



Chapitre d'actes

2011

Open Access

This version of the publication is provided by the author(s) and made available in accordance with the copyright holder(s).

Grid Based Data Processing Tools and Applications for Black Sea Catchment Basin

Gorgan, Dorian; Abbaspour, Karim; Cau, Pierluigi; Bacu, Victor; Mihon, Danut; Giuliani, Gregory; Ray, Nicolas; Lehmann, Anthony

How to cite

GORGAN, Dorian et al. Grid Based Data Processing Tools and Applications for Black Sea Catchment Basin. In: The 6th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications. Prague, Czech Republic. [s.l.] : [s.n.], 2011.

This publication URL: <https://archive-ouverte.unige.ch/unige:18367>

Grid Based Data Processing Tools and Applications for Black Sea Catchment Basin

Dorian Gorgan¹, Karim Abbaspour², Pierluigi Cau³, Victor Bacu¹, Danut Mihon¹,
Gregory Giuliani⁴, Nicolas Ray⁴, Anthony Lehmann⁴

¹ Technical University of Cluj-Napoca, {dorian.gorgan, victor.bacu, vasilie.mihon}@cs.utcluj.ro

² EAWAG, Swiss Federal Institute for Aquatic Science and Technology, karim.abbaspour@eawag.ch,

³ Center for Advanced Studies, Research and Development in Sardinia, pierluigi.cau@gmail.com

⁴ University of Geneva, (giuliani@grid.unep.ch, nicolas.ray@unige.ch, anthony.lehmann@unige.ch)

Abstract—This paper highlights the main functionalities, tools and applications provided by the BSC-OS portal, developed through the enviroGRIDS project. The presentation highlights the main issues of interoperability between geospatial and grid infrastructures, and between different platforms supporting the Earth Science oriented tools and applications. Other subjects such as spatial data management, hydrological model calibration, environmental scenario development and running, satellite image classification and data mining, tools and applications for data reporting and scenarios visualization, and development of Earth Science oriented training materials, are approached by the BSC-OS portal development.

Keywords—*spatial data; grid based processing; hydrological model; satellite images; data visualization; Earth Science; training materials*

I. INTRODUCTION

The subjects of ecologically unsustainable development and inadequate resource management, in the context of climate, land cover and population changes, in the region of Black Sea watershed, are the main concern of the enviroGRIDS (Black Sea Catchment Observation and Assessment System supporting Sustainable Development) FP7 project [1]. It is extremely important to better understand how the quantity and quality of waters will vary in the catchment over the coming decades. The project aims to develop and make available the hydrological model of the Black Sea Catchment by four main achievements: (a) collection of large transnational data sets; (b) adequate management and sharing processes of the environmental data by a dedicated Spatial Data Infrastructure (SDI); (c) distributed computing in order to allow running a high-resolution hydrological model; and (d) provides a set of tools and applications to specialists and citizens in order to access data processing and visualization, and analyze environmental scenarios.

One challenge of the enviroGRIDS project is the interoperability between geospatial and Grid infrastructures by providing extended features of the both technologies. The geospatial technologies offer very specialized functionalities for Earth Science oriented

applications, meanwhile the Grid oriented technology is able to support scalable, distributed, and parallel processing.

The enviroGRIDS system resources are accessible to the large community of users through the BSC-OS (Black Sea Catchment Observation System) portal that provides web applications for data management, hydrologic models calibration and execution, satellite image processing, report generation and visualization, and virtual training center.

This paper presents the BSC-OS portal architecture and the main challenges and issues regarding the development of environmental tools and applications concerning the Black Sea catchment.

The paper is structured as follows. Section 2 presents the works and achievements related with the enviroGRIDS project. Section 3 sketches the portal architecture and the set of tool and application categories. Each of the next six sections describes a tool and application category such as data management, SWAT model calibration and scenario execution, satellite image processing, spatial data visualization and reporting, two demonstrator applications oriented to citizens and decision makers, and virtual training center. The last section concludes on the portal development and future work.

