



Chapitre de livre

2011

Published version

Open Access

This is the published version of the publication, made available in accordance with the publisher's policy.

Distributed Geocomputation for Modeling the Hydrology of the Black Sea Watershed

Ray, Nicolas; Giuliani, Gregory; Gorgan, Dorian; Lehmann, Anthony

How to cite

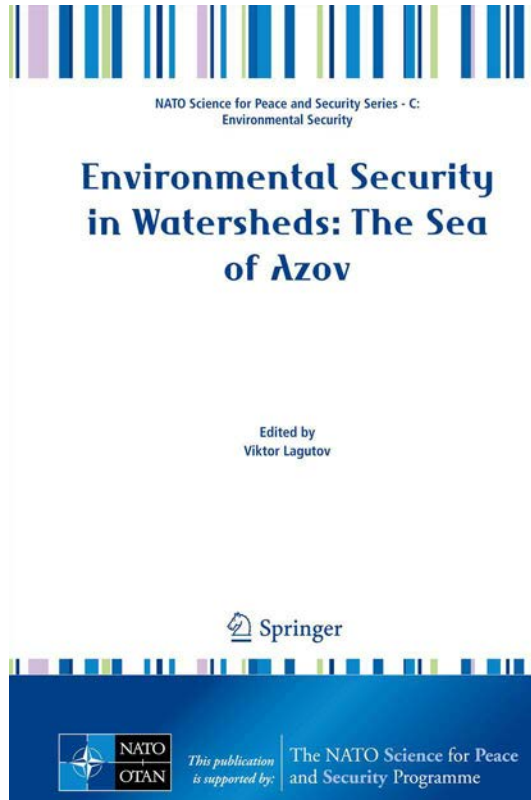
RAY, Nicolas et al. Distributed Geocomputation for Modeling the Hydrology of the Black Sea Watershed. In: Watershed Approach to Environmental Security: Fostering integrated water management in the Azov Sea Basin. Lagutov, V (Ed.). Dordrecht : NATO Science for Peace and Security Series: Environmental Security. Springer, 2011. p. 141–158. doi: 10.1007/978-94-007-2460-0_8

This publication URL: <https://archive-ouverte.unige.ch/unige:18368>

Publication DOI: [10.1007/978-94-007-2460-0_8](https://doi.org/10.1007/978-94-007-2460-0_8)

Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.

ISBN 978-94-007-245 9-4



This chapter was published in the above Springer book. The attached copy is furnished to the author for non-commercial research and education use, including for instruction at the author's institution, sharing with colleagues and providing to institution administration.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the chapter (e.g. in Word or TEX form) to their personal website or institutional repository.

Chapter 8

Distributed Geocomputation for Modeling the Hydrology of the Black Sea Watershed

Nicolas Ray, Gregory Giuliani, Dorian Gorgan, and Anthony Lehmann

Abstract The surface of the Black Sea watershed amounts to about 2 mil. km² with a population of 160 mil. Inhabitants over 25 countries. In light of the current and forthcoming climate, land cover and population changes in this region, it is becoming extremely important to better understand how the quantity and quality of waters will vary in the catchment over the coming decades. To model the hydrology of this catchment, three steps are needed: (1) a large transnational data collection effort, (2) adequate management and sharing processes of the environmental data in a dedicated Spatial Data Infrastructure, and (3) distributed computing in order to allow running a high-resolution model. The EU/FP7 enviroGRIDS project (running 2009–2013) is addressing these steps with a 30-partner consortium mainly located in the Black Sea region. In this paper we are discussing how enviroGRIDS is approaching the various data-related challenges of the project. We particularly address the important issue of sharing data through international initiative such as GEOSS, the specificity of the hydrological modeling tool SWAT (Soil and Water Assessment Tool), and the technical requirement for using Grid computing infrastructures to optimize computationally-intensive simulations.

Keywords Black Sea • Hydrological modeling • SWAT • Web Processing Service • Grid computing • Geospatial data • Spatial Data Infrastructure

N. Ray (✉) • G. Giuliani • A. Lehmann

Institute of Environmental Sciences, Climatic Change and Climate Impacts, enviroSPACE Lab.,
University of Geneva, Battelle – Building D, 7 route de Drize, CH-1227 Carouge, Switzerland

Division of Early Warning and Assessment, Global Resource Information Database – Europe,
United Nations Environment Programme, Châtelaine, Switzerland

University of Geneva, Forel Institute, Versoix, Switzerland
e-mail: nicolas.ray@unige.ch

D. Gorgan

Technical University of Cluj-Napoca, Cluj-Napoca, Romania

8.1 Introduction

Climatic change is becoming a worldwide concern that will affect many areas of human activities. The last report of the Intergovernmental Panel on Climate Change [29, 30] predicts important changes in the coming decades that will not only modify climate patterns in terms of temperature and rainfall, but will also drastically change freshwater resources qualitatively and quantitatively, leading to more floods or droughts in different regions, lower drinking water quality, increased risk of waterborne diseases, and irrigation problems. These changes may trigger socio-economic crises across the globe that need to be addressed well in advance of their occurrences in order to reduce their associated risks.

One region that could particularly be affected by these water-related problems is the watershed of the Black Sea (or Black Sea hydrological Catchment – BSC) (see Fig. 8.1). This catchment has a surface amounting to about 2 mil. km² with a population of 160 mil. Inhabitants over 25 countries. This large watershed is subject to numerous environmental pressures and threats. Inadequate management of wastewater/solid waste, ecological unsustainable industrial activities, inadequate land management, and improper agricultural practices have greatly affected the region in many places [51], and notably the Danube catchment area [44]. These pressures generate several direct consequences such as pollution of surface/groundwater, eutrophication, and accelerated runoff/erosion. These consequences have, in turn, the following main effects: decline in quality of life, human health risks, degradation of



Fig. 8.1 The Black Sea watershed

biodiversity, economic decline, and reduced availability of water. Even if some signs of recovery have been observed in the last years, the Black Sea itself is also affected by severe environmental degradation [8] and eutrophication remains a severe problem.

