2

High-performance Intelligent Computations for Environmental and Disaster Monitoring

2.1 Specifics of Earth observation problems

At present, global climate changes on the Earth made a rational land use, environmental monitoring, prediction of natural and technological disasters, etc the tasks of great importance. The basis for the solution of these crucial problems lies in the integrated use of data of different nature: modeling data, in-situ measurements and observations, and indirect observations such as airborne and space borne remote sensing data [GEOSS, 2010].

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 predict different processes occurring in the atmosphere, land, ocean and sea, etc; they can 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 predictions of natural disasters, environmental monitoring, etc. 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 to rescuers during flooding [Corbley, 1999]. Both space borne microwave and optical data can provide means to detect drought conditions, estimate drought extent and assess the damage 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 etc. [Liang, 2004].

The EO domain is characterized by the large volumes of data that should be processed, catalogued, and archived [Fusco et al, 2003], [Shelestov et al, 2006]. For example, 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, etc. For example, calibration and mosaic composition of 80 images generated by ASAR instrument onboard Envisat satellite takes 3

days on 10 workstations of Earth Science GRID on Demand that is being developed in ESA and ESRIN [Fusco et al, 2003]. 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" and the International Federation of Red Cross.

It should be also noted that the same EO data sets and derived products could be used for a number of applications. For example, information on land use/change, soil properties, meteorological conditions etc. is both important for floods and droughts 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, 2010] and European initiative Global Monitoring for Environment and Security [GMES, 2010]. Services and models that are common for different EO applications (e.g. flood monitoring and crop yield prediction) are shown in Figure 25.

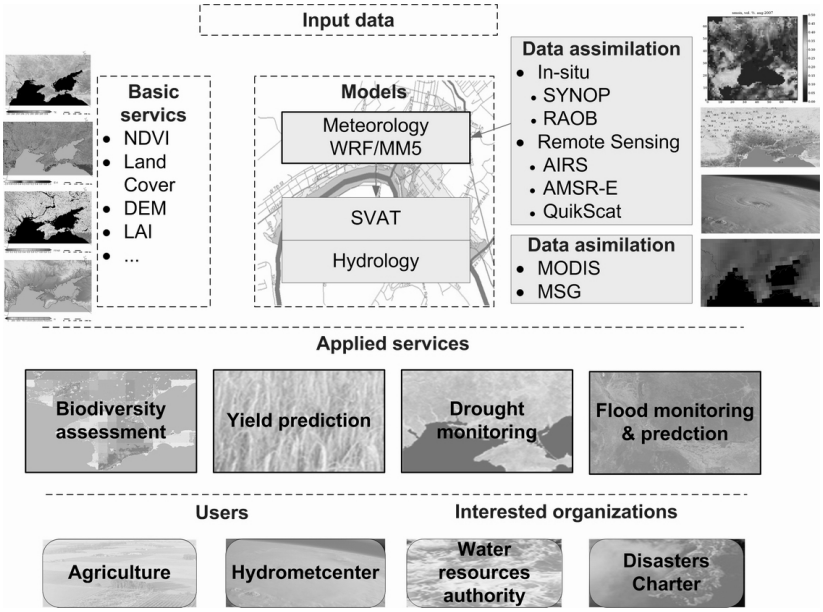

Figure 25. Common services and models for a variety of applications

A considerable need therefore exists for intelligent methods an appropriate infrastructure that will enable the integrated and operational use of multi-source data for different applications domain. From technological point of view, Grids can provide solutions to the above-mentioned problems [Foster and Kesselman, 2004], [Fusco et al, 2003], [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, 2003].

In this part, we focus on the description of the Grid infrastructure that is under the development in 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 vegetation state assessment. We also review issues regarding the integration of the Sensor Web and Grid technologies for flood applications.

2.2 Existing tendencies and initiatives

2.2.1 Challenges

Increasing numbers of natural disasters have demonstrated to the humanity the paramount importance of the natural hazards topic for the protection of the environment and the citizens. Climate change is likely to increase the intensity of rainstorms, river floods, droughts and other extreme weather events. All these problems are among benefit areas of GEOSS (Global Earth Observation System of Systems) aiming to integrate efforts of different countries for exploiting the growing potential of Earth observations to support decision making in an increasingly complex and environmentally stressed world.

Floods. Floods are among the most devastating natural hazards in the world, affecting more people and causing more property damage than any other natural phenomena [CEOSDMSG, 2001]. In the period of time between 1900 and 2006 a total of 415 major flood events occurred in Europe alone, with an average death toll of 22 and 35159 affected people (Source: EM-DAT: The OFDA/CRED International Disaster Database, August 2006).

Ukraine is vulnerable to floods, in particular in the Carpathian region where it occurs almost every year. During the floods in 2001, 9 people were killed and 12000 citizens were evacuated, more than 1500 buildings were destroyed, and more than 30000 buildings were flooded.

In January-February 2008 heavy flooding were affecting a number of southern Africa countries including Mozambique, Zimbabwe, and Zambia. Media provisionally reported that almost 7,000 households have reached resettlement centers and more than 30,000 hectares of crops have been lost (Source: International Charter "Space and Major Disasters").

Drought: In spring-summer, 2007, southern regions of Ukraine were heavily affected by droughts. As a consequence, crops area of approximately 1,4 million ha were totally destroyed, and 8,5 million ha of crops were damaged. Due to this severe drought, financial losses for Ukraine were approximately 100 million of U.S. dollars. The event was considered as a disaster of national level (source: Ministry of Emergent situations of Ukraine).

EU countries are also constantly hit by severe drought. In the last thirty years, EU has been affected by major droughts, in particular in 1989, 1990, 1991, and 2003. The overall impact of droughts in the last thirty years is estimated to 100 000 Million Euro.

Vegetation state assessment: In the following 2 years, Ukraine is planning to launch its own remote sensing satellite Sich-2. One of the applied problems that will be solved using data acquired from Sich's instruments is vegetation health assessment in support of agriculture. Thus, it is needed to develop appropriate method for vegetation state assessment using space-borne remote sensing data.

Computational complexity: It should be stated that efficient monitoring of agricultural resources and natural disasters is almost impossible without the use of Earth Observation (EO) data from space. Satellite observations enable acquisition of data for large and hard-to-reach territories; can provide continuous measurements and human-independent information, etc.

In turn, the EO domain is characterized by large volumes of data that should be processed, catalogued, and archived. For example, GOME instrument onboard Envisat satellite generates nearly 400 Tb data per year. Space Research Institute NASU-NSAU beginning from 2006 have installed EUMETCast system for environmental data dissemination. EUMETCast that is part of global

GEONETCast system of GEOSS enables acquisition of more than 50 Tb of processed and unprocessed information per year. Moreover, the processing of satellite data is carried out not by the single application with monolithic code, but by distributed applications. This process can be viewed as complex workflow that is composed of many tasks: geometric and radiometric calibration, filtration, reprojection, composites construction, classification, products development, post-processing, visualization, etc. For example, calibration and mosaic composition of 80 images generated by ASAR instrument onboard Envisat satellite takes 3 days on 10 workstations of Earth Science GRID on Demand that is being developed in ESA and ESRIN.

2.2.2 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, 2010].

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, 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.

2.2.3 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 hours-a-day.

Based on the gained experience, the European Space Agency (ESA) and the European Space Research Institute (ESRIN) have focused on the development of Earth Observation Grid Processing on-Demand infrastructure (G-POD) [Fusco et al, 2003]. Grid is considered as a comfortable "open platform" for handling computing resources, data, tools, etc., and not limited to only high performing 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, etc. One of the most important applications is the analysis long-term data. For example, the analysis of 8 years of GOME on-board temperatures (overall 525 Gb of data) took less than 2 days on 40 computer elements of ESRIN "Grid-on-demand" structure (overall 38460 files were processed). At present, G-POD

infrastructure consists of more than 150 working nodes with ability to store and handle of about 100 Tb of data.

DEGREE (Dissemination and Exploitation of GRids in Earth science) 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, etc. The following applied problems are within the scope of DEGREE: earthquake analysis, floods modeling and forecasting influence of climate changes on agriculture, etc.

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 disasters monitoring system (http://dmss.tksc.jaxa.jp/sentinel).

The Wide Area Grid (WAG) project is initiated by the CEOS Working Group on Information Systems and Services (WGISS), 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".