II. RELATED WORKS

The enviroGRIDS project aims to use the SWAT model as a high-resolution (i.e., sub-catchment spatial and daily temporal resolution) water balance model to the entire Black Sea Catchment. The model will be calibrated and validated using river discharge data, river water quality data, and crop yield data as in [2]. The Black Sea watershed related hydrological model is very complex due to the highly interconnected and continuously evolving interactions at many spatial and temporal scales, and requires to gather and integrate different sets of environmental data (e.g. physical, chemical, biological) [3]. Other European projects aim environmental related subjects [4]. IS-ENES project develops the European Network for Earth System Modeling (ENES), which calls together the European climate/Earth system modeling

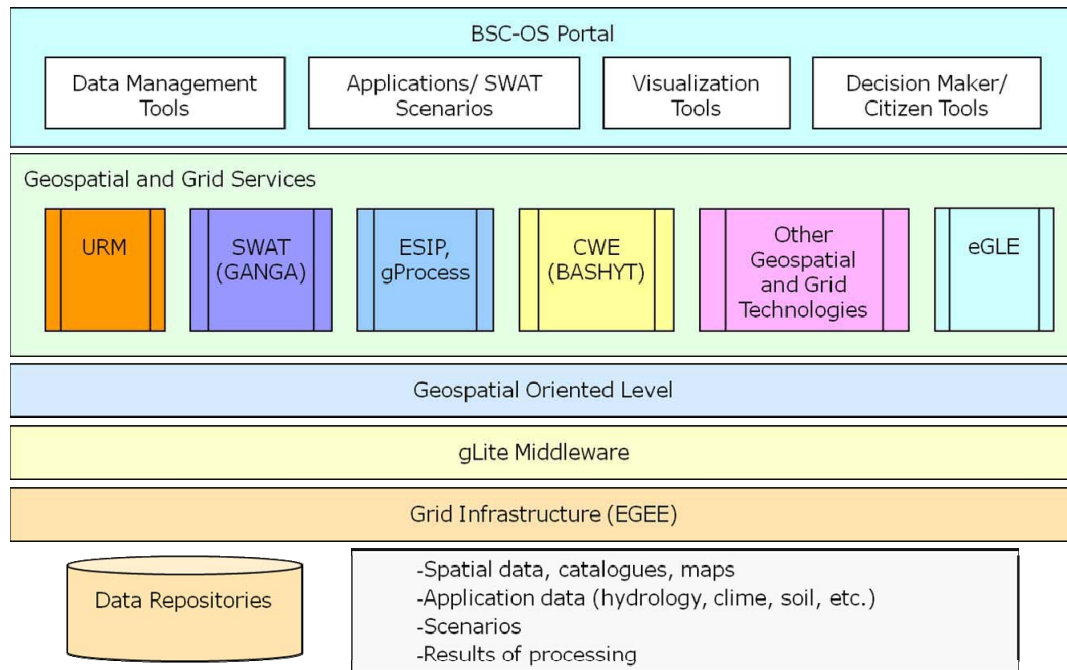


Figure 1. BSC-OS portal architecture.

community in order to work on understanding and prediction of future climate change. The ENSEMBLES project was a joined effort to develop an ensemble prediction system for climate change based on the principal state-of-the-art, high resolution, global and regional Earth System models developed in Europe. The METAFOR project addresses the fragmentation and gaps in availability of metadata for climate data that are currently found in existing repositories. The goal of the DRIHMS project is to systematically build a bridge between the HMR (Hydro-Meteorological Research) and ICT (Information and Computing Technology) communities, and to identify requirements of HMR users and match them to capabilities of the newly developed ICT infrastructure. The GENESI-DEC project aims to provide guaranteed, reliable, easy, effective access to a variety of data, facilities, tools and services to an ever increasing number of Digital Earth users from all disciplines.

The projects EGEE, SEEGRID-SCI, and C3Grid, provide solutions for sharing complex spatial and environmental data sets, and Grid based processing tools and applications. The aim of the C3Grid project for instance, is to create a grid-based working environment for earth system research.

Many other European projects such as SAW-GEO, CYCLOPS, GDI-Grid, GEO-Grid, DEEGREE, DORII, and GENESI-DR address the management of spatial data and environmental tools and applications.

The enviroGRIDS project gathers solutions and experience from all these mentioned projects in order to

approach the particularity of the Black Sea Catchment in terms of SDI, platforms interoperability (i.e. geospatial and Grid, and software platforms like URM, gSWAT, ESIP, eGLE, etc), high resolution models, processing scalability, user interaction usability, and processing efficiency.