The European Community is addressing the crucial problem of water quality and quantity by adopting the Water Framework Directive (WFD) [9, 31]. A key aspect of the WFD is the consideration of the river basin as the working unit, moving away from administrative boundaries such as communes, provinces, districts or countries that often cross across water-related boundaries. This requires the installation of a regulating body for the whole river basin. For many transboundary rivers, these bodies unite representatives of different countries. One extreme case within the Black Sea watershed is the river Danube, which is now regulated by the ICPDR (International Commission for the Protection of the Danube River). The ICPDR is driven by the interests of 19 riparian countries and of the European Union (EU).

Despite efforts to date, the vulnerability of different areas of Europe and beyond to climate change remains poorly addressed. Moreover, there is a strong need to integrate information on land cover and demographic changes in order to better assess population vulnerability to water scarcity. Building spatially-explicit integrated scenarios of climate, land cover and demographic changes for the entire Black Sea catchment is therefore a necessity if one wants to fully understand the future trends in water quantity and quality in this region. However, before exploring the impacts of these scenarios, one needs a good quality spatially-explicit hydrological calibrated with appropriate and sufficient input data.

The aim of this paper is to discuss the enviroGRIDS approach to modeling the hydrology of the entire Black Sea catchment, and especially the two important technical and institutional issues that arose from such an endeavor: (1) the transnational environmental data collection, standardization, and dissemination process, and (2) the need for distributed computing in order to achieve high-resolution hydrological modeling.

8.2 The EnviroGRIDS Project and the Modeling of the Black Sea Catchment

The EnviroGRIDS project (<http://www.envirogrids.net>), funded by the EU 7th Framework Programme for a period of 4 years (April 2009–March 2013) with a consortium of 30 partners, aims at building capacities in the Black Sea region on new international standard 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. The project focuses on the terrestrial part of the catchment and one of its main scientific objectives is to assess how the sustainability of water usage in this catchment may evolve in the future (at 20, 30 and 50 years time horizons). To achieve this, the project aims at building the first full hydrological model for the Black Sea catchment that will allow one to explore the

outcomes of integrated scenarios of change (climate, land cover and demography) on the water quality and quantity of all rivers basins in a comparative way.

To model the hydrology of the catchment, the Soil Water Assessment Tool (SWAT: <http://swatmodel.tamu.edu>) [3] model is used. SWAT is a widely used basin-scale, continuous-time model that integrates various processes such as hydrology, climate, chemical transport, soil erosion, pesticide dynamics and agricultural management. SWAT accounts for variable soil and land cover conditions by subdividing the simulated catchment into sub-areas. The model uses a daily to sub-hourly time step and can perform continuous simulation for a 1–100 year period. SWAT has an ArcGIS (ESRI, Redlands) interface that takes layers of information such as soil, land cover, elevation, and calculates hydrology, erosion and chemical transport both inland and in-stream. About 50 peer-reviewed papers discussed the application of SWAT on pollution loss studies for a wide range of small and large river catchments [17].

SWAT was already used to simulate the hydrology of large-scale body masses such as the African continent [46], the entire U.S. with river discharges data at around 6,000 gauging stations [4], and of 12 large river catchments in India [28]. SWAT is recognized by the U.S. Environmental Protection Agency (EPA) and has been incorporated into the EPA's BASINS (Better Assessment Science Integrating Point and Non-point Sources).

In the enviroGRIDS project, SWAT will be used to apply a high-resolution (i.e., sub-catchment spatial and daily temporal resolution) water balance model to the entire BSC. The BSC model will be calibrated and validated using river discharge data, river water quality data, and crop yield data [1]. As part of the modeling work, uncertainty analysis will also be performed to gauge the confidence on all model outputs. Subsequent analyses of land use change, agricultural management change, and/or climate change can then predict the consequence of various scenarios.

To achieve the building and calibration of the BSC SWAT model, an initial data collection phase (ended in mid-2010) over the catchment was necessary. This data collection phase was a keystone process to ensure the best possible informed model and to discover data gaps. The main outcome of this endeavor was that 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 or to their sensitive nature (as perceived by the national agency owning the data). As a result of this data collection phase, the available data from the enviroGRIDS Consortium were gathered to construct and calibrate a coarse-resolution SWAT model for the full catchment. Recent results obtained with this model (Abbaspour 2010, personal communication) 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. How to best access and make available these data sets to the largest audience possible through international standards is the subject of the following section.

8.3 Sharing Environmental Data in a Transnational Setting

Environmental managers are regularly facing the problem of having to take sound decisions with only partial information, which can translate into inaccurate and inefficient management decisions. Gathering and integrating the vast amount of environmental data generated on a daily basis, but often operated in isolation, therefore appears as an essential and fundamental effort to be taken in order to make sound decisions at all levels, from global to local [40].

Understanding and modeling a hydrological system such as the Black Sea watershed is very complex due to the highly interconnected and continuously evolving interactions at many spatial and temporal scales. These interactions require gathering and integrating different sets of environmental data (e.g., physical, chemical, biological) (GEO [19]). Currently, data accessibility, availability, compatibility, and lack of sufficient resources to analyze these data are among the most frequent difficulties that are negatively influencing the way that scientists, researchers, decision-makers and the general public are accessing and using these data [7, 48]. This is mainly due to the fact that geospatial data are voluminous, geographically distributed, and heterogeneous in term of format. All these factors influence the way that data providers store, publish and deliver environmental data. Moreover, users are often lacking the appropriate computational resources to analyze these data. Current environmental research projects regularly need to handle several terabytes of data and accessing high-performance hardware and specialized software is expensive. This explains why currently data sources are often fragmented, integrating geospatial data to answer a scientific problem is difficult and expensive, and diffusion of geospatial information is problematic and not applied efficiently.

