

Grid and sensor web technologies for environmental monitoring

Nataliia Kussul · Andrii Shelestov · Serhiy Skakun

Received: 12 September 2008 / Accepted: 27 February 2009 / Published online: 20 March 2009
© Springer-Verlag 2009

Abstract In this paper, we focus on the description of the Grid infrastructure that is under the development at the Space Research Institute NASU-NSAU (SRI). We describe several real-world applications that are being solved using the Grid infrastructure: numerical weather prediction (NWP), flood monitoring, and biodiversity assessment. We also review issues regarding the integration of the Sensor Web and Grid technologies for flood applications.

Keywords Biodiversity · Flood · Grid · Numerical weather prediction · Sensor web

Introduction

At present, global climate changes on the Earth made a rational land use, environmental monitoring, forecasting of natural and technological disasters, the tasks of great importance. The basis for the solution of these crucial applied problems consists in the integrated use of data of different nature: modeling data, in situ measurements and observations, and indirect observations such as airborne and spaceborne remote sensing data (GEOSS 2005).

In particular, models can be used to fill in the gaps in the data by extrapolating and estimating necessary parameters to the site of interest; to better understand and forecast

different processes occurring in the atmosphere, land, ocean, and sea; they can also help to interpret measurements and to design new observing systems. In situ measurements are often used for assimilation into models, calibration and validation of both modeling and remote sensing data. Satellite observations have an advantage of acquiring data for large and hard-to-reach territories, as well as providing continuous and human-independent measurements. Many important applications such as monitoring and forecasting of natural disasters, environmental monitoring, heavily rely on the use of Earth observation (EO) data from space. For example, the satellite-derived flood extent is 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). Information on flood extent provided in the near real-time (NRT) can also be used for damage assessment and risk management, and can benefit rescuers during flooding (Corbley 1999). Both spaceborne microwave and optical data can provide means to detect, estimate extent, and assess the damages caused by the drought events (Kogan et al. 2004; Wagner et al. 2007). To assess vegetation health/stress, which is extremely important for agriculture applications, optical remote sensing data can be used to derive biophysical and biochemical variables such as pigment concentration, leaf structure, water content at leaf level, and leaf area index (LAI), fraction of photosynthetically active radiation absorbed by vegetation (FPAR) at canopy level (Liang 2004). Thus, it can be seen that the aforementioned applications require the use of high-performance computing, reliable infrastructure for efficient data management and processing. These applications can benefit from the use of Grid and Sensor Web technologies in the following way: Grid environment can provide access to high-performance resources and enable efficient management of large volumes of data. Sensor Web can provide framework for integration of heterogeneous sensors into a common informational infrastructure.

Communicated by H. A. Babaie

N. Kussul (✉) · A. Shelestov · S. Skakun
Space Research Institute NASU-NSAU,
Glushkov Prospekt 40, building 4/1,
Kyiv 03680, Ukraine
e-mail: inform@ikd.kiev.ua
URL: <http://inform.ikd.kiev.ua>

A. Shelestov
e-mail: inform@ikd.kiev.ua

S. Skakun
e-mail: serhiy.skakun@ikd.kiev.ua

More specifically, the EO domain is characterized by the large volumes of data that should be processed, catalogued, and archived (Fusco et al. 2007; Shelestov et al. 2006). For example, the global ozone monitoring experiment (GOME) instrument onboard Envisat satellite generates nearly 400 Tb data per year (Fusco et al. 2003). The processing of satellite data is carried out not by the single application with a monolithic code, but by the distributed applications. This process can be viewed as a complex workflow (DEGREE 2008) that is composed of many tasks: geometric and radiometric calibration, filtration, reprojection, composites construction, classification, products development, post-processing, visualization. For example, calibration and mosaic composition of 80 images generated by the Advanced Synthetic-Aperture Radar (ASAR) instrument onboard Envisat satellite takes 3 days on ten workstations of the Earth Science GRID on Demand (G-POD) that is being developed at the European Space Agency (ESA) and the European Space Research Institute (ESRIN; Fusco et al. 2007). The mosaic covers all Europe at 90 m resolution, and corresponding products are automatically orthorectified with the Digital Elevation Model (DEM). Dealing with EO data, we have to also consider the security issues regarding satellite data policy, the need for processing in NRT for fast response within international programs and initiatives, in particular the International Charter “Space and Major Disasters” (www.disasterscharter.org) and the International Federation of Red Cross (www.ifrc.org).

It should be also noted that the same EO data sets and derived products can be used for a number of applications. For example, information on land use/change, soil properties, meteorological conditions are both important for flood and drought applications as well as for vegetation state assessment. That is, once we develop interfaces to discover and access the required data and products, they can be used in a uniform way for different purposes and applications. This represents one of the important tasks that are being solved within the development of the Global Earth Observation System of Systems (GEOSS 2005) and European initiative Global Monitoring for Environment and Security (GMES 2004).

A considerable need therefore exists for an appropriate infrastructure that will enable the integrated and operational use of multisource data for different application domain. From a technological point of view, Grids can provide solutions to the aforementioned problems (Foster 2002; Fusco et al. 2007; Shelestov et al. 2006). In this case, a Grid environment can be considered not only for providing high-performance computations but, in fact, can 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). In this context, we may refer to the GENESI-DR project (www.genesi-dr.eu) that aims at building a grid-based infrastructure facing these challenges and supporting GEOSS architecture.

In this paper, we review existing tendencies and initiatives, in particular GEOSS and GMES, and Grid-related projects for EO applications. We will provide the description of the existing Grid infrastructure that is under the development at the Space Research Institute NASU-NSAU (SRI). We will describe several real-world applications that are solved using the Grid infrastructure, namely, numeral weather prediction (NWP), flood monitoring, and biodiversity assessment. We also review issues regarding the integration of the Sensor Web and Grid technologies for flood applications. The paper ends with conclusion remarks.

Existing tendencies and initiatives

GEOSS and GMES

The globalization and integration processes are dominant tendencies in the development of new solutions for complex problems solving. At present, international cooperation efforts are focused on the implementation of GEOSS. GEOSS is a distributed system of systems built on current international cooperation among existing Earth observing and data management systems—in situ and remote sensors and systems (GEOSS 2005).

GMES is a European initiative for the implementation of information services dealing with environment and security; support for emergency management in the case of natural hazards; forecasting for marine zones, air quality, or crop yields; and so on (GMES 2004). The GMES capacity is based on four inter-related components: services, observations from space, in situ, and data integration and information management capacity. The data integration and information management will enable user access and the sharing of information.

In both GEOSS and GMES activities, it is stated that the areas that are data and computationally intensive require high-performance networks and Grid-based computing for the essential data mining, sharing and analyzing, and visualization of the results.

In the following subsection, we briefly describe several projects and initiatives that deal with the application of Grid technology for the EO domain.

Grid projects for EO applications

At present, Grid technologies are widely applied in different domains, in particular the EO domain.

European DataGrid Project (EDG) was the first large European Commission-funded Grid project (www.eu-datagrid.org). Many of the results of EDG project have been included in the European project Enabling Grids for E-science (EGEE). EGEE aims to develop a service grid infrastructure which is available to scientists 24 h a day.

Based on the gained experience, ESA and ESRIN have focused on the development of the Earth observation Grid processing on-demand infrastructure (G-POD; Fusco et al. 2007). Grid is considered as a comfortable “open platform” for handling computing resources, data, tools, and not limited to only high-performance computing. G-POD enables access to different data and products from Envisat satellite (<http://envisat.esa.int>), SEVIRI instrument onboard MSG (Meteosat Second Generation) satellite, and so on. One of the most important applications is the analysis of long-term data. For example, the analysis of 8 years of GOME onboard temperatures (overall 525 Gb of data) took less than 2 days on 40 computer elements of ESRIN “Grid-on-demand” structure (overall 38,460 files were processed; Fusco et al. 2007). At present, G-POD infrastructure consists of more than 150 working nodes with ability to store and handle about 100 Tb of data.

FAIRE is another Grid-based application that is operationally used by ESA in the context of flood mapping. The application takes advantage of Grid technology for NRT data access, calibration, orthorectification, map projection, coregistration. The application is operationally used in the context of the International Charter “Space and Major Events”.

Dissemination and Exploitation of GRids in Earth science (DEGREE) project is a European-funded project that aims to build a bridge linking the Earth Science and Grid communities throughout Europe (DEGREE 2008). Grid is considered to be the appropriate platform for integration of heterogeneous data resources, processing tools, models, algorithms, and so on. The following applied problems are within the scope of DEGREE: earthquake analysis, floods modeling and forecasting, influence of climate changes on agriculture are to be mentioned.

The Japan Aerospace eXploration Agency (JAXA) and the Keio University started establishing the Digital Asia system aimed at semi-real time data processing and analyzing. They use Grid environment to accumulate knowledge and know-how to process the remote sensing data. The Digital Asia project is a part of the Sentinel Asia project that is targeting on building natural disaster monitoring system (<http://dmss.tksc.jaxa.jp/sentinel>).

