: Steffen, Unger (Eds.), K-WfGrid – The Knowledge-based Workflow System for Grid Applications, Proceedings of CGW'06, vol. II. Marian Bubak, pp. 138–145. ISBN:978-83-915141-8-4.

Hluchy, L., Nguyen, G.T., Halada, L., Tran, V.D., 2002. Parallel flood modelling. In: D'Hollander, E.H., et al. (Eds.), Parallel Computing – Advances and Current Issues. Proceedings of the International Conference ParCo 2001. Imperial College Press, pp. 157–164. ISBN:1-86094-315-2.

Horritt, M.S., 2006. A methodology for the validation of uncertain flood inundation models. J. Hydrol. 326, 153–165.

Horton, R.E., 1933. The role of infiltration in the hydrological cycle. Trans. Am. Geophys. Union 14, 446–460.

Kerrou, J., Renard, P., Lecca, G., Tarhouni, J., 2010. Grid-enabled Monte Carlo analysis of the impacts of uncertain discharge rates on seawater intrusion in the Korba aquifer (Tunisia). Hydrol. Sci. J. 55 (8), 1325–1336.

Kojić, M., Filipović, N., Stojanović, B., Ranković, V., Krstić, M., Otašević, L., Ivanović, M., Nedeljković, M., Dimkić, M., Tričković, M., Pušić, M., Boreli-Zdravković, D., Đurić, D., 2007. Finite element modeling of underground water flow with Ranney wells. Water Sci. Technol.: Water Supply 7, 41–50.

Kussul, N., Shelestov, A., Skakun, S., 2008. Grid system for flood extent extraction from satellite images. Earth Sci. Informatics 1 (3–4), 105–117, doi:10.1007/s12145-008-0014-3.

Kussul, N., Shelestov, A., Skakun, S., 2009. Grid and sensor Web technologies for environmental monitoring. Earth Sci. Informatics 2 (1–2), 37–51. doi:10.1007/s12145-009-0024-9.

Lanig, S., Zipf, A., 2009. Interoperable processing of digital elevation models in grid infrastructures. Earth Sci. Informatics 2 (1–2), 107–116. doi:10.1007/s12145-009-0030-y.

Lecca, G., Cau, P., 2009. Using a Monte Carlo approach to evaluate seawater intrusion in the Oristano coastal aquifer: a case study from the AQUAGRID collaborative computing platform. Phys. Chem. Earth 34, 654–661. doi:10.1016/j.pce.2009.03.002.

Lecca, G., Lai, C., Murgia, F., Biddau, R., Fanfani, L., Maggi, P., 2009. AQUAGRID: an extensible platform for collaborative problem solving in groundwater protection. Earth Sci. Informatics 2 (1–2), 83–95. doi:10.1007/s12145-009-0019-6.

Lehmann, A., Giuliani, G., Ray, N., Allenbach, K., Charvat, K., Gorgan, D., Gvilava, M., Bakuradze, T., Sözen, S., Goksel, C., the EnviroGRIDS Consortium, 2009. The Black Sea Catchment observation system built on a grid-enabled spatial data infrastructure. In: Charvat, K., Alberts, M., Horajova, S. (Eds.), INSPIRE, GMES and GEOSS Activities, Methods and Tools Towards a Single Information Space in Europe for the Environment. Wirelessinfo Tehnologiju Attistibas Forums, Riga, pp. 42–58.