Thus, making sense of the vast amount of data and information, and turning them into understandable information is a challenging, but necessary task [26]. Enhancing access to data benefits the wide usage of it and enables scientists to compare results and methods more easily, which improves scientific accountability, credibility and potentially the quality of data. Having environmental data in digital form is essential as it greatly facilitates storage, dissemination, data exchange and sharing, while allowing for faster and easier updates, and giving the users the ability to integrate data from multiple sources. Consequently, digital geospatial data can be thought of as a shared resource that can be maintained continuously [40].

To address the need of sharing environmental data, the concept of Spatial Data Infrastructure (SDI) appears to be an interesting framework [38, 45]. An SDI encompasses data sources, systems, network linkages, standards and institutional issues in delivering geospatial data and information from many different sources to the widest possible group of potential users [11]. SDIs intend to avoid duplication of efforts and expenses by enabling users to save resources and time when trying to acquire or maintain data sets [35, 45]. SDIs can be seen as an integrated information highway which links together environmental, socio-economic and institutional geospatial data resources to provide a movement of data from local to national and global levels [36, 37].

One essential component of SDIs is interoperability: an open science framework allowing scientists and researchers to publish, discover, evaluate and access data [42]. The Open Geospatial Consortium (OGC) aims to develop and provide such standards enabling communication and exchange of information between different systems of different types operated with different software [43]. A SDI committed to interoperability permits widely and effectively exchange of data, maximizes the value and reuse of data and information under its control, and is able to exchange these data and information with other interoperable systems, which allows new knowledge to emerge from relationships that were not envisioned previously.

Different initiatives at the regional and global levels are influencing and promoting the creation of SDIs. These initiatives coordinate actions that promote awareness and implementation of complementary policies, common standards and effective mechanisms for the development and availability of interoperable geospatial data and technologies to support decision making at all scales and for multiple purposes. Such an initiative is the Global Earth Observation System of Systems (GEOSS), a worldwide voluntary effort coordinated by the Group on Earth Observation (GEO) aiming at connecting already existing SDIs and Earth Observation infrastructures. GEOSS is foreseen to act as a gateway between producers of geospatial data and end users, with the aim of enhancing the relevance of Earth observations for the global issues and offering public access to comprehensive information and analyses on the environment (GEO [18, 20, 21, 24]). To support the nine defined Societal Benefit Areas (SBAs, see Fig. 8.2), various mechanisms for data sharing and dissemination are presented in a 10-year Implementation Plan Reference Document (GEO [18]). Any GEO member is volunteer, must endorse data sharing principles, and seeks to agree on “interoperability arrangements” (GEO [20, 21]) to allow its SDIs to communicate with others.

The establishment and implementation of initiatives such as GEOSS follows a more generic approach, the so-called System of Systems (SoS) approach [39]. This approach underpins a multi-disciplinary framework built on existing systems. It allows recognizing the heterogeneity of systems reflecting the diversity of stakeholders involved, while recognizing and specifying arrangements in order to federate these systems that can be very different. Such framework provides interesting features: (1) each component can operate independently (e.g., in order to match their own objective) and can be connected to others component by agreeing and specifying interoperability arrangements, providing flexibility (the overall framework will not fail done if one or more components disappear), (2) it increases the capacity to turn data into information by sharing resources, (3) it provides a holistic approach, (4) it supplements but not supplant existing systems, (5) it is based on Service Oriented Architecture principles, and (6) it can incrementally incorporate new components/systems. However, one of the major benefits of the SoS approach is to allow users to perform functions that cannot be made with any single component [6]. This means that such a system is more than the sum of its parts and it offers the possibility to better understand the complex relationships between the different components of the Earth system. Consequently, such a framework can then offer possibilities for SDIs to extend complement and benefit from capabilities offered by