The Wide Area Grid (WAG) project is initiated by the Working Group on Information Systems and Services (WGISS) of the Committee on Earth Observation Satellites (CEOS), and aims to develop the “horizontal” infrastructure in order to integrate computational, human, intellectual, and

informational resources of the space agencies within a large distributed system. Implementation of geospatial-related services and Grid-enable EO data archives are among the priority tasks in this project (Kopp et al. 2007).

The Space Research Institute NASU-NSAU have created a basic computational Grid infrastructure, provided the proof of concept for the solution of complex problems arising in the space weather, hydro-meteorological modeling, and flood monitoring (Kussul et al. 2008a). The Grid infrastructure is developed within several international projects, namely, INTAS-CNES-NSAU project “Data Fusion Grid Infrastructure,” STCU-NASU projects “Grid Technologies for Multi-Source Data Integration,” and “Grid technologies for environmental monitoring using satellite data.”

Description of Grid infrastructure

Grid infrastructure for EO applications

Currently, the Grid infrastructure integrates the resources of several geographically distributed organizations, in particular:

- Space Research Institute NASU-NSAU (Ukraine) with deployed computational and storage nodes based on Globus Toolkit 4 (<http://www.globus.org>) and gLite 3 (<http://glite.web.cern.ch>) 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 computational resources (approximately 500 processors);
- RSGS-CAS (China) with deployed computational nodes based on gLite 3 middleware and access to geospatial data (approximately 16 processors).

In all cases, the Grid Resource Allocation and Management (GRAM) service (Feller et al. 2007) is used to execute jobs on the Grid resources.

It is also worth mentioning that satellite data are distributed through the Grid environment. For example, ENVISAT WSM data (that are used within the flood application) are stored on the ESA’s rolling archive and routinely downloaded for the Ukrainian territory. Then, they are stored at the Space Research Institute archive that is accessible via the Grid. MODIS data from Terra and Aqua satellites that are used in flood, crop yield, and biodiversity assessment applications are routinely downloaded from the USGS’ archives and stored at the Space Research Institute NASU-NSAU and Institute of Cybernetics of NASU.

Access to the resources of the Grid environment is organized via a high-level Grid portal that have been deployed using GridSphere framework (<http://www.grid-sphere.org>). 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 (Fig. 1).

The workflow of the data processing steps in the Grid (such as transformation, calibration, orthorectification, classification) is controlled by a Karajan engine (<http://www.gridworkflow.org/snips/gridworkflow/space/Karajan>).

The existing architecture of the Grid is shown in Fig. 2.

Visualization of data in grid infrastructure

In order to visualize the results of data processing in the Grid environment, we use an open-source OpenLayers framework (<http://www.openlayers.org>), and UNM Mapserver v5. OpenLayers is a JavaScript library for building rich web-based geographic applications, with no server-side dependencies. OpenLayers implements industry-standard methods for geographic data access, such as the Open Geospatial Consortium's Web Mapping Service (WMS) and Web Feature Service (WFS) protocols.

Mapserver is an Open Source development environment for building spatially enabled Internet applications. It

supports the OGC's WMS standard that enables the creation and display of registered and superimposed map-like views of information that come simultaneously from multiple remote and heterogeneous sources (Beaujardiere 2006).

Having created WMS services for the EO-derived products, we use them in the OpenLayers framework and in Google Earth by generating corresponding Keyhole Markup Language (KML) files.

The examples of results of data processing are given in the next section.

Applications deployed in Grid infrastructure

In this section, we describe in detail EO applications that were deployed in the Grid infrastructure. In particular, we focus on the weather modeling application, flood monitoring, and biodiversity assessment. The motivation for the selection of these applications comes from the following:

1. Numerical weather prediction belongs to computational intensive applications.
2. Flood applications need the fast response to the emergencies, and thus require a reliable infrastructure for data management and processing.

The screenshot shows the GridSphere Portal Framework interface in a Mozilla Firefox browser. The page title is "GridSphere Portal Framework" and the user is logged in as "Mykhailo Korbakov". The main content area is a "Resource Browser Portlet" displaying a "Hardware List" table.

Resource	Name	Address	Description	System	Processor	CPUs
	Portal	gridportal.ikd.kiev.ua	Hosts the GridSphere Portlet Container	-	-	-
	Aurora	gt.ikd.kiev.ua	Front-end to the Aurora cluster	Linux 2.6.9-023stab044.4-smp	x86	4 CPUs
	SKIT-3 (GEOSS segment)	gt.lcyb.kiev.ua	Front-end to the SKIT-3 cluster	-	-	-
	RSGS SE	se.grid.rsgs.ac.cn	Remote Sensing Ground Station Storage Element	Linux 1 Thu Oct 19 11:02:51 CDT 2006 2.4.21-47.0.1.EL	-	-
	RSGS CE	ce.grid.rsgs.ac.cn	Remote Sensing Ground Station Computing Element	-	-	-
	SRI SE	se.grid.ikd.kiev.ua	Space Research Institute Storage Element	Linux 1 Thu Oct 19 11:02:51 CDT 2006 2.4.21-47.0.1.EL	-	-
	SRI CE	ce.grid.ikd.kiev.ua	Space Research Institute Storage Element	-	-	-

The interface also shows a search bar at the bottom with "Done" and "Find" options, and a status bar indicating "110% (ikd.kiev.ua)" and "Open Notebook".

Fig. 1 Portal of the Grid infrastructure

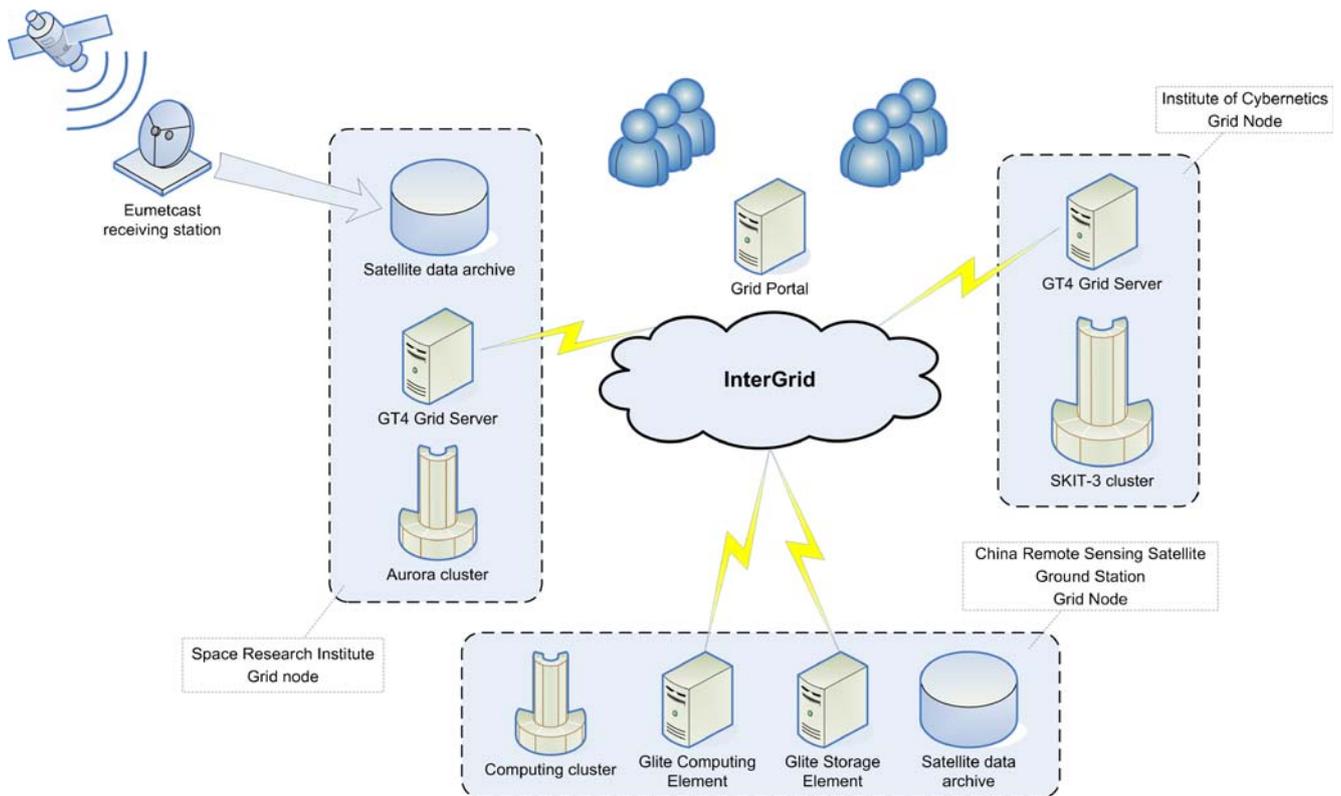


Fig. 2 Architecture of the Grid infrastructure

3. Biodiversity assessment belongs to data intensive application where different data and products are analyzed in order to produce the final product.