Marovic, B., Potocnik, M., Ivanovic, M., 2009. Work binder application service. In: 4th EGEE User Forum/OGF 25 and OGF Europe's 2nd International Event, 2–6

March 2009, Catania, Italy <http://www.bioirc.ac.rs/media/WB-AS-EGEE%20-UF4-2009.ppt>.
Maué, P., Kiehle, C. 2009. Grid technologies for geospatial applications; an overview. Special issue of GIS. Science on Grid Computing. GIS. Science 3, 65–67. ISSN:1430-3663. <http://ww.gis-biz.de>.
Mazzetti, P., Nativi, S., Angelini, V., Verlato, M., Fiorucci, P., 2009. A Grid platform for the European Civil Protection e-Infrastructure: the forest fires use scenario. Earth Sci. Informatics 2 (1–2), 53–62. doi:10.1007/s12145-009-0025-8.
Murgia, F., Biddau, R., Concas, A., Demontis, R., Fanfani, L., Heilmann, Z., Lai, C., Lecca, G., Lorrai, E., Marrocu, M., Marrone, V., Muscas, L., Peneva, E., Piras, A., Pisu, M., Pusceddu, G., Satta, G., Theis, D., Vacca, A., Valera, P., Vallenilla, Ferrara, A.M., Bonomi, E., 2009. GRIDA3—a shared resources manager for environmental data analysis and applications. Earth Sci. Informatics 2 (1–2), 5–21. doi:10.1007/s12145-009-0020-0.
Padberg, A., Greve, K., 2009. Gridification of OGC web services: challenges and potential. Special issue of GIS.Science on Grid Computing. GIS Science 3, 77–81. ISSN:1430-3663. <http://ww.gis-biz.de>.
PCU (Programme Co-ordination Unit), 1999. Strategic Action Plan for the Danube River Basin 1995–2005 – Revision 1999. Danube Pollution Reduction Programme. <http://www.ceit.sk/wwwisis/sap1.htm>.
Rabuffetti, D., 2006. Discharge assimilation in a distributed flood forecasting model. Adv. Geosci. 7, 355–360.
Rabuffetti, D., Barbero, D., 2005. Operational hydro-meteorological warning and real-time flood forecasting; the Piemonte region case study. Hydrol. Earth Syst. Sci. 9, 457–466.
Renard, P., Badoux, V., Petitdidier, M., Cossu, R., 2009. Grid computing for earth science. EOS, Trans. Am. Geophys. Union 90 (14), 117–119.
Sauvagnargues-Lesage, S., Ayral, P.A., 2007. Using GIS for emergency management: a case study during the 2002 and 2003 flooding in south-east France. Int. J. Emergency Manage. 4, 682–703. doi:10.1504/IJEM.2007.015738.
Schuol, J., Abbaspour, K.C., Yang, H., Srinivasan, R., Zehnder, A.J.B., 2008. Modelling blue and green water availability in Africa. Water Resources Res. 44 (W07406), 18.
Shelestov, A., Kussul, N., Skakun S., 2006. Grid technologies in monitoring systems based on satellite data. J. Autom. Inform. Sci. 38, 69–80. doi:10.1615/J.Automat.Inf.Scien.v38.i3.50.
Tavitian, N., Anderson, J., Chiavari, J., Rubel, O., Renshaw, N., Wittich, A., Maas, A., Carius, A., Dupont, C. 2008. Greening the Black Sea synergy. WWF – World Wide Fund for Nature and Heinrich Böll Foundation, Brussels. <http://assets.panda.org/downloads/black_sea_full_report.pdf>.
Thierion, V., Ayral, P.-A., Giesel, J., Sauvagnargues-Lesage, S., Payrastre, O., 2011. Grid technology reliability for flash flood forecasting: end-user assessment. J. Grid Computing, in press, doi:10.1007/s10723-010-9173-9.
Wang, H.F., Anderson, M.P., 1982. Introduction to Groundwater Modeling: Finite Difference and Finite Element Methods. Academic Press. ISBN:012734585X.
Wang, L., Jie, W., Chen, J., 2009. Grid Computing: Infrastructure, Service, and Applications. CRC Press. ISBN:9781420067668.
Winchell, M., Srinivasan, R., DiLuzio, M., Arnold, J. 2007. ArcSWAT Interface for SWAT User's Guide. Blackland Research Center, Temple, Texas. pp. 431.