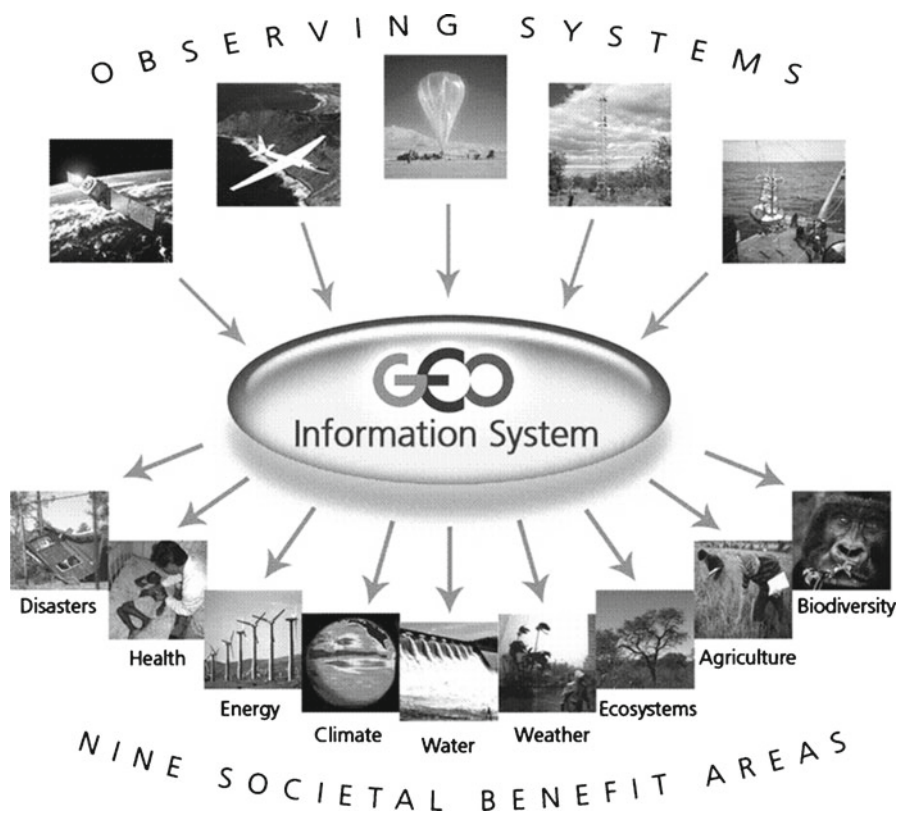


Fig. 8.2 The global earth observation system of systems

other type of infrastructures. For example, distributed computing infrastructures such as Grid architecture can be really beneficial for data processing and management of an ever-increasing amount of high-resolution data.

One of the goals of the enviroGRIDS project is to push for wide adoption of GEO data sharing principle in the Black Sea region, and therefore to register in GEOSS many new environmental data sets on the Black Sea watershed. During the first year of the project, an analysis of data sets and observation systems available within the Black Sea catchment against the enviroGRIDS project requirements revealed spatial and temporal gaps in data coverage, gaps in observation systems, and problems with data accessibility, compatibility and interoperability. It highlighted that large amount of data sets relevant to the project and end-users data needs are available at different scales, from national to regional, European and global. It was also found that access to data is often limited or restricted, particularly at national level, so data accessibility appears to be the main problem preventing effective data usage. With respect to hydrology, major identified problems concerned (1) data gaps on pollution loads to the Black Sea from land based sources,

including rivers, (2) missing observation system to monitor pollutants deposition from atmosphere, (3) gaps on availability of data on pollutants deposition from atmosphere, (4) scarce data from river stations in some regions, (5) limited or restricted access to pollution discharges and water quality data, particularly in non-EU countries, and (6) unsatisfactory spatial resolution of crop yield data from available global data sets, while access to more detailed national agriculture data is limited or restricted. Additionally, the analysis of available data sets revealed the problem of data compatibility at different scales (e.g., European and global land cover classifications) requiring significant efforts, particularly when it is necessary to combine in one application data from different scales. Finally, the fact that most of reported national data sets are not accessible through the Internet and do not have relevant metadata available, points to the necessity of bringing project partners' data and metadata to interoperability standards (e.g., OGC and ISO). This required developing a capacity building strategy on Earth Observation Systems in the Black Sea catchment through improved data collection, management, storage, analyses and dissemination.

For the large adoption, acceptance and commitment to SDI philosophy, enviroGRIDS is therefore seeking to build capacities at human (e.g., education and training of individuals), infrastructure (e.g., installing, configuring, and managing the needed technology) and institutional (e.g., enhancing the understanding within organization and governments of the value of geospatial data to support decision-making) levels. Of particular importance is to show and prove the benefits of sharing interoperable data/metadata through appropriate examples, best practices and guidelines. This will help to strengthen (1) existing observation systems, (2) capacities of decision-makers to use it, and (3) capacities of the general public to understand important environmental, social and economical issues at stake in the region. Additionally, capacity building efforts should aim to convince a maximum of data owners/providers that sharing their data is very good opportunity to become more visible nationally and internationally by joining the effort of GEOSS.

8.4 Distributed Modeling of the Hydrology of the Black Sea Watershed

The push for large-scale high-resolution SWAT modeling, as is foreseen in the enviroGRIDS project, comes at the expense of very large computational needs. This expense proves even greater if there is a need to iteratively run such large models for common practices in modeling such as calibration and uncertainty analysis. However, the very nature of a SWAT hydrological model makes it theoretically possible to split the full computation into sub-units of computation that can be run independently on many computers. A distributed computing architecture can provide the necessary computational and storage resources to achieve the parallel computation of these sub-units.

8.4.1 *Grid Based Computing Infrastructures*

The Grid architecture is one of the main solutions that provide many potential benefits for the modeler:

- it reduces total computation time to get results more rapidly;
- it enables running models for larger geographical area;
- it enables the development and processing of higher-resolution models;
- it allows better assessing model uncertainty and sensitivity by running a larger number of model iterations;
- it supports user scalability by simultaneously running a large number of models.

Altogether, these benefits can participate to a much higher accuracy in model outputs, and can therefore increase the quality of watershed management decisions based on these model outputs.

The term grid computing originated in the 1990s as a metaphor for making computer power as easy to access as an electric power grid [16]. A grid can be defined as a layer of networked services that allows users single sign-on access to distributed collection of resources not centrally controlled. Another, task-oriented, definition of a grid is a cluster of loosely coupled, networked computers acting in concert to perform very large tasks. There are basic concepts lying behind grid: long term collaboration, user and provider communities and security. To use a grid infrastructure, users belonging to different administrative organizations are typically grouped into a specific user community, called a Virtual Organization (VO), a group of people who share a data-intensive goal. This group of users 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). The acceptance in a VO authorizes users to access the resources based on the policies of the VO. Moreover, an application that is intended to be run on the grid must go through a so-called “gridification” process. This process intends to generate a grid application that interacts with grid services to achieve requirements that are specific to a particular VO or user.