Numerical weather modeling

Forecasting meteorological parameters represents one of the core services for a number of applications (e.g., floods, droughts, agriculture). Currently, we run the Weather Research and Forecasting model (WRF; Michalakes et al. 2004) in operational mode for the territory of Ukraine. The meteorological forecast is generated every 6 h with a spatial resolution of 10 km. Forecast range is 72 h. The horizontal grid dimensions are 200×200 points with 31 vertical levels. We use NCEP GFS (Global Forecasting System) forecast as boundary conditions for the WRF model. This data is available via Internet through the National Operational Model Archive and Distribution System (NOMADS system).

The workflow of the WRF model run is composed of the following steps (Fig. 3): (1) data acquisition; (2) data pre-processing, computation of forecast using WRF model and data post-processing; (3) visualization of the forecast.

Data acquisition To run WRF model, it is necessary to obtain boundary and initial conditions for the territory of

Ukraine. This data can be extracted from GFS model forecast. To get the required data, the dedicated script was developed. This script downloads global forecast every 6 h. To decrease the data volume, our script uses special Web-service capable of selecting subsets of the GFS data for the territory of Ukraine. The acquired data is transferred to the storage subsystem and marked as unprocessed (i.e., it has to be processed by the WRF model). After the GFS data has been downloaded, the Karajan script initializes a workflow for the data pre-processing, WRF run, and data post-processing.

Data pre-processing step is intended to transform the downloaded data into the format that is used to run the WRF model. GFS data is delivered in the GRIB format in the geographical projection. This data is transformed into the internal WRF format by the `grib_prep.exe` command, warped into the Lambert Conformal Conic projection (by executing `hinterp.exe` command) and vertically interpolated using the `vinterp.exe` command. (`grib_prep.exe`, `hinterp.exe` and `vinterp.exe` commands are tools from WRF Standard Initialization (SI) package.) The results of these transformations are stored in the netCDF format. After that, the `real.exe` command is used to produce initial and boundary conditions for WRF model run. The inputs to `real.exe` command are GFS data in netCDF format and WRF configuration file (`namelist.input`).

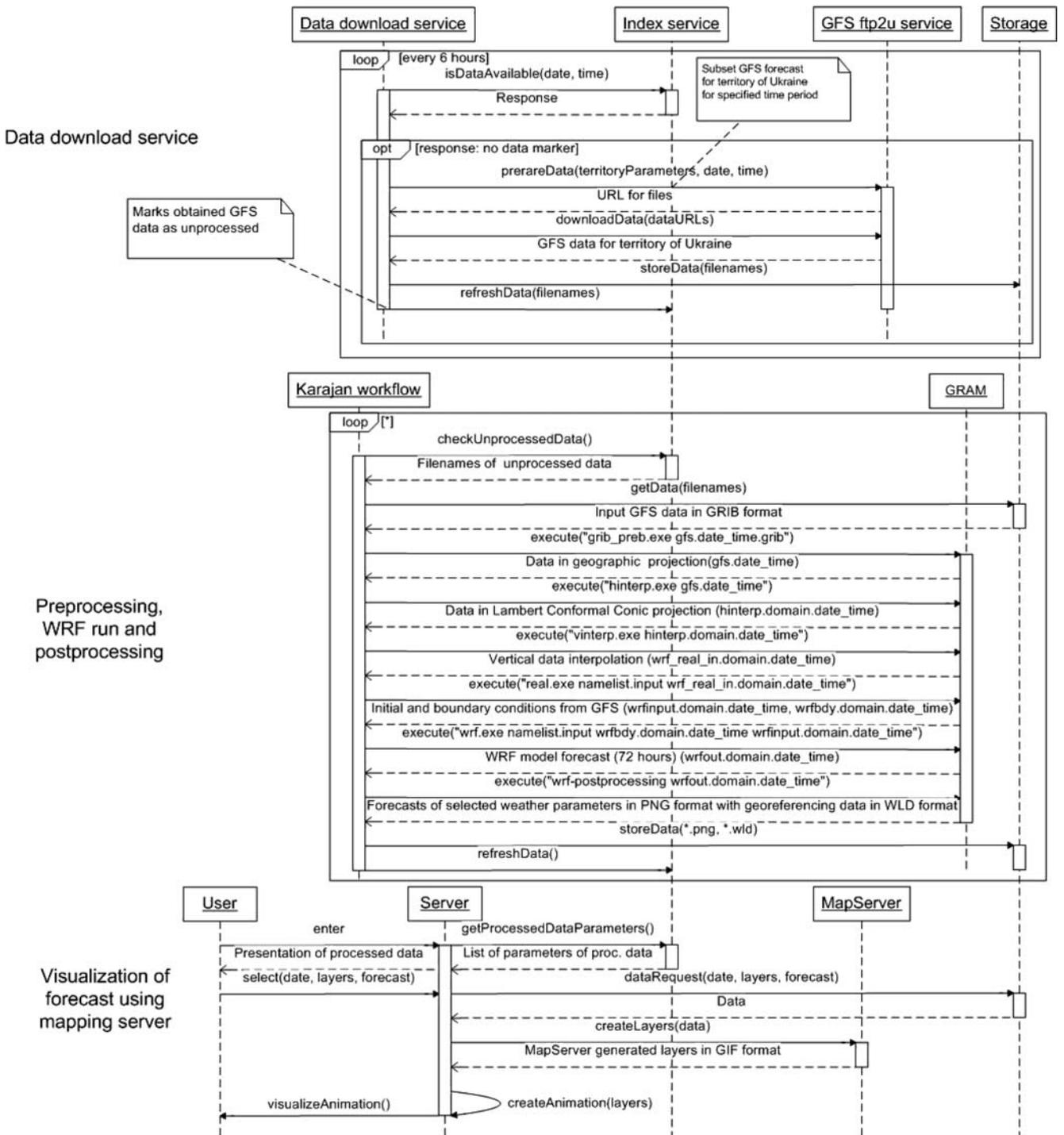


Fig. 3 UML sequence diagram (Larman 2004) for the NWP application

Data processing step consists in performing WRF run using wrf.exe command. The output of the command is the forecast of the meteorological parameters. This is the most computationally intensive task.

After WRF model run, post-processing step is carried out. For specified weather parameters and for each forecast frame (3 h), a graphic representation (in PNG format) of

spatial distribution is created. Additionally, special files containing georeferencing information are created (files with *.wld extension). The results of the post-processing phase are used to visualize the WRF forecast via the mapping service. This service is available via <http://dos.ikd.kiev.ua>, and provides the users animations of the weather forecast (Fig. 4). The service provides tools to select a