In this paper we present different approaches to multi-source data integration for the solution of complex applied problems, in particular flood mapping and vegetation state estimation using satellite, modeling and in-situ data. Since these applications are data- and computation-intensive, we use Grid computing technologies. In such a case computational and informational resources are geographically distributed and may belong to different organizations. For this purpose, we also investigate benefits and approaches to the integration of satellite-based monitoring systems.

2.2.4 Scientific approaches

There are some methods for solving of aforementioned problems. Below we examine them in details.

Flood monitoring and prediction: Hydrologic and hydrodynamic models play a major role in assessing and forecasting flood risk. Model's predictions of potential flood extent can help emergency managers to develop contingency plans well in advance of an actual event to help facilitate a more efficient and effective response. One of the main stages of flood prediction is runoff-rainfall simulation. Traditionally for this stage, lumped or semi-distributed hydrological models were used (for instance HSPF model [Singh, 1995]. In general, such models have several parameters that are subject for calibration using input-output time series and/or expert's knowledge. The modern way of hydrological prediction is application of distributed physically based models, e.g. TOPKAPI model [Liu and Todini, 2002]. Such models are better suited for representing

heterogeneous hydrological features and for using gridded meteorological data available from modern regional Numerical Weather Prediction models.

These models require several types of data as input, such as rainfall amount/intensity, water extent, land use, soil type and moisture, Digital Elevation Models (DEM), etc. Complex terrain and land use in many regions result in a requirement for high spatial resolution data over very large areas, which can only be practically obtained by remote sensing systems.

Remote sensing data are widely used for flood extent extraction, since it is impractical to acquire the flood area through field observations. Flood extent can be used for hydraulic models to reconstruct what happened during the flood and determine what caused the water to go where it did, for damage assessment and risk management, and can benefit to rescuers during flooding. In order to extract flood extent from satellite imagery we can use data in both optical and microwave range of electromagnetic emission.

The flood extent maps using optical sensors can be extracted using information provided in visible and infrared channels. Different vegetation indices, such as NDVI (Normalized Difference Vegetation Index), could also be used for these purposes. However, the use of optical imagery is limited by severe weather conditions, in particular clouds.

In turn, SAR (synthetic aperture radar) image acquisition is independent of daytime and weather conditions. The use of SAR data for flood extent mapping is motivated by the fact that smooth water surface provides no return to antenna in microwave spectrum and appears black in SAR imagery. Existing methods for flood extent mapping are based on the use of multitemporal technique (http://earth.esa.int/ew/floods/), pixel-processing methods with threshold [Cunjian et al, 2001], [De Chiara et al, 2006]. The authors of the project have developed neural network method for flood extent extraction [Kussul et al, 2007] that is based on image segmentation with sliding window.

Therefore, there exist sophisticated methods for flood extent extraction from satellite imagery. However, in order to provide comprehensive system for flood monitoring and forecasting one need to integrate data different from different sources: modeling, satellite, and in-situ measurements.

Drought monitoring. Both radar and optical data can provide means to detect drought condition, estimate drought extent and assess damage caused by drought events. The temporal resolution of current low resolution optical data as well as wide swath low and medium resolution radar data is enough to monitor vegetation condition and is close to be enough to monitor moisture changes. For instance, MERIS and MODIS data are available once per day for middle latitudes and ASAR WS/GM data are available with temporal resolution up to 2 images per week.

Estimation of drought condition using radar data is possible due to radar's sensitivity to soil/vegetation moisture content. However, the complete decoupling soil and vegetation scattering effects is hard using current C-band single polarization wide swath data. Due to this drought monitoring using radar data can be provided using time series analysis of ASAR WS/GM backscatters [Wagner et al, 2007].

Drought monitor can be done using optical data, e.g. MERIS and MODIS VIS/NIR to create vegetation indices, MODIS TIR to monitor surface temperature. Methods for drought monitoring using NOAA AVHRR data were developed in [Kogan et al, 2004].

Drought monitoring will benefit from merging both optical and microwave remote-sensing data with comprehensive Land Surface Models driven by meteorological data from regional Numerical Weather Prediction models.

Vegetation state assessment: To assess vegetation health/stress the derivation of several biophysical and biochemical variables from optical remote-sensing data was considered by scientific community. Such variables include pigment concentration (e.g. chlorophyll a+b), leaf structure, dry matter content (e.g. lignin, cellulose, protein), water content at leaf level and leaf area index (LAI), leaf angle distribution (LAD), fraction of photosynthetically active radiation absorbed by vegetation (FPAR) at canopy level [Liang, 2004].

Roughly, two main approaches were investigated. The first is empirical or physically based derivation of biophysical parameters from so-called spectral indexes. For instance, relations between LAI and Normalized Difference Vegetation Index (NDVI) and between reflectance in NIR/SWIR domain with vegetation water content were established [Carlson and Ripley, 1998], [Gao, 1996] [Ceccato et al, 2002].

The second approach consists in inversion of physically based leaf, canopy, and atmosphere models. These models were used to estimate structural parameters as LAI and FPAR [Knyazikhin et al, 1998] and biophysical variables such as water content [Zarco-Tejada et al, 2003].

As a conclusion, existing methods for solution of aforementioned applied problems are quite fragmentary and designed to work with some particular sensor data. Significant progress in monitoring of floods, drought, and vegetation's state can be achieved through simultaneous use of data from different sensors, in-situ observations, and modeling approach. Moreover, the same data and core services (e.g. land cover/land use) could be used as inputs for different applications. For instance, regional Numerical Weather Prediction models are valuable for both drought and flood monitoring, as well as assessment of vegetation's state and drought conditions will benefit from using Land Surface Models. Grid technologies will provide the platform for development of such methods and deployment of core and applied services.

2.3 Data assimilation approach

As we can see from overview, there is an urgent need for operational services solving environment-monitoring problems using heterogeneous data. Such approach is implemented within GMES program. Within Ukrainian segment of GEOSS/GMES, we develop new methods for integration of data of different nature (in-situ measurements, modeling, and remote sensing) and Grid technologies for their implementation and data visualization. The particular tasks that are solved are as follows:

1. Development of new method for data integration, in particular remote sensing data from space, in-situ measurements, and modeling data
2. Development of Grid-technologies for heterogeneous data integration
3. Application of developed methods to agricultural and natural disaster monitoring

The overall flowchart of Ukrainian segment of GEOSS/GMES (models, methods, information flows) is depicted in the Figure 26. In this scheme, filled rectangles represent models, methods, and processes. Rounded dotted rectangles represent input data for models and methods or results of previous processing, while rounded rectangles with solid line show end-users.

We consider given blocks in details.