III. BSC-OS PORTAL ARCHITECTURE

The BSC-OS portal consists of a set of web applications through which the users access the system resources such as spatial data, hydrologic models, environmental scenarios, data processing tools, visualization facilities, environmental reports, and training materials (Figure 1).

There are five categories of users such as data providers, earth science specialists, decision makers, citizens, and system administrators.

The main user tool and application categories provided by the portal are: *Data management* – provides the user with spatial data management and operations. The user may enter data and metadata, visualize, modify, update, and remove spatial data from the data repositories; *Hydrologic model management* – provides the Earth Science specialists with hydrologic model configuration, scenario and model development, model calibration and scenario running. One of the water quality models used in the enviroGRIDS project is SWAT (Soil Water Assessment Tool) [5]. It is a model designed to estimate impacts of land management practices on water quantity and quality in complex watersheds. The SWAT model requires specific information about weather, soil

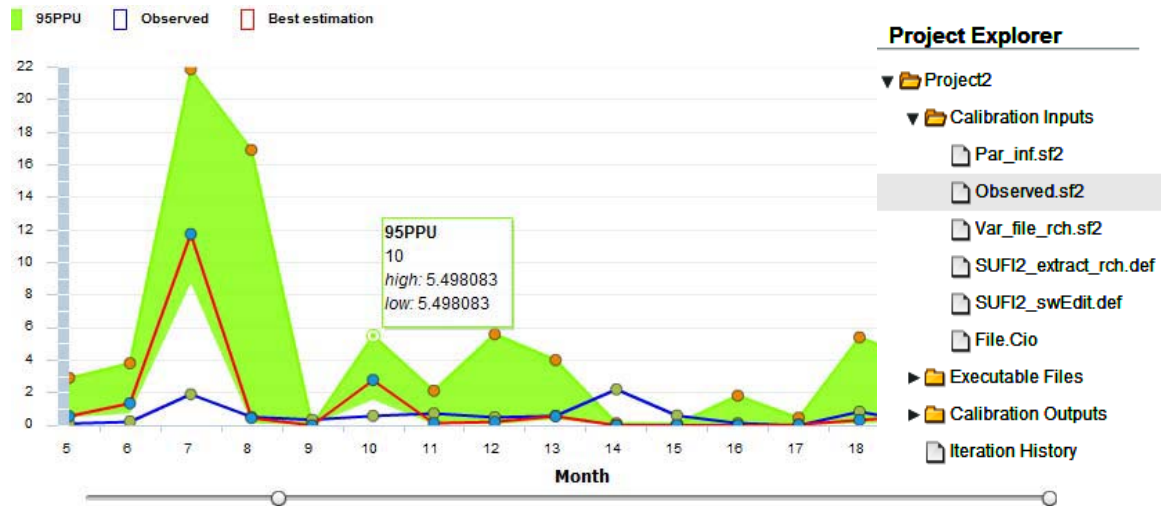


Figure 2. Detailed visualization of the gSWAT calibration results.

properties, topography, vegetation, and land management practices of the watershed; *Satellite data processing* – the specialist may process satellite data and images in order to search for relevant information (e.g. land cover, vegetation, water, soil composition, etc); *Data visualization and report* – the specialists visualize various spatial data in different formats and views and compose environmental reports for decision makers and citizens; *Decision maker and citizen applications* – provide the decision makers with the interactive and graphical tools to access the private environmental reports. The user may visualize data that make possible statistical analysis and predictions; *Virtual training center* – supports the specialists to develop Earth Science oriented training materials and the users to execute the lessons.

The regular users visualize the reports generated by the specialists as results of executing environmental scenarios. The input data for the reports are built up by the specialists by running hydrological models of the Black Sea catchment area and by processing related satellite data. All data sets required for building up the hydrological models, environmental scenarios, and spatial models are provided and entered into the system by the data providers.