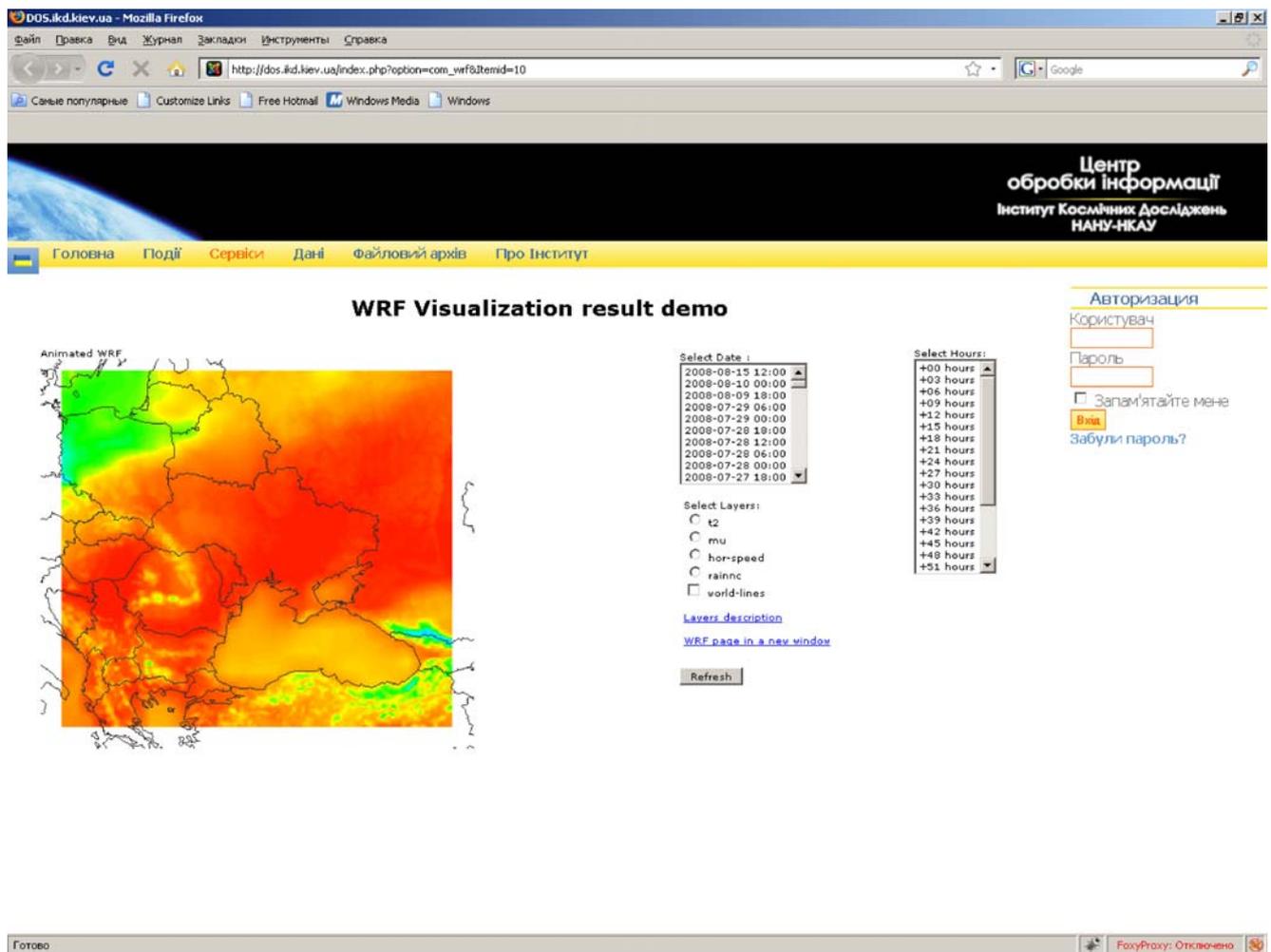


Fig. 4 The example of land temperature forecast using WRF model

forecast time, forecast frames (up to 72 h ahead), and weather parameters to be displayed. Selected by the user information is packed into the request to the server. To process the request, all required data (in PNG and WLD formats) is retrieved from the storage subsystem and passed to the mapping server in order to create the maps. Maps are further processed by the script to generate weather animation in GIF format. Finally, this animation is presented at user side.

We have also tested the performance of the WRF model in dependence of the number of computational nodes. For test purposes, we used the parallelized version 2.2 of the WRF model with a model domain identical to those used in operational NWP service ($200 \times 200 \times 31$ gridpoints with horizontal spatial resolution 10 km). Parallelization was implemented using the message passing interface (MPI). We observed almost linear productivity growth within increasing number of computation nodes. For instance, eight nodes of the SCIT-3 cluster of the Grid infrastructure gave the performance increase in 7.09 times (of 8.0 theoretically possible) when

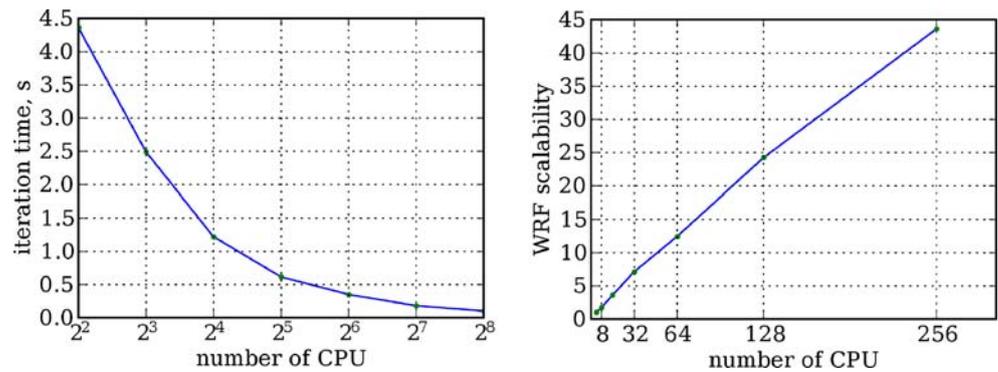
compared to the single node. The use of 64 nodes increases the performance in 43.6 times (see Fig. 5).

The single iteration of the model run corresponds to the forecast of meteorological parameters 1 min in advance. Hence, 3 days forecast requires completion of 4,320 iterations. That is, when using one node of SCIT-3 cluster of the Grid infrastructure, it takes 5.16 h to provide 3 days forecast. In turn, the use of 64 nodes of the cluster makes possible to reduce the overall computing time up to 7.1 min.

Flood extent extraction from SAR imagery

One of the most important problems associated with a flood monitoring is a flood extent extraction, since it is impractical to determine the flood area through field observations. We have developed a neural network approach to flood extent extraction from synthetic aperture radar (SAR) imagery (Kussul et al. 2008b). In contrast to optical data, SAR measurements from space are indepen-

Fig. 5 The results of WRF performance on the SCIT-3 cluster: computation time for one iteration (*left*); acceleration of the WRF model with respect to a number of nodes (*right*)



dent of daytime and weather conditions and can provide valuable information to monitoring of flood events. Neural network is used to segment and classify the image on two classes: “Water” and “No water.” As inputs to neural network, we used a moving window of image pixels intensities. We applied our approach to determine flood extent from SAR images acquired by three different sensors: ERS-2/SAR (spatial resolution 8 m) for the river Tisza, Ukraine (2001); ENVISAT/ASAR WSM (Wide Swath Mode, spatial resolution 150 m) and RADARSAT-1 (spatial resolution 25 m) for the river Huaihe, China (2007). The size of the window depended on the satellite instrument imaging mode. For example, for data acquired by Envisat/ASAR in Wide Swath Mode, we used a 3-by-3 window; for ERS-2 and RADARSAT-1 data, we used a 7-by-7 window. Classification rates for independent testing data sets were

85.40%, 98.52%, and 95.99% for ERS-2/SAR, ENVISAT/ASAR WSM and RADARSAT-1 data, respectively (Kussul et al. 2008b).

We developed a parallel version of our method and deployed it in the Grid infrastructure. Parallelization of the image processing is performed in the following way: SAR image is split into the uniform parts that are processed on different nodes using the OpenMP Application Program Interface (www.openmp.org). The use of the Grids allowed us to considerably reduce the time required for image processing. In particular, it took approximately 10–30 min (depending on image size) to process a single SAR image on a single workstation. The use of Grid computing resources allowed us to reduce the time to less than 1 min. The example of the flood extent extraction product is shown in Fig. 6.

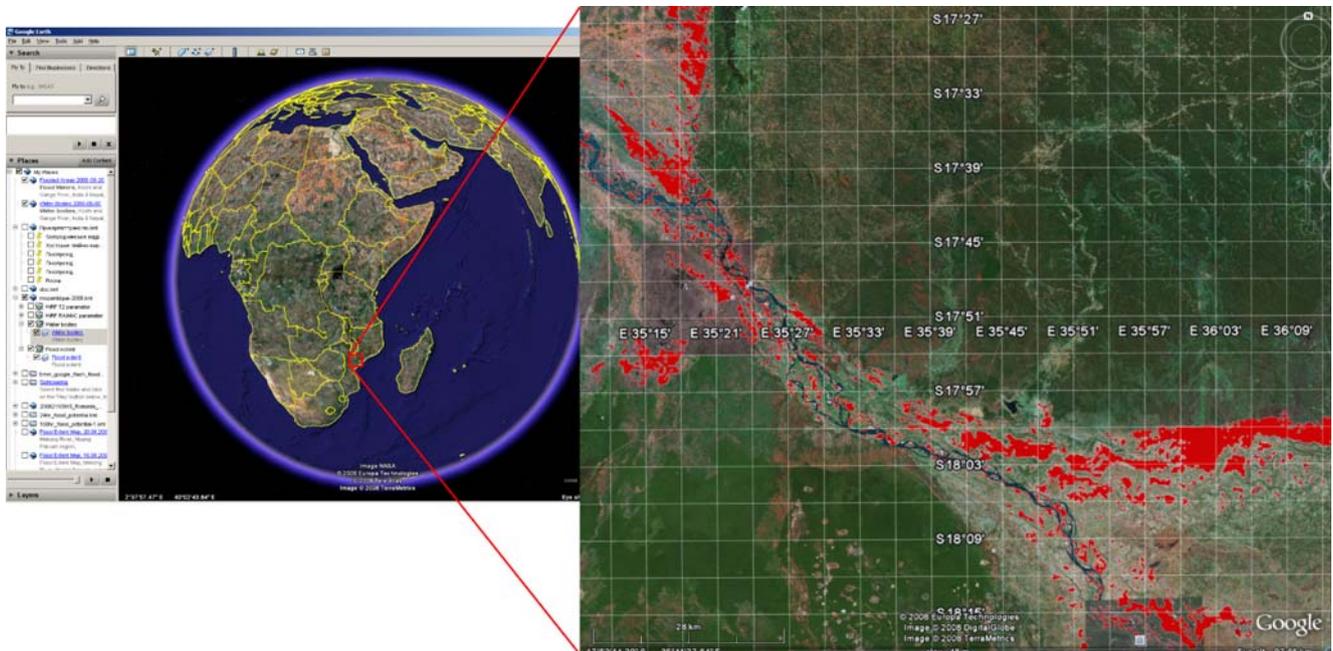


Fig. 6 Visualization of the results of image processing for ENVISAT/ASAR WSM data during flooding on the river Zambezi, Mozambique (February 2008). Flood extent is shown with *red color*

Land biodiversity assessment

In the framework of the innovative project of the National Academy of Sciences of Ukraine, scientists from the Scientific Centre for Aerospace Research of the Earth (CASRE) and the SRI NASU-NSAU have jointly developed a Web service for land biodiversity assessment for the Pre-Black Sea region of Ukraine (Popov et al. 2008) using EOS data products (King et al. 2004).