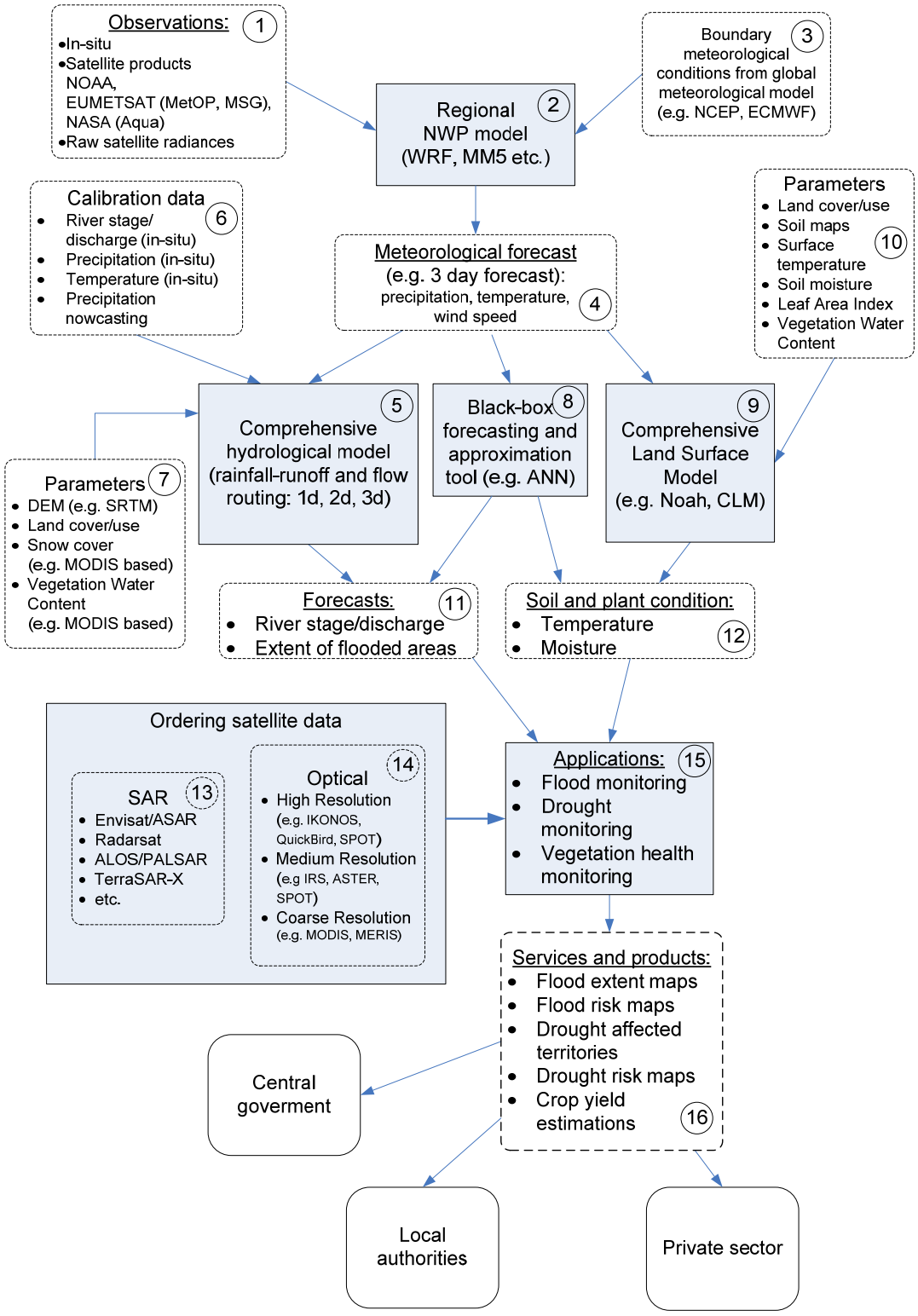

Figure 26. Overall data flowchart at Ukrainian segment of GEOSS/GMES

✓ Meteorological observations

Within Ukrainian segment of GEOSS/GMES, it is planned to create data assimilation system for regional Numerical Weather Prediction model (NWP). In particular, the following satellite data will be assimilated:

— NOAA (microwave & IR instruments: AMSU, MHS, HIRS etc.);

— EUMETSAT (MetOP instruments, MSG geostationary data);

— NASA (Aqua satellite: AMSR-E, AIRS data).

This will increase the accuracy of regional meteorological forecasts and as consequence will increase the quality of solving of several applied problems namely flood monitoring and forecasting, drought monitoring and vegetation health monitoring.

✓ **Regional NWP model**

Because of previous joint STCU-NASU project "GRID technologies for environmental monitoring using satellite data" (2005-2007), we have adapted regional NWP model for the territory of Ukraine. Now SRI runs Weather Research&Forecasting NWP model in operational mode for the territory of Ukraine (see http://dos.ikd.kiev.ua/index.php?option=com_wrf).

✓ **Boundary meteorological conditions from global meteorological model**

To create regional NWP forecasts it is necessary to obtain forecast frames from global meteorological models. This data is used to specify boundary conditions of regional model (vertical profiles of wind, temperature, humidity, pressure etc.). Currently global NCEP forecasts using GFS model are easily available in Internet. In this project, global forecasts will be obtained using NOAA NOMADS system (http://nomad5.ncep.noaa.gov/).

✓ **Meteorological forecasts**

Meteorological forecasts provided by NWP model will be used as inputs for hydrological models, Land Surface Models and additionally they will be used for initiation of retrieving additional datasets in case of possible natural disasters (blocks #13 and #14). For example, using 3 day forecast we could order satellite images at least in 3 days in advance before actual flood event occurs. Without using of forecasts to order satellite data, timely will be much more complex because satellite operators need some time to reprogram satellite and make changes in queue schedule and priorities. Depending on satellite, such procedure can take from few hours to several days. Additionally it should be noted that acceleration of retrieving of satellite data will significantly increases the price of the images.

✓ **Comprehensive hydrological model**

In Ukraine, a detailed hydrological model for river Tisza basins was deployed. Using this model in the framework of the project, we can obtain detailed forecasts for flood dynamics, water levels, and flooded areas. Other basins require adaptation of existing models using additional in-situ observation for calibration.

✓ **Calibration data**

These data are required to estimate parameters in both physically based and empirical black-box hydrological models. Project members have access to such data for Tisza river basin (Carpathian mountain region of Ukraine). For other regions, data from local authorities is required. Additionally

we can use satellite based rainfall estimations, for instance, using geostationary data (e.g. MSG or GOES).

- ✓ **Parameters of hydrological models**

Some of these parameters can be estimated using remote-sensing data. For instance, we will use digital elevation model (DEM) data from SRTM project that have spatial resolution of 90m and available for free while buying more precise DEM as needed. Land cover/land use maps will be obtained for local watershed using medium/coarse resolution imagery. Additionally, some state variables of hydrological models can be estimated from optical/radar data (for instance, MODIS snow cover or LAI (Leaf Area Index) products).

- ✓ **Black-box forecasting and approximation tool**

Such tools that are based on Artificial Neural Networks (ANN) can be used as additional source of information complementary to comprehensive environmental models. For instance, such ANN-based models are useful for rainfall-runoff simulations. In addition, this tool will be used for fusion of model-based, in-situ and satellite-based retrievals of environmental parameters (temperature, moisture, etc.). For these purposes, modular neural networks can be applied. Modular NN combine different modules, which can be neural networks with various parameters in order to exploit advantages of modules and to improve the global performance.

- ✓ **Comprehensive Land Surface Model (LSM)**

Such model describes interactions between atmosphere, soil and vegetation including such processes as infiltration, evapotranspiration, soil moisture and heat transport. With appropriate meteorological forcing and soil/land cover data; these models are capable to predict soil temperature and moisture profiles, surface temperature, plant water content, snowpack etc. LSMs are commonly included into meteorological models to provide bottom boundary conditions (for instance, Noah LSM within WRF, MM5 or NCEP Global Forecast System models, Community Land Model (CLM) within Community Climate System Model) but can be run in so-called off-line mode (decoupled from meteorological model). In the latter case, LSM can be run with high spatial resolution (up to 1 km). SRI has experience to operate Noah LSM in coupled mode within WRF modeling system. In the framework of the project, we will develop method for assimilation of satellite data into such models. The results of assimilation will be used for creation of several products (block #12) that will become the basis for end-user services (block #16), in particular for drought indicators and vegetation stress estimations.

- ✓ **Parameters of LSM**

As in the case of hydrological models used for flood prediction, several LSM's parameters can be estimated using remote-sensing data. Within the proposed project, we will use land cover/land use maps, leaf area index, surface temperature, soil moisture. These data will be assimilated into LSM using intelligent techniques (in particular by ANN) and evolutionary computations (e.g. genetic algorithms).

✓ Hydrological forecasts

This block provides results of hydrological model. Ideally, we would like to have forecasts of flooded areas. Worse case if we have only river stage/discharge data. In the case of flood, we can issue alert message to the local authorities and order remote sensing data in advance to estimate flooded areas during the flood (blocks #13 and #14).

✓ Soil and plant condition

Soil and plant condition are obtained as a merge of remote-sensing retrievals, in-situ data and results of modeling of land surface. Such data will be used to produce dedicated products in the field of drought and plant condition monitoring (block #16).

✓ SAR (synthetic aperture radar) data from space-borne instruments

SAR imagery are most valuable satellite data for estimation of flooded areas due to all-weather SAR functioning. At first stage, we propose to use Envisat/ASAR (ESA) and ALOS/PALSAR (JAXA) data. These sensors are included into International Charter "Space and Major Disasters" (http://www.disasterscharter.org/main_e.html). In addition, these data can be obtained via ESA Cat-1 projects "Wide Area Grid Testbed for Flood Monitoring using Spaceborne SAR and Optical Data" (#4181) in which SRI takes part.

✓ Optical data

Taking into account possible cloud cover problems we can use optical imagery for flood extent estimation as well as to assess state of vegetation. Within this project, coarse resolution sensors such as MODIS from Terra satellite or MERIS from Envisat satellite will be used. Medium/fine resolution imagery can be ordered to produce local/regional products and services.