Many different Grid infrastructures exist and are available for scientists worldwide. These grids can be classified in international, national, or field-specific sets (see <http://www.gridcafe.org/grid-powered-project.html>). In the enviroGRIDS project, we are using the Worldwide LHC Computing Grid [50]. The WLCG was implemented within the framework of The Enabling Grids for E-science (EGEE) series of projects funded by the European Commission and that started in March 2004 and officially ended on April 2010 [13]. The coordination of the Grid infrastructure is now taken over by the European Grid Initiative (see [14]), which is the future sustainable computing Grid infrastructure in Europe. The goal of the establishment of EGI is to move from a project-based funding of the European grid infrastructure (as was the case in the EGEE project) to a sustainable country-based funding. Currently, the majority of users of the WLCG come from the High Energy Physics (HEP) community for which the WLCG is a necessity to analyse the data generated by the Large Hadron Collider

(LHC) experiment at the European Organization for Nuclear Research (CERN) in Geneva, Switzerland. However, a growing number of WLCG users come from many other scientific disciplines, such as the biomedical field, earth science, astrophysics, fusion science, etc. In late 2010, the WLCG had about 260 resource centres in 55 countries, with more than 150,000 CPU cores, 28 PB of disk storage, 38 PB of tape storage, more than 14,000 registered users, and more than 300,000 jobs/day. These figures are likely to increase considerably over the coming months/years.

8.4.2 *EnviroGRIDS Computing Infrastructure*

The time dedicated to the gridification of particular software is only one of the several aspects to take into consideration before deciding to engage project resources into Grid computing. In enviroGRIDS, an important issue was to ensure the sustainability of the pool of computing resources that enviroGRIDS partners could use during, but also beyond, the project duration. Two mitigation actions were taken to minimize the risk of losing access to computational resources. First, we decided to build a dedicated enviroGRIDS VO. This VO is composed of computational nodes belonging to enviroGRIDS partners either already part of the WLCG or willing to join after appropriate software installation. The current enviroGRIDS VO has started the first experiments on developing tools and applications on resources provided by the Technical University of Cluj-Napoca (512 core processors and 12 TB storage), but additional resources provided by enviroGRIDS partners, universities and research institutes (UPB, ICI, UVT from Romania) should join soon. An advantage of the enviroGRIDS VO is that it may facilitate incorporation of new computing resources in the EGEE from the Black Sea countries. Some institutions in this region may indeed be more willing to share their resources for their use in a dedicated Black Sea project, rather than joining a more generic VO (e.g., Earth Science Research VO). Finally, administrating its own VO permits more flexibility in term of software installation and resource allocations within the VO.

The second mitigation action was to ensure not to be restricted to a given Grid platform or Grid middleware (i.e., the software running Grid services). The new European Middleware Initiative (EMI) aims to improve and standardize the dominant existing various middleware in order to produce one simplified and interoperable middleware [15]. EMI attempts to unify a few Grid platforms such as ARC [2], gLite [33], Unicore [47] and dCache [12]. The EMI platform will empower the EGI infrastructure with more stable, useable and manageable software.

8.4.3 *SWAT Model Parallelization for Distributed Processing*

Although a recent survey [32] indicates that Grid technology in hydrology has been successfully tested to improve flood prediction and ground-water resources

management, only minimal efforts on reduction of computation time have been made in the past for SWAT modeling [49]. However, the intrinsic model representation in SWAT makes it theoretically well suited for gridification. A watershed modeled using SWAT is indeed partitioned into different required and/or optional objects of subunits such as sub-basins, reaches/main channel segments, impoundments/reservoirs on the main channel network and point sources. Watershed sub-basins are the first level of the subdivision. These sub-basins are defined by geographical positions in the watershed and are spatially related to one another [41]. All sub-basins drain into the river network where water is routed from upstream to downstream reaches. The land area in a sub-basin may be divided into hydrologic response units (HRUs) that are portions of a sub-basin that possess unique land use, management, or soil attributes [41]. Unlike in the case of sub-basins, no spatial relationship or interaction can be specified among HRUs. Sediment, chemicals or nutrient loadings from each HRU are computed independently and then summed up to determine the total load from a sub-basin. A watershed model should also incorporate one reach or main channel associated with each sub-basin. This channel carries loadings from the sub-basin or outflow from the upstream reach segments into the downstream network of the watershed in the associated reach segment.

By simulating each sub-basin independently on a separate Grid node and ensuring proper hydrological network routing at the end, it is therefore possible to decrease the total simulation time. We recently explored various ways of achieving this with SWAT running on the WLCG infrastructure [52]. Our results showed a clear potential for using SWAT on the Grid, but only with large models. For smaller models, the various overheads of running on the Grid (e.g., splitting of the model, submission time, queuing in remote nodes, merging of results) are costly in term of total computational time, and in such case running the model locally is more efficient.

In parallel, we also developed gSWAT [5], a web-based application that will be accessible through the project main portal (see below). This application allows the calibration of SWAT models and the executions of different scenarios. The development of such a web interface is extremely important if one seeks wide adoption of Grid technology by the SWAT and other user communities. User-friendly interfaces to the Grid must indeed be developed to mask the underlying complexity of the Grid architecture.