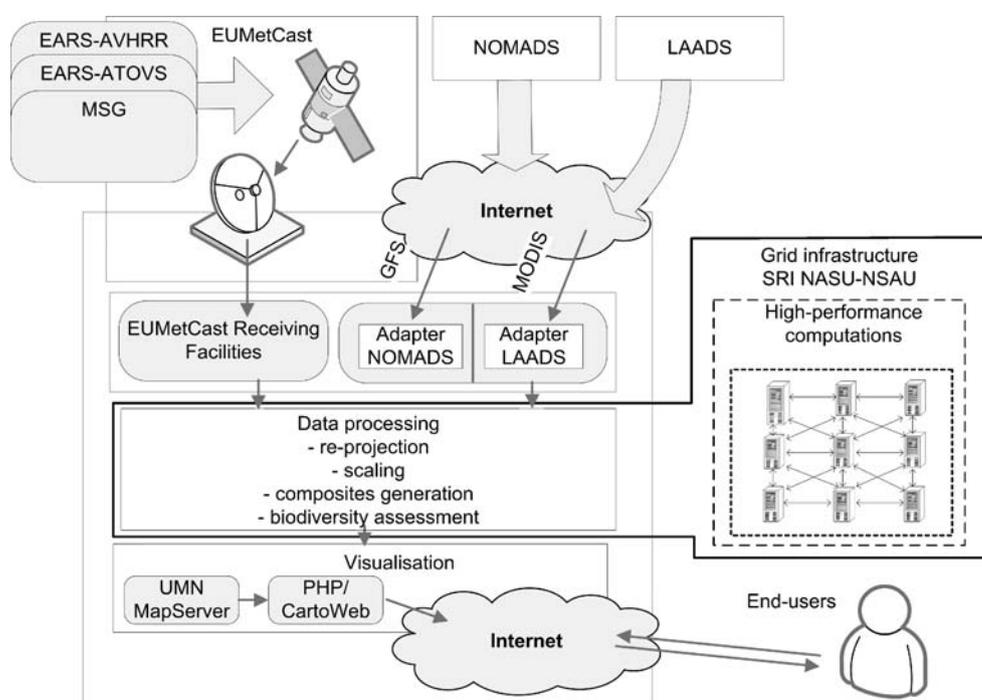
Biodiversity is associated with a number of abiotic and biological factors that can be identified using remote sensing data. These factors include: landscape types, geographical latitude/altitude, climate conditions (such as mean daily temperatures, humidity), structure, and primary productivity of a vegetation mantle (Hansen and Rotella 1999). These factors can be estimated using EO data from space (Popov et al. 2008). The workflow for biodiversity estimation consists of the following steps: data acquisition, data processing, and visualization. Figure 7 shows the overall architecture of the service with information flows.

Special system was developed in order to acquire satellite data on regular basis. This system operationally monitors for the new products and provides automatic data acquisition from different sources: Level 1 and Atmosphere Archive and Distribution System (LAADS), Land Processes Distributed Active Archive Center (LP DAAC), and National Snow and Ice Data Center (NSIDC). The acquired data are stored in the data archive of SRI. The detailed UML sequence diagram for the data acquisition step is shown in Fig. 8.

After the required data has been acquired, the data is re-projected to a conical Albert projection and scaled to the spatial resolution of 250 m. Since we use data from multiple sources, different tools were applied for the re-projection and scaling purposes. In particular, we used MODIS Swath Reprojection Tool, MODIS Reprojection Tool, and Geospatial Data Abstraction Layer (GDAL) library (<http://www.gdal.org>). Since biodiversity index represents a parameter that is estimated for the time range, it is required to calculate average values for the parameters influencing biodiversity. For this purpose, average composites of images were created. Using these composites and solar irradiation acquired from SRTM DEM v2, we estimated the biodiversity index using the fuzzy model (Popov et al. 2008). The resulting product is a georeferenced file in GeoTIFF format showing biodiversity index over the given region. The workflow of the data processing step is controlled by the Karajan engine while the data are processed on the computational resources of the Grid system using the GRAM service (Feller et al. 2007). The detailed UML sequence diagram for the data processing and visualization steps is shown in Fig. 9.

The proposed Web service is implemented on the basis of OGC standards, Web Map Service 1.1.1 (<http://www.opengeospatial.org/standards/wms>) and Web Coverage Service 1.0 (<http://www.opengeospatial.org/standards/wcs>). The developed Web service is accessible via Internet through the address <http://biodiv.ikd.kiev.ua> (Fig. 10). It represents a current distribution of the potential biodiversity and allows monitoring each of the factors that influence biodiversity.

Fig. 7 Overall architecture of the service with information flows



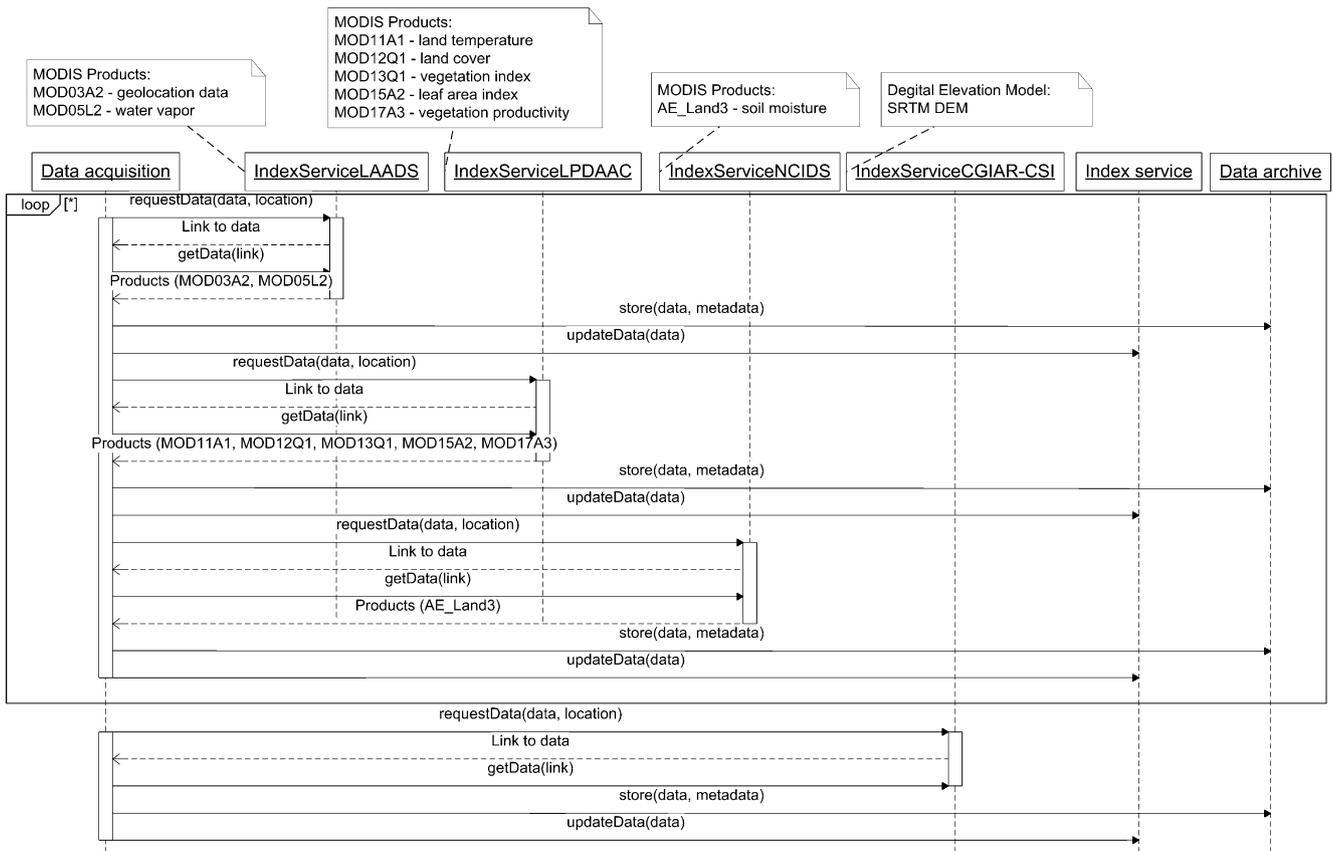


Fig. 8 UML sequence diagram for the data acquisition step of the biodiversity assessment procedure

Summarizing, we may point out the following benefits of using Grid technologies for the described applications. Within the meteorological application, the use of the Grid system resources made it possible to considerably reduce the time required for the model run (up to 43.6 times). It is especially important for the cases when one needs to tune the model and adapt it to the specific region and thus run the model multiple times to find the best configuration and parameterization. For the flood application, Grids also allowed us to reduce the overall computing time required for satellite image processing, and made possible the fast response within international programs and initiatives concerned with emergencies. As to the biodiversity application, the benefits of the Grids come from the ability to manage large volumes of data, and to provide high-performance computations, since the analysis of historical data is required.