✓ Applications

Within Ukrainian segment of GEOSS/GMES, we focus on the following applications:

– flood monitoring;

– drought monitoring;

– vegetation state monitoring.

✓ Services and products

The following products and services are provided for the central government, local authorities and private sector:

– flood extent maps;

– flood risk maps;

– areas affected by drought;

– drought risk maps;

– crop yield estimation for agricultural regions.

2.4 Applications

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

(i) numerical weather prediction belongs to computational intensive applications;

(ii) flood applications need the fast response to the emergencies, and thus require a reliable infrastructure for data management and processing;

(iii) vegetation state estimation belongs to data intensive application where different data and products are analyzed in order to produce the final product and requires intelligent data assimilation techniques.

Prediction of meteorological parameters represents one of the core services for a number of applications (e.g. floods, droughts, agriculture, etc). Currently, we run the Weather Research and Forecasting model (WRF) (Michalakes et al. 2004) in operational mode for the territory of Ukraine. The meteorological forecasts are generated every 6 hours with a spatial resolution of 10 km. Forecast range is 72 hours. The horizontal grid dimensions are 200x200 points with 31 vertical levels. We use NCEP GFS (Global Forecasting System) forecasts as boundary conditions. This data is available via Internet though the NOMADS system (National Operational Model Archive & Distribution System).

The workflow of the model run is composed of the following steps (Figure 27):

(i) data acquisition;

(ii) data pre-processing, computation of forecasts using WRF model and data post-processing;

(iii) visualization of the predicted parameters.

Data acquisition: To run WRF model, it is necessary to obtain boundary and initial conditions for territory of Ukraine. This data can be extracted from GFS model forecasts. To get the required data, the dedicated script was developed. This script downloads global forecasts every 6 hours. 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 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 vinderp.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).

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

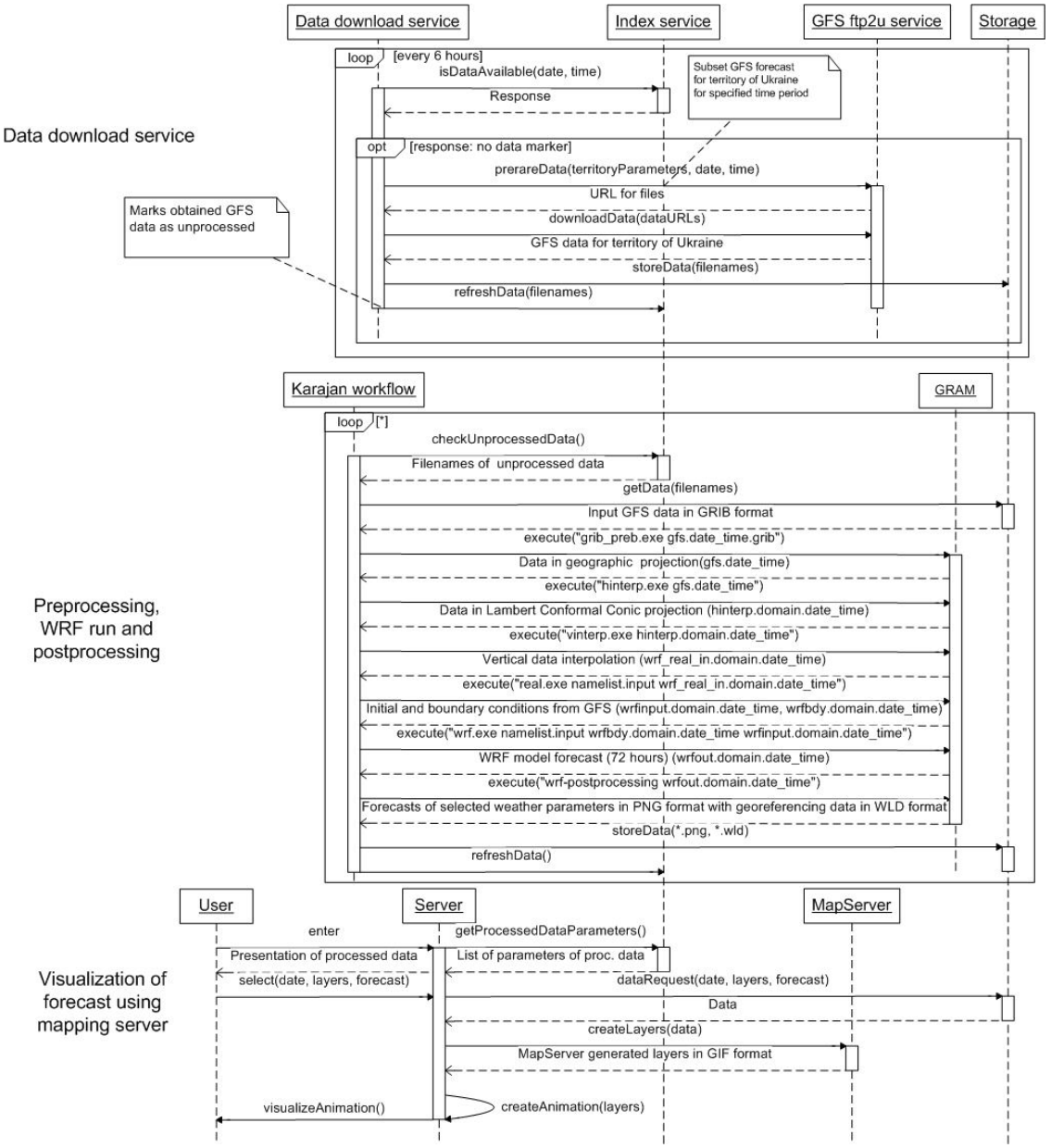

Figure 27. UML sequence diagram [Larman, 2004] for the NWP application

After WRF model run, **post-processing step** is carried out. For specified weather parameters and for each forecast frame (3 hours), 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 forecasts via the mapping service. This service is available via http://dos.ikd.kiev.ua, and provides to the users animations of the weather forecasts (Figure 28).

The service provides tools to select a forecast time, forecast frames (up to 72 hours ahead), and weather parameters to display. 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 epy storage subsystem and passed to epy 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.

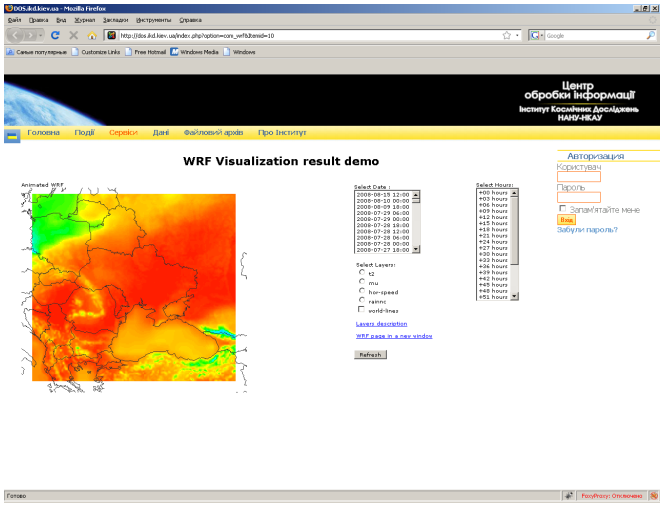

Figure 28. The example of land temperature forecasts using WRF model

We have also tested the performance of the WRF model in dependence of the number of computational nodes. For test purposes, we used the WRF model version 2.2 with a model domain identical to those used in operational NWP service (200x200x31 grid points with horizontal spatial resolution 10 km). We observed almost linear productivity growth within increasing number of computation nodes. For instance, 8 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 Figure 29).