8.4.4 BSC-OS Portal

One challenge of the enviroGRIDS project is the interoperability between geospatial and Grid infrastructures. The geospatial technologies offer very specialized functionality for Earth Science oriented applications, while the Grid oriented technology is able to support distributed and parallel processing. The enviroGRIDS system resources are accessible to the large community of users through the BSC-OS (Back Sea Catchment Observation System) Portal that provides Web applications for data management, hydrological models calibration and execution,

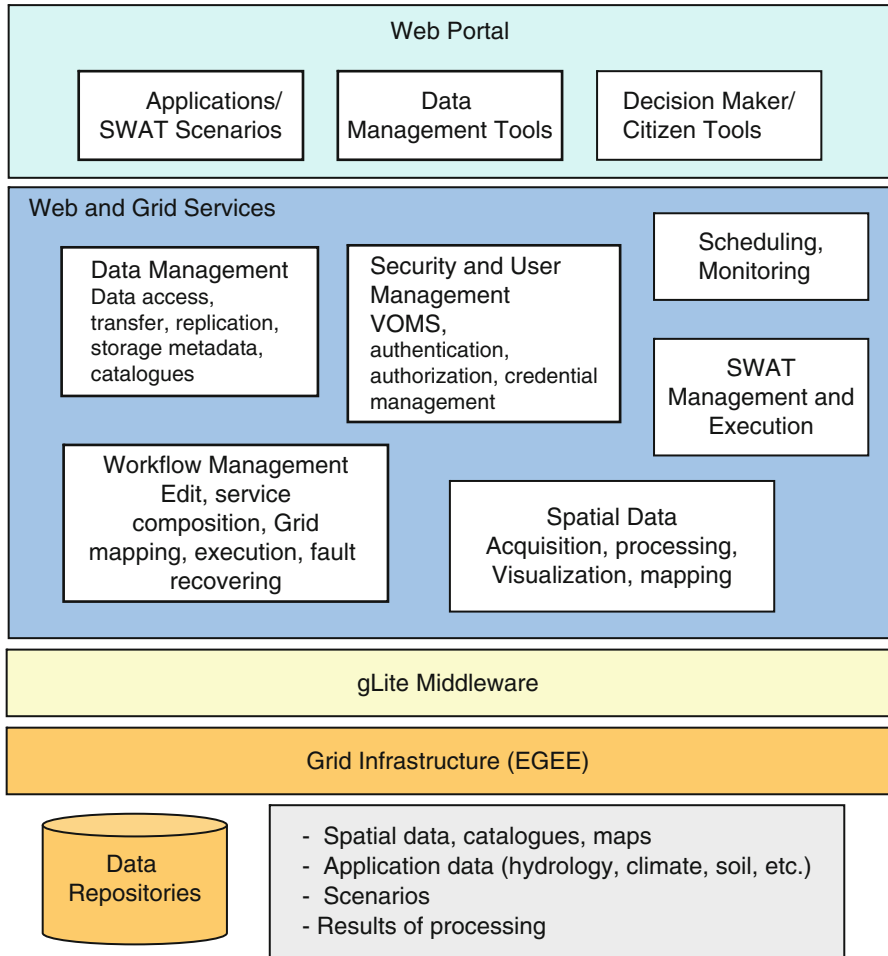


Fig. 8.3 EnviroGRIDS functional Layers. The lower level is the data level. The Grid system is provided by the EGEE infrastructure, on which the gLite middleware is running. The middle levels consist of a set of various services and platforms supporting the basic functionality. The upper level provides tools and applications to end users with appropriate graphical user interfaces

satellite image processing, report generation and visualization, and virtual training center (see Fig. 8.3).

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

The main user 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;
- Hydrological model management – provides the Earth Science specialists with hydrologic model configuration, scenario and model development, model calibration and scenario running;
- 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 analyses and predictions;
- Virtual Training Center – supports the development of Earth Science oriented training materials based on the Grid processing.

The BSC-OS Portal provides a set of tools and applications to the users interested in environmental studies and predictions. The regular users visualize the reports generated by the specialists as results of executing environmental scenarios. The input materials 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. The portal gathers services provided by various technologies such as gSWAT [5], Collaborative Working Environment (CWE) [34], Uniform Resource Management (URM) [10], gProcess and ESIP Platforms [27].

Finally, Grid technology can promote 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 [32]. In the enviroGRIDS project, the use of the 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 is enabling regional organizations to take advantage of EnviroGRIDS to analyze large trans-boundary environmental data sets in a harmonized way to support the conceptualization and implementation of environmental and relevant sustainable development policies.

8.5 Some Remaining Challenges

Several challenges are foreseen in the second part of the enviroGRIDS project (ending in 2013). We have discussed in this paper the necessity and benefits of sharing environmental data through interoperable SDIs. We have also examined how Grid computing can help in tackling large spatially-explicit hydrological models such as

the Black Sea watershed. If we want the joint benefit of achieving optimized data management and massive data processing, we need to address the real challenge of integrating SDIs and Grid computing. As we discussed previously (see [23]), this integration means developing a Grid-enabled SDI (or gSDI), and it is not straightforward because Grid and SDI architectures differ in many ways. To take full advantage of a gSDI, research and development is necessary in order to: (1) develop so-called mediation mechanisms [25] to be able to launch simulations on different computing back ends (e.g., local computer, local or remote cluster, Grid infrastructure) through standardized procedures, (2) develop plugins for various GIS clients such as ArcGIS and GRASS to allowing users to seamlessly access different computational resources depending on the task (e.g., data retrieval, processing or map making), (3) evaluate the capabilities offered by Desktop Grids that provide access to unused resources of desktop computers (CPU and storage) within a local network, (4) evaluate the potential of Grids in term of data management because Grids offer capabilities that are currently not (or only partially) provide by SDIs (e.g., distribution storage, data replication, data stored as close as possible to components that access them, security, and efficient data moving protocols). This last point requires making Grid middleware spatially-enabled, and then implementing OGC standards and interfaces directly into them.