Sensor web and Grid integration: main problems and possible solutions

Decision makers in emergency response situations (e.g., floods, droughts) need rapid access to the existing data,

the ability to request and process data specific to the emergency, and tools to rapidly integrate the various information sources into a basis for decisions. The flood forecasting and monitoring scenario presented here is being implemented within the GEOSS AIP-2 (Architecture Implementation Pilot Phase-2, <http://www.ogcnetwork.net/AIPIlot>). It uses precipitation data from the Global Forecasting System (GFS) model and NASA's Tropical Rainfall Measuring Mission (TRMM, <http://trmm.gsfc.nasa.gov>) to identify the potential flood areas. Once the areas have been identified, we can request satellite data for the specific territory for flood assessment. These data can be both optical (like EO-1, MODIS, SPOT) and microwave (Envisat, ERS-2, ALOS, Radarsat-1).

From technological view point, the scenario is implemented using the Sensor Web (Moe et al. 2008; Mandl et al. 2006) and Grid. The integration of sensor networks with Grid computing brings out dual benefits (Chu et al. 2006): (1) sensor networks can off-load heavy processing activities to the Grid and (2) Grid-based sensor applications can provide advance services for smart-sensing by deploying scenario-specific operators at runtime.

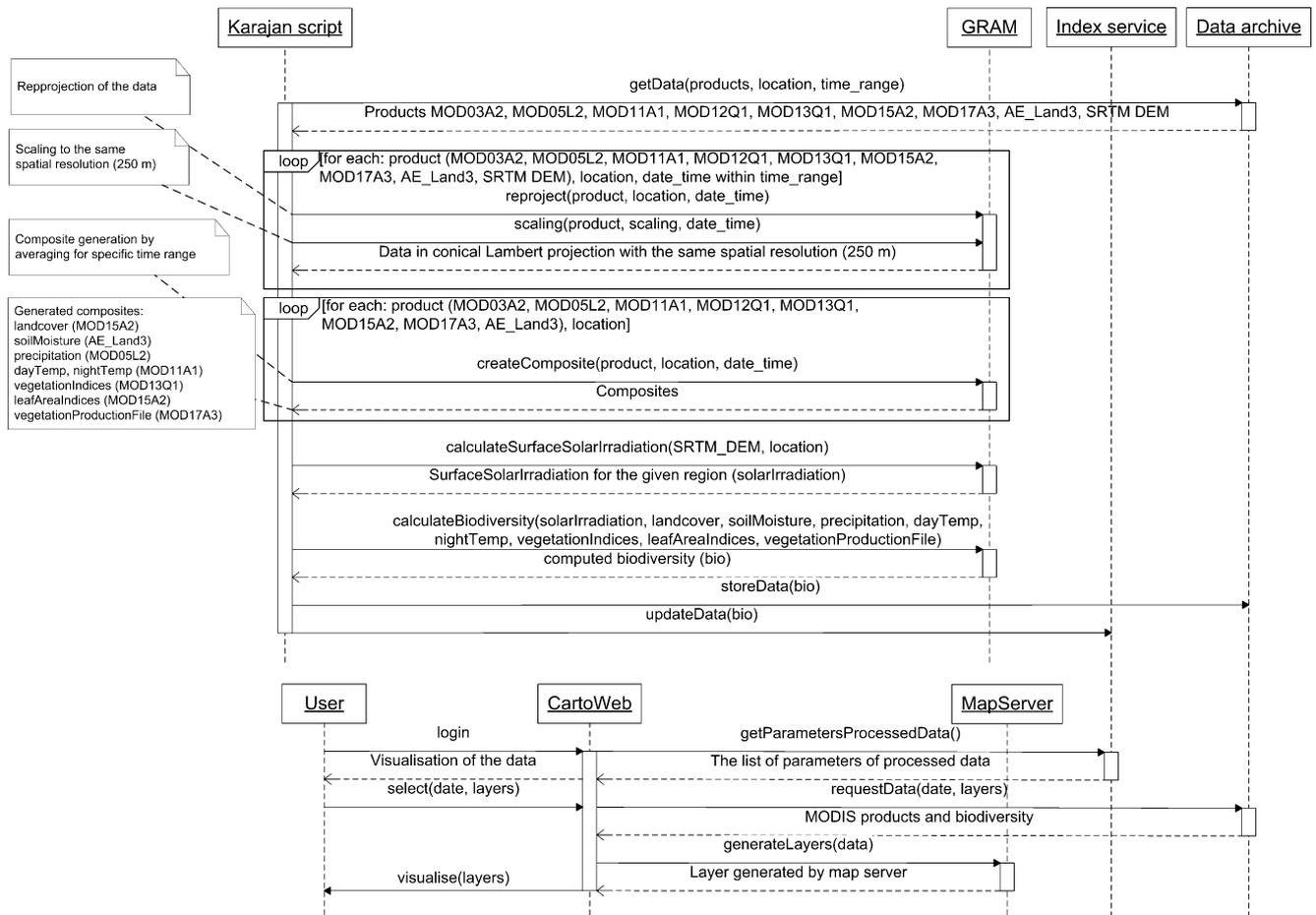


Fig. 9 UML sequence diagram for the data processing and visualization steps of the biodiversity assessment procedure

Sensor web paradigm

Sensor Web is an emerging paradigm and technology stack for integration of heterogeneous sensors into common informational infrastructure (Moe et al. 2008; Mandl et al. 2006). The basic functionality required from such infrastructure is remote data access with filtering capabilities, sensors discovery, and triggering of events by sensors conditions.

Sensor Web is governed by the set of standards developed by the Open Geospatial Consortium (Botts et al. 2007). At present, the following standards are available and approved by the consortium:

- OGC Observations and Measurements (<http://www.opengeospatial.org/standards/om>)—common terms and definition for Sensor Web domain;
- Sensor Model Language (<http://www.opengeospatial.org/standards/sensorml>)—XML-based language for describing different kinds of sensors;
- Transducer Model Language (<http://www.opengeospatial.org/standards/tml>)—XML-based language for describing the response characteristics of a transducer;

- Sensor Observations Service (<http://www.opengeospatial.org/standards/sos>)—an interface for providing remote access to sensors data;
- Sensor Planning Service (<http://www.opengeospatial.org/standards/sps>)—an interface for submitting tasks to sensors.

There are also standards drafts that are available from Sensor Web working group but not yet approved as official OpenGIS standards:

- Sensor Alert Service—service for triggering different kinds of events basing of sensors data;
- Web Notification Services—notification framework for sensor events.

Sensor Web paradigm assumes that sensors could belong to different organizations with different access policies or, in broader sense, to different administrative domains. However, existing standards stack does not provide any means for enforcing data access policies leaving it to underlying technologies. One possible way for handling informational security issues in Sensor Web is presented in the next subsections.