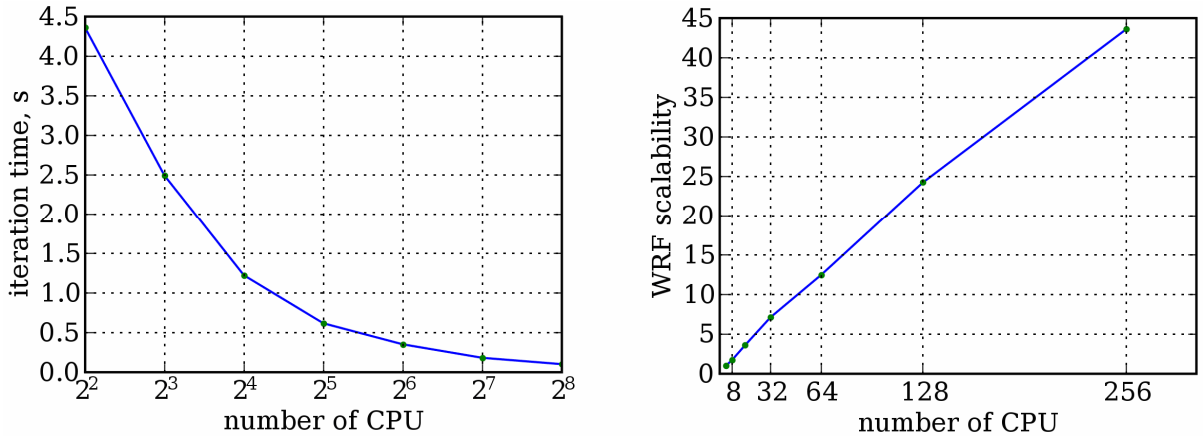

Figure 29. The results of WRF performance on the SCIT-3 cluster:
computation time for 1 iteration (left);
acceleration of the WRF model with respect to a number of nodes (right)

2.4.1 Flood prediction and mapping from satellite imagery

In recent decades, the number of hydrological natural disasters has considerably increased. According to [Scheuren et al, 2008], we have witnessed in recent years a strengthening of the upward trend, with an average annual growth rate of 8.4% in the 2000 to 2007 period. Hydrological disasters, such as floods, wet mass movements, represent 55% of the overall disasters reported in 2007, having a tremendously high human impact (177 million victims) and causing high economic damages (24.5 billion USD) [Scheuren et al, 2008].

EO data from space can provide valuable and timely information when one has to respond to and mitigate such emergencies as floods. From satellite imagery, we can determine flood areas, since it is impractical to provide such information through field observations. The use of optical imagery (in visible and infrared range) for flood mapping is limited by severe weather conditions, in particular by the presence of clouds. In turn, synthetic aperture radar (SAR) measurements from space are independent of daytime and weather conditions and can provide valuable information to monitoring of flood events. This is mainly due to the fact that smooth water surface provides no return to antenna in microwave spectrum and appears black in SAR imagery [Rees, 2001].

Flood mapping procedure from SAR imagery represents a complex workflow and consists of the following steps. The first step consists in re-constructing a satellite imagery taking into account the calibration, the terrain distortion using digital elevation model (DEM) and providing exact geographical coordinates. The second step is image segmentation, and the third step consists in the classification to determine the flood extent.

In this subsection we describe a neural network approach to flood mapping from satellite SAR imagery that is based on the application of self-organizing Kohonen's maps (SOMs) [Kohonen, 1995], [Haykin, 1999]. The advantage of using SOMs is that they provide effective software tool for the visualization of high-dimensional data, automatically discover of statistically salient features of pattern vectors in data set, and can find clusters in training data pattern space, which can be used to classify new patterns [Kohonen, 1995]. We applied our approach to the processing of data acquired from different satellite SAR instruments (ERS-2/SAR, ENVISAT/ASAR, RADARSAT-1 and RADARSAT-2) for different flood events: river Tisza, Ukraine and Hungary (2001); river Huaihe, China (2007); river Mekong, Thailand and Laos (2008); river Koshi, India and Nepal (2008); river Norman, Australia (2009); and river Zambezi, Mozambique (2008) and Zambia (2009).

To this end, different methods and approaches were proposed to flood mapping using satellite imagery:

- multi-temporal technique (http://earth.esa.int/ew/floods);
- threshold segmentation [Cunjian et al, 2001];
- statistical active contour model [Horritt, 1999];
- edge-detection techniques [Niedermeier et al, 2000];
- analysis of time-series of SAR images [Martinez and Le Toan, 2007].

The following shortcomings of the existing approaches can be identified: manual threshold selection and parameters identification; statistical models require a priori knowledge of image statistical properties; application of complex models for noise (speckle) reduction; no spatial neighborhood between pixel is considered. A more detailed description of the existing techniques is given in [Kussul et al, 2008a].

Data set description. We applied our approach to the processing of remote-sensing data acquired from different satellite SAR instruments for different flood events:

- ERS-2/SAR: flood on Tisza river (Ukraine), 2001;
- ENVISAT/ASAR Wide Swath Mode (WSM): river Huaihe, China, 2007; river Zambezi, Mozambique, 2008; river Mekong, Thailand and Laos, 2008; river Koshi, India and Nepal, 2008; Ha Noi City, Vietnam, 2008; river Zambezi, Zambia, 2009;
- RADARSAT-1: river Huaihe, China, 2007;
- RADARSAT-2: river Norman, Queensland, Australia, 2009 (see Figure 30).

Data from European satellites (ERS-2 and ENVISAT) were provided from the ESA Category-1 project "Wide Area Grid Testbed for Flood Monitoring using Spaceborne SAR and Optical Data" (№4181). Data from RADARSAT-1 satellite were provided from the Center of Earth Observation and Digital Earth (China). RADARSAT-2 data were provided by the Canadian Space Agency (CSA) within the GEOSS Architecture Implementation Pilot Phase 2. (AIP-2, www.ogcnetwork.net).

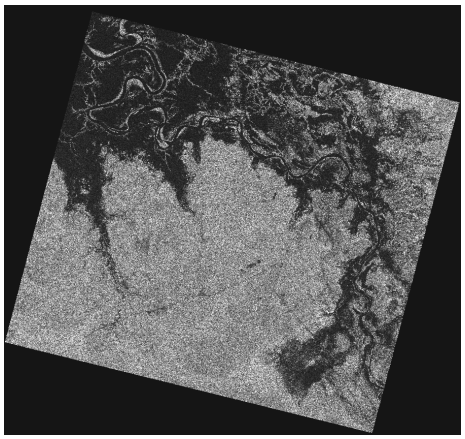

Figure 30. SAR image acquired from RADARSAT-2 satellite during the flood on the river Norman, Australia(14.02.2009)

A pixel size and ground resolution of ERS-2 imagery (in ENVISAT format, SLC – Single Look Complex) were 4 m and 8 m, respectively; for ENVISAT imagery – 75 m and 150 m; and for RADARSAT-1 imagery – 12.5 m and 25 m; for RADARSAT-2 imagery – 3 m both. We used auxiliary data to derive information on water bodies (Landsat-7/ETM+, European Corine Land Cover CLC 2000) and topography (SRTM DEM v.3).

Neural network is built for each SAR instrument separately. In order to train and test neural networks, we manually selected the ground-truth pixels with the use of auxiliary data sets that correspond to both territories with the presence of water (we denote them as belonging to a class "Water") and without water (class "No water"). For ENVISAT/ASAR instrument, data from Chinese flood event were used to construct and calibrate the neural network. This neural network, then, was used to produce flood maps for other flood events. Collected ground-truth data were randomly divided into the training set (which constituted 75% of total amount) and the testing set (25%). Data from the training set were used to train the neural networks, and data from the testing set were used to verify the generalization ability of the neural networks, i.e. the ability to operate on independent, previously unseen data sets [Haykin, 1999].

Methodology description: Our flood mapping workflow with input and output data is shown in Figure 31 [Kussul et al, 2008a].

SOM is a type of artificial neural network that is trained using unsupervised learning to produce a low dimensional (typically two-dimensional), discretized representation of the input space of the training samples, called a map [Kohonen, 1995], [Haykin, 1999]. The map seeks to preserve the topological properties of the input space. SOM is formed of the neurons located on a regular, usually 1- or 2-dimensional grid. Neurons compete with each other in order to pass to the excited state. The output of the map is a, so-called, neuron-winner or best-matching unit (BMU) whose weight vector has the greatest similarity with the input sample **x**.

The network is trained in the following way: weight vectors $\mathbf{w}_j$ from the topological neighborhood of BMU vector i are updated according to [Kohonen, 1995], [Haykin, 1999]

$$i(\mathbf{x}) = \arg\min_{j=\overline{1,L}} \left\| \mathbf{x} - \mathbf{w}_j \right\|,$$

$$\mathbf{w}_j(n+1) = \mathbf{w}_j(n) + \eta(n) h_{j,i(\mathbf{x})}(n)(\mathbf{x} - \mathbf{w}_j(n)), j = \overline{1,L} \quad (1)$$

where η is learning rate (see Eq. 3), $h_{j,i(\mathbf{x})}(n)$ is a neighborhood kernel around the winner unit i, **x** is an input vector, $\|\bullet\|$ means Euclidean metric, L is a number of neurons in the output grid, n denotes a number of iteration in the learning phase.