IV. DATA MANAGEMENT

URM (Uniform Resource Management System) allows users to search and share spatial and non-spatial information and establish a network to encourage a broader community to adopt and support the GEOSS concept of data sharing for a more sustainable environment [6]. The URM Geoportal is not the one integrated solution, but a set of modules and services, which are able to communicate through interoperable services defined by OGC (Open Geospatial Consortium), and W3C (World Wide web Consortium). URM

Geoportal consists of four basic blocks interconnected through metadata:

1. *Metadata management* is supported by the MicKa toolset for editing and management of metadata for spatial information, web services and other sources;
2. *Data management* by the DataMan application. It supports the import, export, and management of spatial data in files or databases, for both raster (IFF/GeoTIFF, JPEG, GIF, PNG, BMP, ECW) and vector (ESRI Shapefile, DGN, DWG, GML) data types;
3. *Data visualization*, provided by the MapMan software tool. It supports publication of spatial compositions from locally stored data with external WMS (web Map Service), WFS (web Feature Service) data services;
4. *Content management* (publishing of context and connection with social networks) supported by the SimpleCMS toolset.

V. SWAT MODEL CALIBRATION AND SCENARIO EXECUTION

The SWAT model supports the specialists on making predictions on the effects of management decisions on water, sediment, nutrient and pesticide yields with reasonable accuracy on large, engaged river basins [7]. The data package of the model could be quite large (up to 20 thousands) and its running requires great storage capacity and high power computation resources.

A. gSWAT Application

The gSWAT application has been developed in enviroGRIDS project and available through the BSC-OS Portal in order to support the development, calibration and execution of the SWAT model [8]. Grid based computation infrastructure is the basic solution for parallel

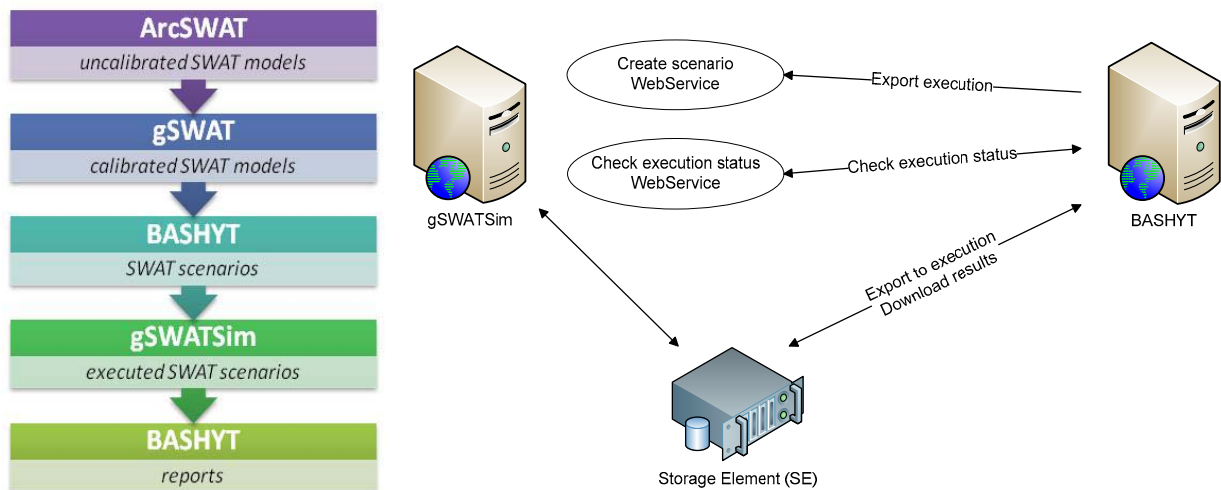


Figure 3. Interaction flow between gSWATSim and BASHYT.

and distributed processing of the hydrological model in the gSWAT application.

It is developed as a web application that hides to the user the complexity of the Grid infrastructure (Figure 2). The application provides support for scalable models in terms of geographical area, modeling resolution, and number of users. Multicore architecture and GPU cluster based solutions are explored as well in order to speed up and optimize the hydrological model processing [9].

B. SWAT Oriented Services

gSWATSim is a server side extension of the gSWAT platform that is exposed as a collection of REST web Services supporting the user to create new projects (i.e. new scenarios), modify some information related to projects (e.g. project name, description, etc.), run environmental scenarios, upload results to visualization module (i.e. BASHYT), and visualize the execution status of scenarios.