Another challenge stems from the world of social media/networks (e.g., Facebook, LinkedIn, Twitter, Flickr, YouTube) that allows users to share information and rapidly proliferate content, which is currently changing the field of Science (see [22] and references therein). Social media users can now easily share any report, image, map or any other geospatial information with their network. It can be expected that this kind of medium becomes an increasingly important way to communicate information on the environment, and potentially also to generate new data. The latter is exemplified by the geotagging capabilities of pictures published on Flickr (e.g., identifying pictures on a map) or mapping of tweets (i.e., Twitter messages) in order to spatially analyze this data and make sense of it (e.g., in a crisis situation). In the context of enviroGRIDS, we are exploring ways in which social media could help in building capacities and knowledge in the field of data sharing principles. These media can potentially foster communication, sharing and collaboration around emerging problems, help disseminate best practices, and influence the development of new methodologies and tools.

At the global and regional levels, initiatives such as GEO/GEOSS need the support and engagement of different stakeholders from the different communities involved. However, with the growing availability of server/client tools to set up a local SDIs, the risk is high to see a growing number of unconnected systems. Hence, promoting and raising awareness on GEO/GEOSS and their related benefits is essential in order to concentrate and coordinate efforts on joining the vision of standardized and interoperable SDIs. To achieve this objective, commitment, capacity building, interoperability, harmonization, and answering the needs and requirements of communities are essential elements to strengthen the engagement of all stakeholders.

Finally, environmental security linked to water quantity and quality in the Black Sea watershed is bound to the scientifically sound forecasting of its state under forthcoming climatic, land cover and populations changes. This forecasting needs a well-calibrated hydrological model, which can only be achieved through sufficient availability of data from river stations. Moreover, the Water Framework Directive needs a lot of observational data to assess the quality of the rivers based on water and biodiversity samples. Obtaining these data is still problematic in many of the non-EU countries on the Eastern side of the Black Sea watershed. Focusing our efforts in these countries, while understanding the institutional particularities of each of them, is a big challenge in the enviroGRIDS project. However, communicating the data sharing principles and benefits to these countries' stakeholders is important because the information shared by different water authorities across Europe should find its way to the Water Information System for Europe (WISE) and the Biodiversity Information System for Europe (BISE). This would be able to present the water-related environmental quality in all regions of Europe in a comparable way, which could greatly facilitate research and enhance the quality of environmental management. The unique natural features and identity of the Black Sea region well deserve a concerted effort and a state-of-the-art information system in order to preserve its value and assess its vulnerability to the global changes we are facing.

Acknowledgments We thank the European Commission "Seventh Framework Programme" for funding the enviroGRIDS project (Grant Agreement n 227640).

References

1. Abbaspour KC, 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
2. Advanced Resource Connector (ARC) (2011) <http://www.nordugrid.org/middleware>
3. Arnold J, Srinivasan R, Muttiah R, Williams J (1998) Large area hydrologic modeling and assessment – part 1: model development. *J Am Water Resour Assoc* 34(1):73–89
4. Arnold JG, Srinivasan R, Muttiah RS, Allen PM (1999) Continental scale simulation of the hydrologic balance. *J Am Water Resour Assoc* 35:1037–1051
5. Bacu V, Mihon D, Rodila D, Stefanut T, Gorgan D (2011) Grid based architectural components for SWAT model calibration. In: Proceedings of the 2011 international conference on High Performance Computing and Simulation (HPCS 2011), pp. 193–199, Istanbul, Turkey, July 4–8, 2011
6. Béjar R, Latre MA, Noguera-Iso J, Muro-Medrano PR, Zarazaga-Soria FJ (2009) Systems of systems as a conceptual framework for spatial data infrastructures. *Int J Spat Data Infrastruct Res* 4:201–217
7. Bernard L, Craglia M (2005) SDI: from spatial data infrastructure to service driven infrastructure. Workshop on cross-learning on Spatial Data Infrastructures (SDI) and Information Infrastructures (II), Twente
8. Borysova O, Kondakov A, Paleari S, Rautalahti-Miettinen E, Stolberg F, Daler D (2005) Eutrophication in the Black Sea region; Impact assessment and Causal chain analysis. University of Kalmar, Kalmar