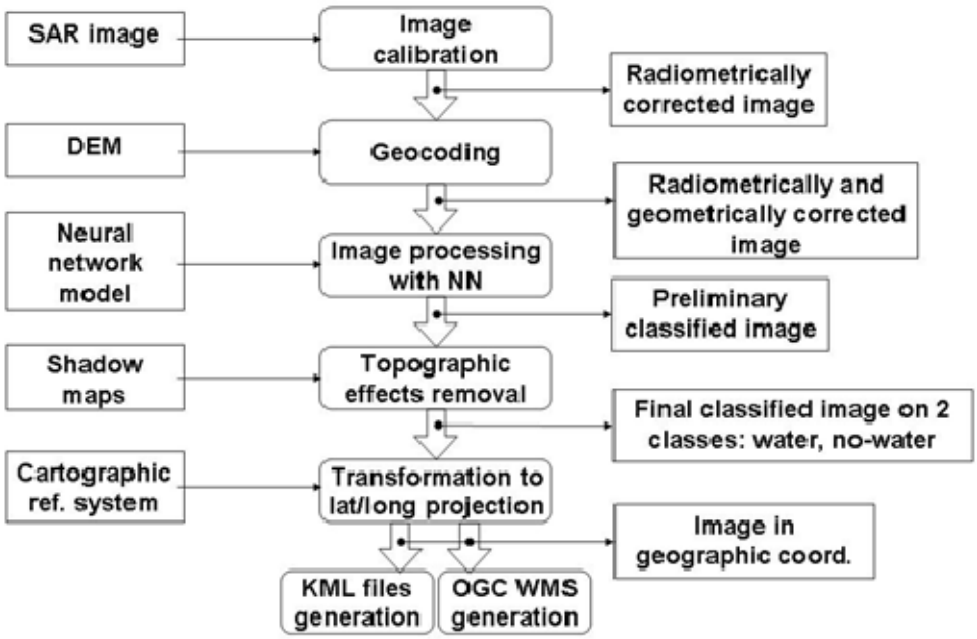

Figure 31. Flood mapping from SAR satellite imagery: workflow

The neighborhood kernel function $h_{j,i(\mathbf{x})}(n)$ is taken to be the Gaussian

$$h_{j,i(\mathbf{x})}(n) = \exp\left(-\frac{\|r_j - r_{i(\mathbf{x})}\|}{2\sigma^2(n)}\right) \tag{2}$$

where $r_j, r_{i(\mathbf{x})}$ are the vector locations in the display grid of the SOM, $\sigma(n)$ corresponds to the width of the neighborhood function, which is decreasing monotonically with the regression steps.

For learning rate, we used the following expression:

$$\eta(n) = \eta_0 \cdot e^{-\frac{n}{\tau}}, \eta_0 = 0.1 \tag{3}$$

where τ is a constant. The initial value of 0.1 for learning rate was found experimentally.

Kohonen's maps are widely applied to the image processing, in particular image segmentation and classification [Kohonen, 1995], [Haykin, 1999]. Prior neural network training, we need to select image features that will be give to the input of neural network. For this purpose, one can choose original pixel values, various filters, Fourier transformation etc. In our approach, we used a moving window with backscatter coefficient values for ERS-2 and ENVISAT images and digital numbers (DNs) for RADARSAT-1/2 image as inputs to neural network. The output of neural network, i.e. neuron-winner, corresponds to the central pixel of moving window. In order to choose appropriate size of the moving window for each satellite sensor, we ran experiments for the following windows size: 3-by-3, 5-by-5, 7-by-7, 9-by-9, and 11-by-11.

We, first, used SOM to segment each SAR image where each pixel of the output image was assigned a number of the neuron in the map. Then, we used pixels from the training set to assign each neuron one of two classes ("Water" or "No water") using the following rule. For each neuron, we calculated a number of pixels from the training set that activated this neuron. If maximum number of these pixels belonged to class "Water", then this neuron was assigned "Water" class. If maximum number of these pixels belonged to class "No water", then this neuron was assigned "No water" class. If neuron was activated by neither of the training pixels, then it was assigned "No data" class.

Results of image processing: In order to choose the best neural network architecture, we ran experiments for each image varying the following parameters: (i) size of the moving window for images that define the number of neurons in the input layer of the neural network; (ii) number of neurons in the output layer, i.e. the sizes of 2-dimensional output grid. Other parameters that were used during the image processing are as follows:

- neighborhood topology is hexagonal;
- neighborhood kernel around the winner unit is the Gaussian function (see Eq. 2);
- initial learning rate is set to 0.1;
- number of the training epochs is equal to 20.

The initial values for the weight vectors are selected as a regular array of vector values that lie on the subspace spanned by the eigenvectors corresponding to the two largest principal components of the input data [Kohonen, 1995].

We applied our approach to determine flood areas from SAR images acquired by the following instruments: ERS-2/SAR, ENVISAT/ASAR, and RADARSAT-1. Classification rates for these sensors using independent testing data sets were 85.40%, 98.52% and 95.99%, respectively.

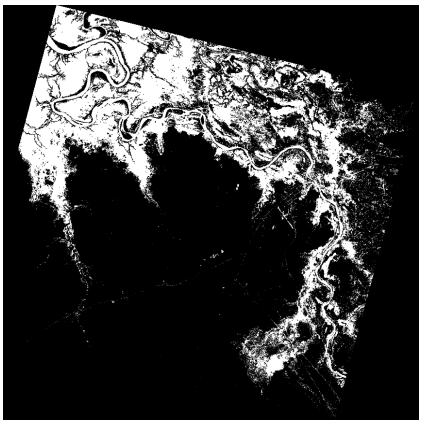

Figure 32. The resulting flood extent shown with white color for the river Norman, Australia (RADARSAT)

For the images with higher spatial resolution (i.e. ERS-2 and RADARSAT-1), the best results were achieved for larger moving window 7-by-7. In turn, for the ENVISAT/ASAR WSM image, we used the moving window of smaller size 3-by-3. The use of higher dimension of input window for the ENVISAT image led to the coarser resolution of the resulting flood extent image and reduced classification rate.

The example of resulting flood extent map derived from RADARSAT-2 data acquired for the river Norman, Australia (see Figure 30) is shown in Figure 32.

Implementation: We developed a parallel version of our method and deployed it at 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 reduce considerably the time required for image processing. In particular, it took approximately 30 min 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.

2.4.2 Vegetation State Estimation

Estimation of vegetation state from satellite data has proved to be very helpful for agriculture monitoring, climate modeling, natural disasters management [Liang, 2008]. Parameters that can be estimated using optical data include Leaf Area Index (LAI), Fraction of Photosynthetic Active Radiation (FPAR), leaf pigment concentration, water concentration. Here, we will focus on plant moisture estimation from satellite data. This is very important for drought monitoring that becomes one of the major disasters in agricultural countries like Ukraine. For example, drought in Ukraine in 2007 resulted in $100 millions losses.

Water shortage in plants and plant stress in general can be detected by optical satellite data. Vegetation moisture determination is possible mainly due to significant differences in reflectance in Shortwave Infrared band of electromagnetic spectrum (SWIR) of vegetation under water stress and under normal conditions. However, in solar optical domain vegetation reflectance is controlled not only by moisture but also by several other factors: leaf structure, pigment concentration, LAI, soil reflectance [Liang, 2004]. Due to this plant moisture, estimation is far from trivial.

This estimation task is a massive parallel problem since estimation has to be performed on the per pixel basis. In addition, even if the problem is not computationally complex for a single pixel, it has to be solved for each pixel of the satellite imagery. For current moderate resolution sensors such as MODIS 1 million pixels has to be processed per day, and new satellite systems such as RapidEye will deliver billions pixels per day. Nevertheless, this problem is highly parallelizable and, thus, is a good candidate to be executed in a Grid environment.