C. SWAT Model Development and Running

The hydrological model could be developed, calibrated and run through various approaches based on the gSWAT, gSWATSim services, and BASHYT platforms. The specialist could use the following solutions:

1. *gSWAT application* - The specialist develops the SWAT model by using ArcSWAT and ArcView tools on his desktop. By using the gSWAT application the user uploads the model onto the gSWAT server and performs interactively the calibration of the model [8]. The user controls the convergence to the optimal calibration (i.e. parameters, simulations, and iterations) by interactive techniques provided through the web graphical user interface (Figure 2). Finally the user may download the resulted calibrated model.
2. *gSWAT and BASHYT tools* - The applications “work together” by separate working sessions that

are connected just at the data level. The main advantage of this solution is the independency of the tools. The user performs the following steps: develop the SWAT model just in BASHYT, and then downloads the archived SWAT files and metadata. Now, follows the calibration by gSWAT as in the first solution. Finally the user uploads the results into BASHYT and visualizes the environmental information.

3. *gSWATSim services* - The applications work together through a common Storage Element and dedicated web Services (Figure 3). The working session is in BASHYT through which the user develops the model and defines the scenario. The user exports to gSWATSim the data onto the Storage Element, then through the dedicated web Service the execution environment is customized, and the scenario is executed. Scenario execution progress can be monitored directly in BASHYT. Finally after the execution, the results are available automatically into BASHYT for visualization. In this solution the user does not need to switch between the applications. BASHYT accesses a new functionality available through gSWATSim services, which allows the execution and the monitoring of running scenarios.

VI. SATELLITE IMAGE PROCESSING

Satellite images could reveal information on land cover, precipitations, geographic areas, pollution, and natural phenomena. Spatial and environment related data could be acquired by imagery classification that is actually a data mining throughout the multispectral bands. It is a multivariable process taking in account satellite image types (e.g. MODIS, Landsat), particular geographic area, soil composition, vegetation cover, and generally the

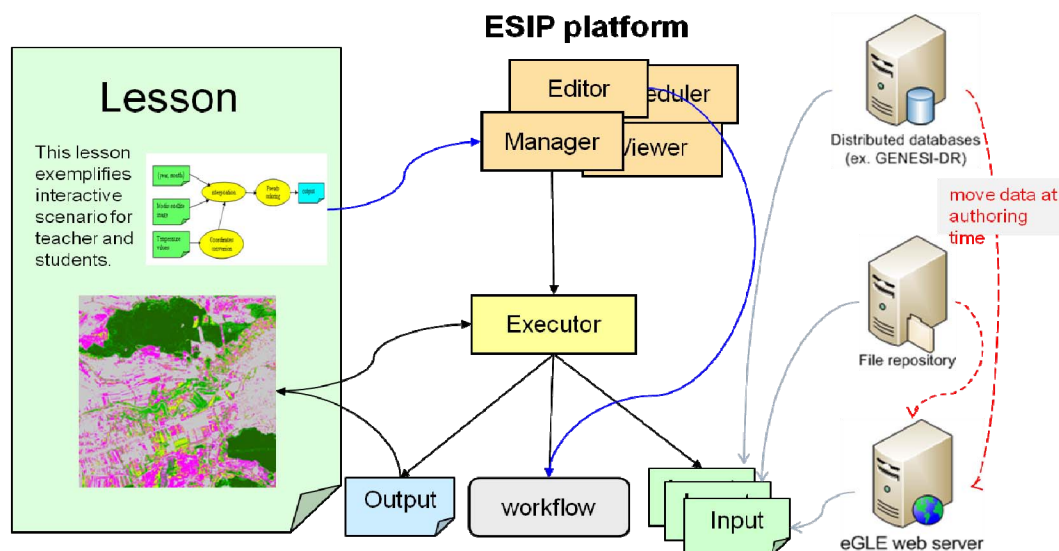


Figure 4. Grid based execution of the Earth Science oriented lesson.

context (e.g. clouds, snow, and season). All these specific and variable conditions require flexible tools and applications to support an optimal search for the appropriate solutions.

One of the platforms supporting the development of the Grid oriented applications for satellite image processing is ESIP (Environment oriented Satellite Data Processing Platform) [10]. ESIP supports a workflow based flexible description of the satellite images complex processing over the Grid. Actually ESIP includes as well the gridified GRASS functionality [11].