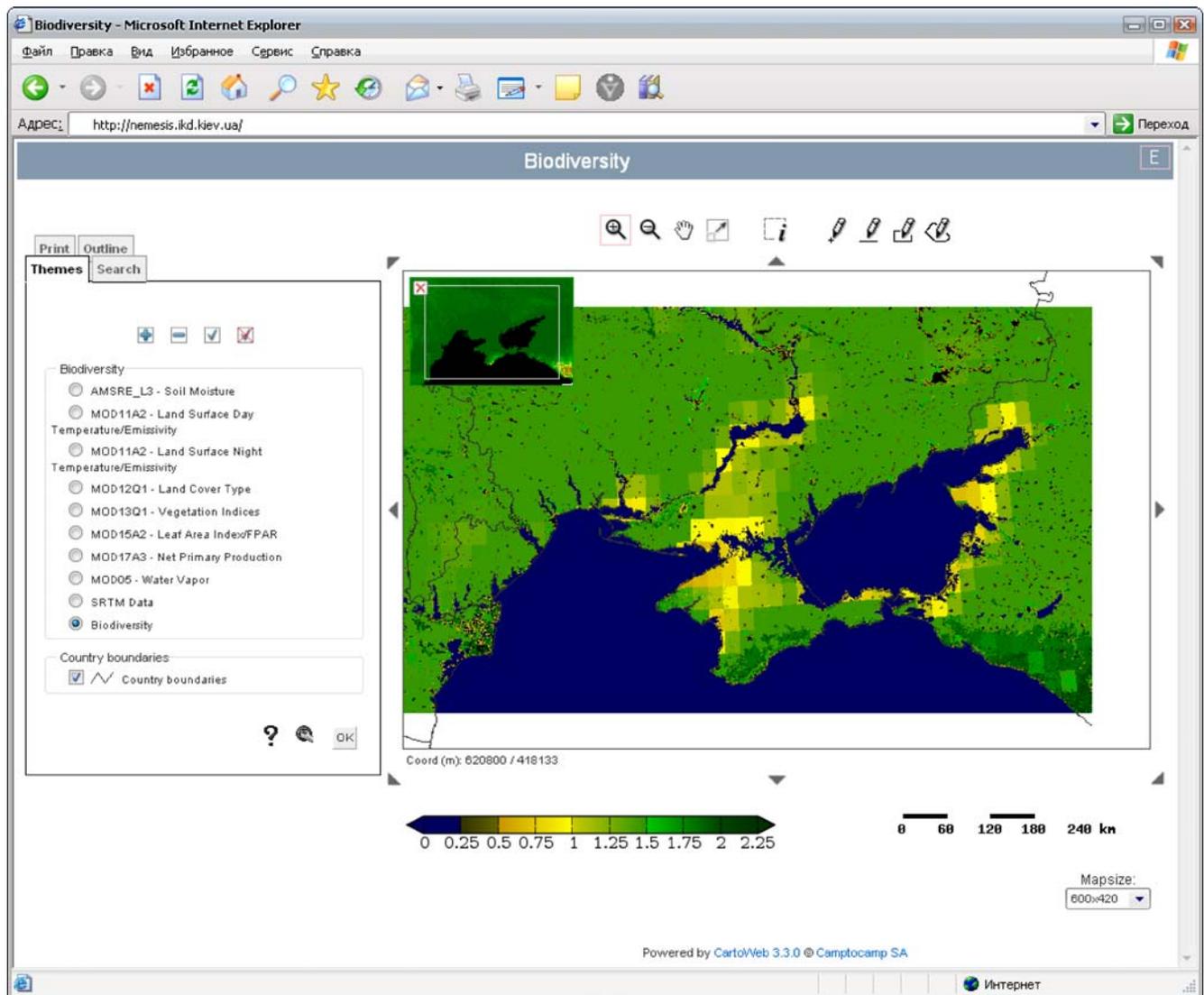


Fig. 10 Demonstration of Web service for biodiversity assessment using EOS data products for the Pre-Black Sea region of Ukraine

Sensor web flood use case

One of the most challenging problems for Sensor Web technology implementation is global ecological monitoring in the framework of GEOSS. In this paper, we consider the problem of flood monitoring using satellite remote sensing data, in situ data, and results of simulations.

The problem of flood monitoring by itself consumes data from many heterogeneous data sources such as remote sensing satellites (we are using data of ASAR, MODIS, and MERIS sensors), in situ observations (water levels, temperature, humidity). Floods forecasting is adding the complexity of physical simulation to the task.

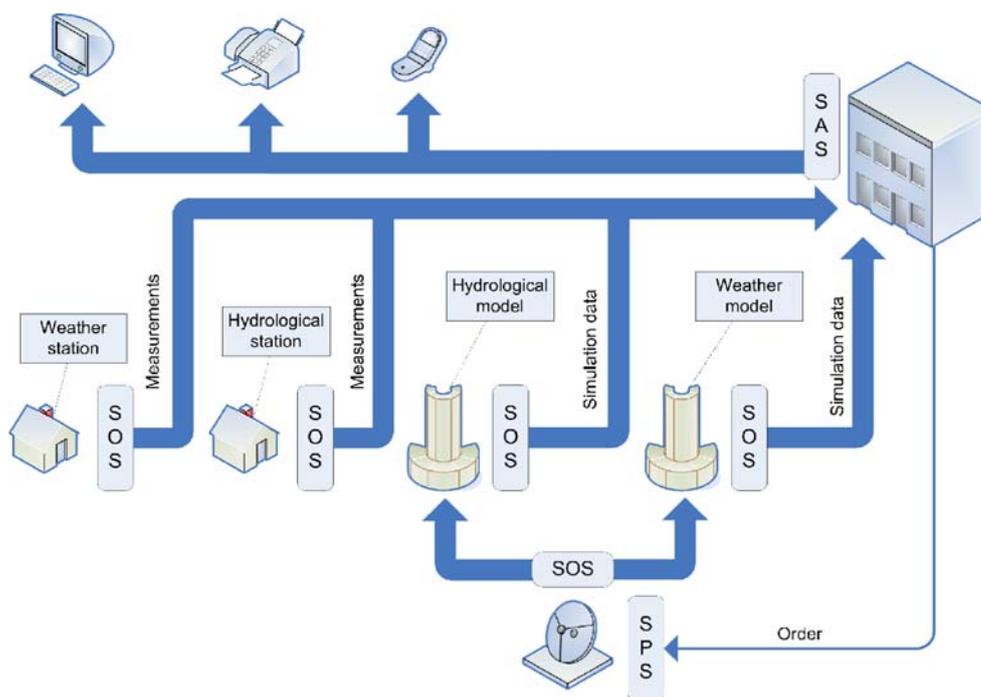
The Sensor Web perspective of this test case is depicted in Fig. 11. It shows collaboration of different

OpenGIS specifications of the Sensor Web. The data from different sources (numerical models, remote sensing, in situ observations) is accessed through the Sensor Observation Service (SOS). Aggregator site is running the Sensor Alert Service to notify interested organization of possible flood event using different communication mean. Aggregator site is also sending orders to satellite receiving facilities using the Sensor Planning Service (SPS) to get satellite imagery only available by preliminary order.

Sensor Web SOS gridification

Sensor Web services like SOS, SPS, and SAS can benefit from the integration with the Grid platform like Globus Toolkit (<http://www.globus.org>). Many Sensor Web fea-

Fig. 11 Sensor Web perspective of flooding test case



tures can take advantage of the Grid platform services, namely:

- Sensors discovery could be performed through the combination of Index Service and Trigger Service;
- High-level access to XML description of the sensors and services could be made through queries to the Index Service;
- Grid platform provides a convenient way for the implementation of notifications and event triggering using corresponding platform components (Humphrey et al. 2005);
- Reliable File Transfer (RFT) service (Allcock et al. 2005) provides reliable data transfer for large volumes of data;
- Globus Security Infrastructure (Welch et al. 2003) provides enforcement of data and services access policies in a very flexible way allowing implementation of desired security policy.

We have developed a testbed SOS Service using Globus Toolkit as a platform. Currently, this service works as a proxy translating and redirecting user requests to the standard HTTP SOS server (see Fig. 12). The current version uses client-side libraries for interacting with the SOS provided by the 52North in their OX-Framework. The next version will also include in-service implementation of SOS-server functionality.

Grid service implementing SOS provides the interface specified in the SOS reference document. The key difference between the standard interfaces and Grid-based implementations of the SOS lies in the encoding of service

requests. The standard implementation uses custom serialization for the requests and responses, and the Grid-based implementation uses standard SOAP encoding.

To get advantage of the most Globus features, the SOS service should export service capabilities and sensor descriptions as WSRF resource properties (Foster 2005). Traditionally, the implementation of such properties requires translation between XML Schema and Java code. However, the XML Schema of the SOS and related standards, in particular GML (Humphrey et al. 2005), is a very complex one, and there are no available program tools able to generate Java classes from it. We have solved this problem by storing service capabilities and sensor descriptions data as Document Object Model (DOM) Element objects, and using custom serialization for this class provided by the Axis framework that is used by the Globus Toolkit. Using this approach, we cannot access particular elements of the XML document in object-oriented style. However, the SOS Grid service is acting as a proxy between the user and SOS implementation, so it does not need to modify XML directly. With resource properties

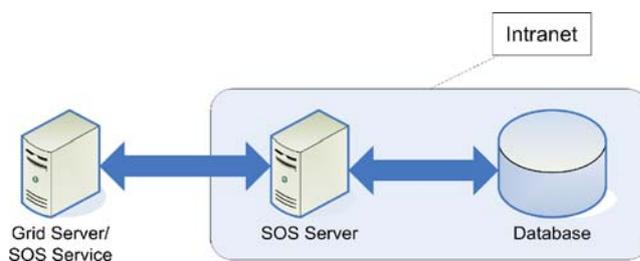


Fig. 12 Grid-based SOS service implementation

defined in this way, we can access it using standard Globus API or command line utilities.