Earlier approaches to vegetation moisture estimation were based on so-called Vegetation Indexes [Ceccato et al, 2002], [Gao, 1996]. Index is a simple combination of reflectance in different bands of satellite image, which has increased sensitivity to target variable like moisture content and low sensitivity to other factors. For example, one of the popular indexes is a Normalized Difference Water Index (NDWI):

$$NDWI = \frac{\rho_{0,8} - \rho_{1,6}}{\rho_{0,8} + \rho_{1,6}} \tag{4}$$

where $\rho_{0,8}$ and $\rho_{1,6}$ are reflectance value in Near Infrared band (NIR) and SWIR band.

Vegetation Indexes uses only a limited number of spectral bands (2-3) while modern sensors like MODIS, MERIS have 7-15 bands. In addition, indexes remain only indirect measures of target variables, and additional regressions have to be used to estimate it. Usually, such regressions require additional calibration using local data, which further complicates utilization of Vegetation Indexes. That is why, at present, the modern way to estimate vegetation parameters is based on more sophisticated approach – physical modeling of satellite signal using canopy radiative transfer models [Liang, 2004].

Problem statement: Under modeling approach, the estimation problem is considered as inverse to the problem of simulation of satellite signal. For the latter task the wide range of models exists [Liang, 2004], among which several models (like PROSPECT [Feret et al, 2008] and SAIL [Verhoef et al, 2007]) are widely used in remote sensing. For our purpose we will formulate radiative transfer model as a mapping $h : \mathbf{R}^{n_x} \to \mathbf{R}^{n_d}$ that maps state of vegetation $x \in X \subset \mathbf{R}^{n_x}$ into reflectance in different bands $h(x) \in D \subset \mathbf{R}^{n_d}$:

$$d = h(x) + h(x)\varepsilon \tag{5}$$

where d is measurement vector and ε is noise vector. This problem is characterized by multiplicative noise [Bacour et al, 2006].

For instance, for PROSPECT leaf radiative transfer model the dimension of x is four $x = (N, C_{ab}, C_w, C_m)^T$, where N — leaf structure parameter, while C_{ab}, C_w, C_m — concentration of chlorophyll, water and dry matter. Dimension of model output vector $h(x)$ is 2100, however for remote sensing purposes model output has to be aggregated to be comparable with current multispectral sensors. So usually the dimension of observation vector d is much smaller, for instance for MODIS sensor it will be 7.

In this chapter, the Bayesian approach to inverse problems is considered [Tarantola, 2005]. Within this approach, uncertainty in a priory estimate of state vector x and in process of measurement of reflectance vector $h(x)$ has probabilistic nature. Let $\mathbf{x}$, $\mathbf{d}$, $\boldsymbol{\varepsilon}$ — random vectors of a priory estimate of model input, observations and noise in observations, $p(x)$, $p(d)$ and $p(\varepsilon)$ — densities of probability distributions of these vectors. It is assumed that random vectors $\mathbf{x}$ and $\boldsymbol{\varepsilon}$ are independent, while densities $p(x)$, $p(\varepsilon)$ and function h is such, that random vectors $\mathbf{x}$ and $\mathbf{d}$ have common density $p(x,d)$ and components of these vectors have variance.

The solution of inverse problem is conditional density of model input $\mathbf{x}$ with respect of known value of observations vector $\mathbf{d}$ [Tarantola, 2005]:

$$p(x \mid d) \propto p(d \mid x) p(x), \quad x \in \mathbf{R}^{n_x}, d \in \mathbf{R}^{n_d} \tag{6}$$

However, for practical purposes we have to estimate some properties of above conditional density, like mean, standard deviation, median, most probable value etc.

Neural network method to solve inverse problem: There are several methods to estimate properties of (6): Monte-Carlo [Qingyuan et al, 2005], variational [Bacour et al, 2002], lookup tables [Combal et al, 2002] and neural networks [Bacour et al, 2006]. However, in recent years neural networks gain a lot of attention due to their ability to approximate arbitrary continuous function and computational efficiency [Haykin, 1999].

To solve inverse problem (6) within traditional neural network approach the approximation $f: D \to X$ of inverse mapping to $h: X \to D$ is constructed using neural network, for instance Multilayer Perceptron (MLP). This is performed through minimization of quadratic functional:

$$J(w) = \frac{1}{2} \sum_i \left\| x_i - f(d_i, w) \right\|^2 \tag{7}$$

where function $f(\cdot, w)$ is defined by neural network with weight coefficients w, $\left\{ (d_i, x_i), i = \overline{1,n} \right\}$ is learning sample set created via sampling from density $p(x,d)$.

It can be shown (see for instance [Bishop, 1996], [Kravchenko, 2009]) that given sufficient number of learning samples neural network with quadratic error criteria will approximate conditional mean $\mathrm{E}[\mathbf{x} \mid \mathbf{d} = d] = \int x p(x \mid d) dd$ of network output x given input d. Therefore, in the framework traditional neural network approach we can obtain only point estimate of parameters. To overcome this deficiency of traditional neural networks for inverse problem solving we propose to apply neural networks with non-quadratic error criteria, such as Mixture Density Networks (MDN) [Bishop, 1996]. Such networks allow modeling of conditional density $p(x \mid d)$ as a mixture of Gaussian densities.

$$p_{MDN}(x \mid d,w) = \sum_{l=1}^{L} \alpha_l(d,w) \cdot \phi\left(x; m_l(d,w), \sigma_l(d,w)\right) \tag{8}$$

where $\phi(x;m,\sigma) = \frac{1}{\left(\sqrt{2\pi}\sigma\right)^{n_x}} \exp\left(-\frac{\|x-m\|^2}{2\sigma^2}\right)$ — Gaussian density with mean m and diagonal covariance matrix $\sigma^2 \mathrm{I}$, α_l — mixture coefficients ($\sum_l \alpha_l = 1$), L — number of elements of mixture. Functions $\alpha_l(d,w)$, $m_l(d,w)$ and $\sigma_l(d,w)$ are constructed using MLP with modified output layer. MDN is learned through minimizing the following error criteria:

$$J(w) = \frac{1}{n} \sum_{i=1}^{n} -\ln p_{MDN}(x_i \mid d_i, w) \tag{9}$$

Unlike MLP, MDN with even one Gaussian component in mixture can approximate both conditional mean and variance of $p(x \mid d)$ [Kravchenko, 2009].

Numerical experiment with PROSPECT model: Here we will demonstrate use of MDN to solve inverse problem of leaf moisture estimation. To formulate forward problem we will use PROSPECT leaf radiative transfer model. In this case x vector consists of 4 parameters: $x = (N, C_{ab}, C_w, C_m)^T$, while observation vector d consists of seven leaf reflectances in MODIS-like spectral bands. To pose inverse problem we will assume uniform a priory density $p(x)$ and independent Gaussian noise model for $\boldsymbol{\varepsilon}$ (5% standard deviation). To estimate plant moisture we will use MDN with 7 neurons in input layer, 5 neurons in hidden layer and one-dimensional mixture containing one Gaussian component. This network is used to estimate mean and variance of conditional density $p(C_w \mid d)$. Increasing number of mixture's components or number of neurons in hidden layer does not improve the quality of solution in this problem.

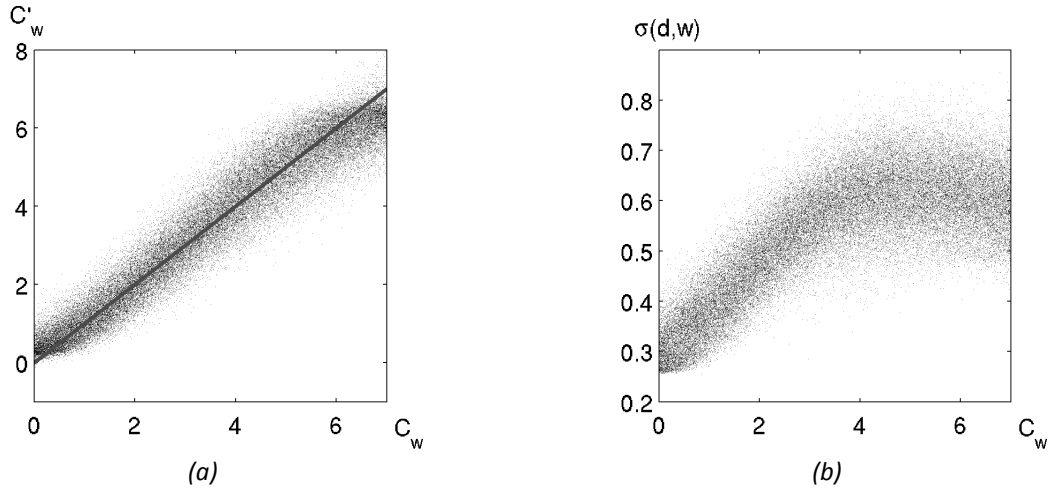

Figure 33. a) scatter plot of estimated leaf moisture $C_w^{'}$ and true C_w;
b) dependency of estimate of standard deviation of leaf moisture $\sigma(d,w)$ w.r.t. real leaf moisture C_w

Scatter plot of conditional mean of leaf moisture $C_w^{'} = m_1(d_i, w)$ estimated by MDN given observation d_i and true value C_w is shown on Figure 33a (identical dependency is shown by strait

line), while dependency of estimate of standard deviation of leaf moisture $\sigma(d_i, w)$ given observation d_i with respect to true value C_w is shown in Figure 33b. Standard deviation is increased with increase of moisture C_w and stabilized for large C_w (4-7 cg/cm^2). This is in accordance with the fact that sensitivity of SWIR reflectance is decreased for large leaf moisture values.