The ESIP based applications have been developed according with the methodology reported in [12]. The BSC-OS portal publishes the GreenLand web application concerning the satellite image processing, according the requirements defined for the Istanbul area in Turkey, Rioni River in Georgia, and generally the Black Sea area.

VII. DATA VISUALIZATION AND REPORTING

BASHYT (The Basin Scale Hydrological Tool) [13] is a web based interface to SWAT that works in tandem with ArcSWAT and AvSWAT [14]. It can be used to manage many watersheds/scenarios at once and exposes on the web a template to produce environment oriented applications. The applications can be edited directly through the browser and it is using its own authentication and authorization layer. BASHYT implements the Driving forces-Pressures-States-Impacts-Responses paradigm and is able to produce reports on environmental states that can be visualized in different ways.

In BASHYT the SWAT models are stored into a relational database. A preprocessing step is required to import raw data (vector, raster and tabular data) into the system. After importing SWAT models BASHYT could

offer tables, charts, and maps in a transparent way to the end users.

VIII. CITIZENS ORIENTED APPLICATIONS

Two demonstrator web applications for citizens will be developed within the enviroGRIDS project and available through the BSC-OS Portal. The first application, which is related to near real time dissemination of environmental data to citizens, a flood forecasting demonstrator is going to be applied on the Somes Mare catchment in northern Romania. For the second application, related to long term planning in river basins a demonstrator for long-term planning of remediation strategies regarding flooding, sediment and ecosystem problems along the Danube River section between the towns of Braila and Tulcea has been selected.

The first application is supported by a HEC-HMS [15] hydrological model executed over the Grid infrastructure, and the second one is supported by a SOBEK 1D/2D [16] hydrodynamic model of flow and sediment transport.

Actually the necessary data is already available and the model development is ongoing for both case studies. Geospatial data are available through the enviroGRIDS URM Portal by standard OGC services, while for water-related time series data the emerging WaterML standard will be used.

On the client side, for both applications the main interfaces are map-based (e.g. OpenLayers, Google maps, and Google Earth platforms), over which the additional data are overlaid as spatially distributed data, or point data containing time series of modeled results.

IX. TRAINING MATERIALS

The BSC-OS Portal provides the virtual training center based on eGLE (GiSHEO eLearning Environment),

developed initially through the GiSHEO project [17]. The training system has as generic users the teacher and the student. The teacher is the Earth Science specialist who authors teaching materials and coordinates the training sessions. The student is the trainee who accesses the teaching objects organized by lessons in order to get presentations, experiment algorithms on spatial data, process satellite images, execute environmental scenarios, and visualize reports already prepared by the specialists.

The teacher may use the Grid based execution to process satellite images (Figure 4), to execute specific algorithms through workflow descriptions or to visualize previously created teaching resources (i.e. already processed satellite images, geographical maps, diagrams, algorithm workflow descriptions, etc.). The students have only the ability to execute the lessons according to the constraints established by the teacher. Depending on the interaction specified level, they could as well be allowed to describe and experiment new workflows (i.e. algorithms, scenarios) or choose different input data (e.g. satellite images, parameters) for existing ones.

X. CONCLUSIONS

The development of the BSC-OS portal and generally the research through the enviroGRIDS project have revealed a lot of challenges regarding the gathering data into a dedicated SDI, interoperability between geospatial and Grid infrastructures, and interoperability between platforms developed by different partners (e.g. URM, gSWAT, ESIP, gProcess, gLite, BASHYT, and eGLE), huge spatial data sets involved in the development of hydrological models and environmental scenarios (e.g. Danube, Black Sea Catchment, Istanbul, and Rioni River in Georgia), security and access management in different platforms, application development in distributed and heterogeneous systems, etc.

Another issue the portal development has to face is the compatibility with new technologies and functional requirements. One main concern is the compatibility with the new European Middleware Initiative (EMI) that aims to improve and standardize the dominant existing middlewares in order to produce one simplified and interoperable middleware [18]. EMI attempts to unify a few Grid platforms such as ARC, gLite, Unicore and dCache. The EMI and Globus platforms will empower the EGI (European Grid Infrastructure) with more stable, useable and manageable software.