Conclusions

In this paper, we presented the Grid infrastructure that is under the development in the Space Research Institute NASU-NSAU. The Grid integrates computational and storage resources of the geographically distributed organizations: the Space Research Institute NASU-NSAU, the Institute of Cybernetics NASU, and the China's Remote Sensing Satellite Ground Station of CAS. The use of Grid technologies for the EO domain is motivated by the need to make computations in the near real-time for fast response to natural disasters and to manage large volumes of satellite data. The EO applications are also characterized by the complex workflow that have to be managed and controlled.

We reviewed some issues regarding the integration of the Grid and the Sensor Web technologies. Such integration can bring out dual benefits: (1) sensor networks can off-load heavy processing activities to the Grid and (2) Grid-based sensor applications can provide advance services for smart-sensing by deploying scenario-specific operators at runtime.

We described several real-world applications that benefit from the use of the Grid infrastructure. The applications included: numeral weather prediction that is computationally intensive, flood applications that require fast response to the emergencies, and biodiversity assessment that requires the analysis and integration of large volumes of data in order to derive the final product.

Acknowledgements This work is supported by ESA CAT-1 project "Wide Area Grid Testbed for Flood Monitoring using Spaceborne SAR and Optical Data" (No. 4181); joint project of INTAS, the Centre National d'Etudes Spatiales (CNES) and the National Space Agency of Ukraine (NSAU), "Data Fusion Grid Infrastructure" (Ref. Nr 06-1000024-9154); joint project of the Science and Technology Center in Ukraine (STCU) and the National Academy of Sciences of Ukraine (NASU), "Grid Technologies for Multi-Source Data Integration" (no. 4928), and the Ministry of Education and Science of Ukraine, "Development of Integrated Remote Sensing Data Processing System using Grid Technologies" (no. M/72–2008).

References

- Allcock W, Bresnahan J, Kettimuthu R, Link M (2005) The Globus Striped GridFTP Framework and Server. In Proc ACM/IEEE SC 2005 Conf on Supercomputing. doi:10.1109/SC.2005.72
- Beaujardiere J (ed) (2006) OpenGIS® Web Map Service Implementation Specification. Open Geospatial Consortium Inc.
- Botts M, Percivall G, Reed C, Davidson J (2007) OGC Sensor Web Enablement: Overview and High Level Architecture (OGC 07-165) Accessible via http://portal.opengeospatial.org/files/?artifact_id=25562 Accessed 10 Sept 2008
- Chu X, Kobialka T, Durnota B, Buyya R (2006) Open sensor web architecture: core services. In: Proc of the 4th Int Conf on Intelligent Sensing and Information Processing (ICISIP). IEEE Press, Piscataway, New Jersey, USA, pp 98–103
- Corbley KP (1999) Radar Imagery Proves Valuable in Managing and Analyzing Floods Red River flood demonstrates operational capabilities. Earth Observation Magazine, vol. 8, num. 10. Accessible via http://www.eonline.com/Common/Archives/1999dec/99dec_corbley.html Accessed 10 Sept 2008
- DEGREE – Earth Science White Paper on Grids (2008) Accessible via http://www.eu-degree.eu/DEGREE/internal-section/wp6/DEGREE-D6.1.2_v2.8.pdf Accessed 10 Sept 2008
- Feller M, Foster I, Martin S (2007) GT4 GRAM: A Functionality and Performance Study. Accessible via <http://www.globus.org/alliance/publications/papers/TG07-GRAM-comparison-final.pdf> Accessed 10 Sept 2008
- Foster I (2002) The grid: a new infrastructure for 21st century science. Phys Today 55(2):42–47
- Foster I (2005) Globus Toolkit Version 4: Software for Service-Oriented Systems. In: IFIP International Conference on Network and Parallel Computing, Springer-Verlag LNCS 3779, pp 2–13
- Fusco L, Goncalves P, Julian L, Fulcoli M, Terracina A, D'Acunzo G (2003) Putting Earth-Observation Applications on the Grid. ESA Bulletin 114:86–90
- Fusco L, Cossu R, Retscher C (2007) Open grid services for envisat and earth observation applications. In: Plaza AJ, Chang C-I (eds) High performance computing in remote sensing, 1st edn. Taylor & Francis, New York, pp 237–280
- GEOSS, Global Earth Observation System of Systems (2005) 10-Year Implementation Plan: reference document. ESA Publication Division, Netherlands
- GMES, Global Monitoring for Environment and Security (2004) Establishing a GMES capacity by 2008 - (Action Plan (2004–2008)). Communication from the Commission to the European Parliament and the Council, Brussels, COM (2004) 65 final
- Hansen AJ, Rotella JJ (1999) Abiotic factors. In: Hunter ML (ed) Maintaining biodiversity in forest ecosystems. Cambridge University Press, Cambridge, pp 161–209
- Horritt MS (2006) A methodology for the validation of uncertain flood inundation models. J of Hydrology 326:153–165
- Humphrey M, Wasson G, Jackson K, Boverhof J, Rodriguez M, Bester J, Gawor J, Lang S, Foster I, Meder S, Pickles S, McKeown M (2005) State and events for web services: a comparison of five WS-resource framework and WS-notification implementations. In: Proc 4th IEEE Int Symp on High Performance Distributed Computing (HPDC-14), Research Triangle Park, NC
- King MD, Closs J, Spangler S, Greenstone R, Wharton S, Myers M (2004) EOS data products handbook (Vol 1). NASA Goddard Space Flight Center, Greenbelt
- Kogan F, Stark R, Gitelson A, Adar E, Jargalsaikhan L, Dugrajav C, Tsooj S (2004) Derivation of pasture biomass in Mongolia from AVHRR-based vegetation health indices. Int J Remote Sens 25 (14):2889–2896
- Kopp P, Petiteville I, Shelestov A, Li G (2007) Wide Area Grid (WAG). In: Proc of the 7th Ukrainian Conf on Space Research (National Flight and Control Center, Evpatoria, Ukraine) p 209
- Kussul N, Shelestov A, Korbakov M, Kravchenko O, Skakun S, Ilin M (2008a) Grid infrastructure for satellite data processing in Ukraine. Int J Inf Techn and Knowledge 2(1):69–76
- Kussul N, Shelestov A, Skakun S (2008b) Grid system for flood extent extraction from satellite images. Earth Science Informatics 1(3–4):105–117
- Larman C (2004) Applying UML and patterns: An introduction to object-oriented analysis and design and iterative development, 3rd edn. Prentice Hall PTR, Upper Saddle River, NJ
- Liang S (2004) Quantitative remote sensing of land surfaces. Wiley

- Mandl D, Frye SW, Goldberg MD, Habib S, Talabac S (2006) Sensor Webs: Where they are today and what are the future needs? In: Proc Second IEEE Workshop on Dependability and Security in Sensor Networks and Systems (DSSNS 2006), pp 65–70. doi:10.1109/DSSNS.2006.16
- Michalakes J, Dudhia J, Gill D, Henderson T, Klemp J, Skamarock W, Wang W (2004) The weather research and forecast model: software architecture and performance. In: Proc of the 11th ECMWF Workshop on the Use of High Performance Computing In Meteorology (25–29 October 2004, Reading U.K) Accessible via http://www.wrf-model.org/wrfadmin/docs/ecmwf_2004.pdf Accessed 10 Sept 2008
- Moe K, Smith S, Prescott G, Sherwood R (2008) Sensor Web Technologies for NASA Earth Science. In: Proc of 2008 IEEE Aerospace Conference, pp 1–7. doi:10.1109/AERO.2008.4526458
- Popov M, Kussul N, Stankevich S, Kozlova A, Shelestov A, Kravchenko O, Korbakov M, Skakun S (2008) Web service for biodiversity estimation using remote sensing data. *Int J Digital Earth* 1(4):367–376
- Shelestov A, Kussul N, Skakun S (2006) Grid technologies in monitoring systems based on satellite data. *J of Automation and Inf Sci* 38(3):69–80
- Wagner W, Pathe C, Sabel D, Bartsch A, Kuenzer C, Scipal K (2007) Experimental 1 km soil moisture products from ENVISAT ASAR for Southern Africa. In: Proc of ENVISAT Symposium 2007, Montreux, Switzerland, SP-636
- Welch V, Siebenlist F, Foster I., Bresnahan J, Czajkowski K, Gawor J, Kesselman C, Meder S, Pearlman L, Tuecke S (2003) Security for grid services. In: Proc 12th IEEE Int Symp on High Performance Distributed Computing, pp 48–57