9. CEC (2000) Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for the community action in the field of water policy. <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2000:327:0001:0072:EN:PDF>
10. Charvat K, Kafka S, Splichal M, Alberts M, Martinsone A, Horak P, Vlk M, Bruins P (2008) URM concept for sharing information inside of communities, WCCA 2008, Tokyo
11. Coleman DJ, McLaughlin JD, Nichols S (1997) Building a spatial data infrastructure. In: 64th permanent congress meeting of the Fédération Internationale des Géomètres (FIG), Singapore, p 11
12. dCache project (2011) <http://www.dcache.org/>
13. Enabling Grids for E-science project (EGEE) (2010) <http://www.eu-egee.org/>
14. European Grid Infrastructure (EGI) (2011) <http://www.egi.eu/>
15. European Middleware Initiative (EMI) (2011) <http://www.eu-emi.eu/>
16. Foster I, Kesselman C (1999) The grid: blueprint for a new computing infrastructure. In: Foster I, Kesselman C (eds) *Globus: a toolkit-based grid architecture*. Morgan Kaufmann, San Francisco, pp 259–278
17. Gassman PW, Reyes MR, Green CH, Arnold JG (2007) The soil and water assessment tool: historical development, applications, and future research directions. *Trans Am Soc Agric Biol Eng* 50(4):1211–1250
18. GEO Secretariat (2005) Global earth observation system of systems 10-year implementation plan reference document. GEO, Geneva. <http://www.earthobservations.org/documents/10-Year%20Plan%20Reference%20Document.pdf>
19. GEO Secretariat (2007) The full picture. GEO, Geneva. http://www.earthobservations.org/documents/fullpicture/fp_ps_p136_we_cl.pdf
20. GEO Secretariat (2007) Strategic guidance for current and potential contributors to GEOSS. GEO, Geneva. http://www.earthobservations.org/documents/portal/25_Strategic%20Guidance%20Document.pdf
21. GEO Secretariat (2007) Tactical guidance for current and potential contributors to GEOSS. GEO, Geneva. http://www.earthobservations.org/documents/portal/24_Tactical%20Guidance%20for%20current%20and%20potential%20contributors%20to%20GEOSS.pdf
22. Gewin V (2011) Social media: self-reflection, online. *Nature* 471(7340):667–669
23. Giuliani G, Ray N, Lehmann A (2011) Grid-enabled Spatial Data Infrastructure for environmental sciences: challenges and opportunities. *Future Gener Comput Syst* 27(3):292–303. doi:10.1016/j.future.2010.09.011
24. Giuliani G, Ray N, Schwarzer S, De Bono A, Peduzzi P, Dao H, Van Woerden J, Witt R, Beniston M, Lehmann A (2011) Sharing environmental data through GEOSS. *Int J Appl Geospat Res* 2(1):1–17. doi:10.4018/jagr.2011010101
25. Giuliani G, Nativi S, Lehmann A, Ray N (in review) WPS mediation: an approach to process geospatial data on different computing backends. *Comput Geosci*
26. Gore A (1998) The digital earth: understanding our planet in the 21st century. *Photogramm Eng Remote Sens* 65(5):528–530
27. Gorgan D, Bacu B, Rodila D, Pop P, Petcu D (2010) Experiments on ESIP – environment oriented satellite data processing platform. *Earth Sci Inform J* 3(4):297–308, ISSN: 1865–0473
28. Gosain AK, Rao S, Basuray D (2006) Climate change impact assessment on hydrology of Indian River basins. *Curr Sci* 90:346–353
29. IPCC (2007) *Climate change 2007: impacts, adaptation and vulnerability*. Contribution of Working Group II to the Fourth Assessment Report of the IPCC. Cambridge University Press, Cambridge
30. IPCC (2007) *Climate change 2007: mitigation of climate change*. Contribution of Working Group III to the Fourth Assessment Report of the IPCC. Cambridge University Press, Cambridge
31. Kaika M (2003) The water framework directive: a new directive for a changing social, political and economic European framework. *Eur Plan Stud* 11(3)
32. Lecca G, Petitdidier M, Hluchy L, Ivanovic M, Kussul N, Ray N, Thieron V (2011) Grid computing technology for hydrological applications. *J Hydrol* 403:186–199. doi:10.1016/j.jhydrol.2011.04.003

33. Lightweight Middleware for Grid Computing (gLite) (2011) <http://glite.cern.ch/>
34. Manca S, Soru C, Cau P, Meloni G, Fiori M (2009) A multi model and multiscale, GIS oriented Web framework based on the SWAT model to face issues of water and soil resource vulnerability. In: Presentation at the 5th International SWAT Conference, August 3–7, Boulder, Colorado, USA
35. Mansourian A, Rajabifard A, Valadan Zoj MJ, Williams D (2004) SDI for disaster management to support sustainable development. Paper presented at the Map Asia, Beijing
36. Masser I (2005) The future of spatial data infrastructures. ISPRS Workshop on service and application of spatial data infrastructure, Hangzhou
37. Masser I (2005) GIS worlds: creating spatial data infrastructures. ESRI Press, Redlands
38. Masser I (2007) Building European spatial data infrastructure. ESRI Press, Redlands
39. Nativi S, Bigagli L (2009) Discovery, mediation, and access services for earth observation data. *IEEE J Sel Top Appl Earth Obs Remote Sens* 2(4):233–240. doi:10.1109/jstars.2009.2028584
40. Nebert DD (2005) Developing spatial data infrastructure: the SDI cookbook. GSDI Association, Chapel Hill. http://memberservices.gsd.org/files/?artifact_id=655
41. Neitsch SL, Arnold JG, Kiniry J, Williams JR (2005) Soil and water assessment tool theoretical documentation, version 2005. Grassland, Soil and Water Research Laboratory, Agricultural Research Service and Blackland Research Center, Texas Agricultural Research Station, Temple, Texas
42. Open Geospatial Consortium (2004) The havoc of non-interoperability. OGC, White Paper, http://portal.opengeospatial.org/files/?artifact_id=5097
43. Open Geospatial Consortium (2008) OGC reference model. OGC http://portal.opengeospatial.org/files/?artifact_id=31112
44. PCU (1999) Strategic action plan for the Danube River basin 1995–2005 – Revision 1999. Danube Pollution Reduction Programme, Programme Co-ordination Unit, Vienna
45. Rajabifard A, Williamson IP (2001) Spatial data infrastructures: concept, SDI hierarchy and future directions. In: *Geomatics'80*, Tehran, Iran
46. Schuol J, Abbaspour KC, Yang H, Srinivasan R, Zehnder AJB (2008) Modelling blue and green water availability in Africa. *Water Resour Res* 44, W07406:1–18
47. Uniform Interface to Computing Resources (UNICORE) (2011) <http://www.unicore.eu/>
48. Vandembroucke D (2010) Spatial data infrastructures in Europe: state of play spring 2010. K.U. Leuven, Leuven. <http://inspire.jrc.ec.europa.eu/reports/stateofplay2009/INSPIRE%20NSDI%20SoP%20-%20Summary%20Report%202010%20-%20v4%204.pdf>
49. Whittaker G (2004) Use of a Beowulf cluster for estimation of risk using SWAT. *Agron J* 96(5):1495–1497
50. Worldwide LHC Computing Grid (WLCG) (2011) <http://lcg.web.cern.ch/LCG/>
51. WWF (2008) Greening the Black Sea synergy. EU Regional Office, Brussels
52. Yalaw SG, van Griensven A, Ray N, Betrie GD, Kokoszkiwicz L (in review) Parallel computing of a large scale SWAT model. *Environ Model Softw*