We have to develop extended and high resolution models and scenarios, to improve the tool and application functionality, and to improve the user interaction techniques with spatial data models.

The service oriented architecture, multicore, GPGPU based systems, Cloud processing are other technologies that we intend to explore in order to extend the functionality, application usability, system efficiency, and to improve the performance of data processing.

ACKNOWLEDGMENT

This research is supported by the FP7 enviroGRIDS Project funded by the European Commission, through the Contract 226740.

REFERENCES

- [1] enviroGRIDS Project, <http://www.envirogrids.net/>
- [2] K.C. Abbaspour, J. Yang, I. Maximov, R. Siber, K. Bogner, J. Mieleitner, J. Zobrist, R. Srinivasan, "Spatially-distributed modelling of hydrology and water quality in the pre-alpine/alpine Thur watershed using SWAT," *Journal of Hydrology*, 2007, 333:413-430.
- [3] *GEO secretariat (2007a) The Full Picture*. GEO, Geneva, Switzerland. http://www.earthobservations.org/documents/fullpicture/fp_ps_p136_we_cl.pdf
- [4] *European Commission, Research and Innovation - Environment*, <http://ec.europa.eu/research/environment/>
- [5] J.G. Arnold, R. Srinivasan, R.S. Muttiah, and J. R. Williams. 1998. Large Area Hydrologic Modeling and Assessment: Part I. Model Development. *Journal of the American Water Resources Association* 34(1), pp.73-89.
- [6] K. Charvat, S. Kafka, M. Splichal, M. Alberts, "URM for agriculture, environmental education, and knowledge sharing," *WCCA 2008 - 6th World Congress on Computers in Agriculture*, Tokyo, Japan, 24 - 27 August, 2008, pp. 455-460.
- [7] Soil and Water Assessment Tool, *Official SWAT web site*, <http://www.brc.tamus.edu/swat/index.html>
- [8] V. Bacu, D. Mihon, D. Rodila, T. Stefanut, D. Gorgan, "Grid based architectural components for SWAT model calibration," *HPCS 2011 - International Conference on High Performance Computing and Simulation*, 4-8 July, Istanbul, Turkey, pp. 193-198.
- [9] D. Rodila, V. Bacu, D. Gorgan, "Comparative analysis of multicore and Grid based execution of SWAT model," *3PGCIC 2011 - Sixth International Conference on P2P, Parallel, Grid, Cloud and Internet Computing*, October 26-28, 2011 (accepted paper).
- [10] D. Gorgan, V. Bacu, D. Rodila, F. Pop, D. Petcu, "Experiments on ESIP - environment oriented satellite data processing platform," in *Earth Science Informatics Journal*, Springer, Vol.3/4, 2010, pp. 297-308.
- [11] GRASS - *Geographic Resources Analysis Support System* project (2011), <http://grass.osgeo.org/>
- [12] D. Gorgan, T. Stefanut, V. Bacu, D. Mihon, D. Rodila, "Grid based environment application development methodology," in *Large-Scale Scientific Computing, SpringerJournal LNCS 5910*, pp.499-506
- [13] S. Manca, C. Soru, P. Cau, G. Meloni, M. Fiori, "A multi model and multiscale, GIS oriented web framework based on the SWAT model to face issues of water and soil resource vulnerability," *5th International SWAT 2009 Conference*, August 3-7. Available at <http://www.brc.tamus.edu/swat/4thswatconf/docs/rooma/session5/Cau-Bashyt.pdf>
- [14] Di Luzio, M., Srinivasan R., Arnold J.G. 2004. "A GIS-coupled hydrological model system for the watershed assessment of agricultural nonpoint and point sources of pollution," *Transactions in GIS*, 8(1): 113-136.
- [15] The Hydrologic Modeling System (HEC-HMS), <http://www.hec.usace.army.mil/software/heh-hms/>
- [16] SOBEK Suite, Delaters Systems, <http://www.deltaressystems.com/hydro/product/108282/sobek-suite>
- [17] D. Gorgan, T. Stefanut, V. Bacu, "Grid based training environment for earth observation," Published by Springer-Verlag in *LNCS 5529*, pp. 98-109.
- [18] EMI - European Middleware Initiative (2011), <http://www.eu-emi.eu/>