Validation results: To validate our algorithm we used LOPEX leaf optical properties database (Leaf Optical Properties EXperiment). This database contains over 1250 plant reflectance spectra. For validation purpose, 330 fresh leaf spectra of 66 plant species at different moisture level were used. Spectra were aggregated using MODIS band relative spectral response functions. Figure 34a shows the scatter plot of estimated leaf moisture ($C_w^{'}$) and observed (C_w), while Figure 34b shows the histogram of moisture estimation error normalized by estimate of standard deviation $\delta = \left(C_w^{'} - C_w\right) / \sigma(d_i, w)$. Most of the departures (90%) are located in $[-2;2]$ interval (in $\pm 2\sigma$ interval) that confirms adequacy of standard deviation estimates using MDN.

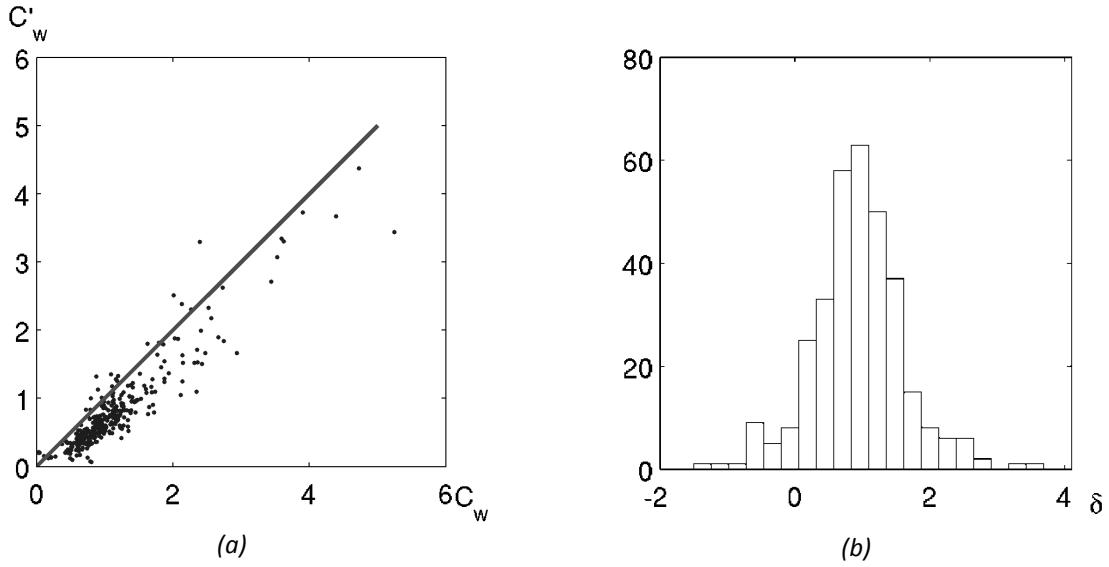

Figure 34. a) scatter plot of estimated leaf moisture $C_w^{'}$ and true C_w ; b) histogram of normalized errors δ

2.5 Levels of integration: main problems and possible solutions

Modern tendencies of globalization and development of the "system of systems" GEOSS lead to the need of integration of heterogeneous satellite-based monitoring systems. Integration can be done at different levels: (i) data exchange level, (ii) task management level. Data exchange is supposed to provide infrastructure for sharing data and products. This infrastructure enables data integration where different entities provide various kinds of data to support joint solution of complex problems (Figure 35). Task management level envisages running applications at distributed computational resources provided by different entities (Figure 36). Since many of the existing satellite monitoring system rely on Grid technologies appropriate approaches and technologies should be evaluated and developed to enable Grid system integration (so called InterGrid).

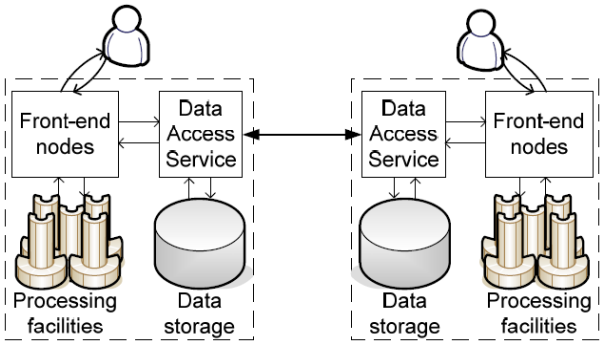

Figure 35. Data integration level

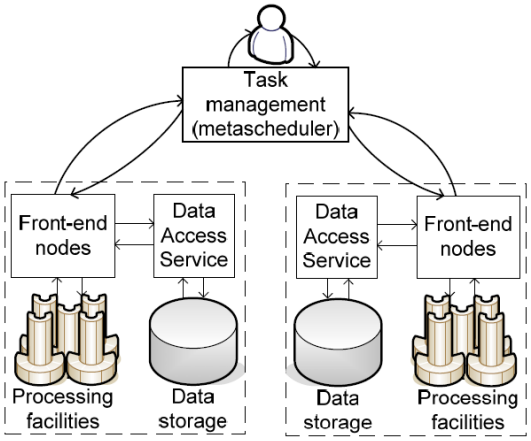

Figure 36. Task management level

This section highlights main challenges and possible solutions for satellite monitoring systems integration at both levels, and provides the case studies for both cases.

Integration at data exchange level could be done by using common standards for EO data exchange, common user interfaces, and common data and metadata catalogues. Considering the task management level, the following problems additionally should be tackled: the use of joint computational infrastructure; development of jobs submission and scheduling algorithms; load monitoring enabling; security policy enforcement.

✓ Data exchange level

At present the most appropriate standards for data integration is Open Geospatial Community (OGC) standards. Data visualization issues can be solved by using the following set of standards: WMS (Web Map Service), SLD (Style Layer Descriptors), and WMC (Web Map Context). OGC's WFS (Web Feature Service) and WCS (Web Coverage Service) standards provide uniform ways for data delivery. In order to provide interoperability at the level of catalogues CSW (Catalogue for Web) standard can be applied.

Since data are stored at geographically, distributed sites there can be issues regarding optimization of visualization schemes. In general, there are two possible ways for distributed data visualization: centralized visualization scheme and distributed visualization scheme. Advantages and faults of each scheme were described in [Shelestov et al, 2008].

This approach is implemented in the International vegetation state estimation system, developed jointly by Space Research Institute NASU-NSAU, Space Research System of Russian Academy of Science and Institute of Informatics of Slovak Academy of Science.

✓ Task management level

In this subsection, we present main issues and possible solutions for Grid-system integration. Main prerequisite of such kind of integration is certificates trust. It could be done, for example, through EGEE infrastructure that nowadays brings together the resources of more than 70 countries. Another problems concerned with different Grid systems integration are as follows: enabling data transfers

and high-level access to geospatial data; development of common catalogues; enabling jobs submission and monitoring; enabling information exchange.

Data transfer: GridFTP is an appropriate and reliable solution for data transfer. The only limitation is the requirement of transparent LAN (local area network) infrastructure.

Access to geospatial data: High-level access to geospatial data can be organized in two possible ways: using pure WSRF services or using OGSA-DAI container. Each of this approach has its own advantages and weaknesses. Basic functionality for WSRF-based services can be easily implemented (with proper tools), packed, and deployed. Nevertheless, advanced functionality such as security delegation, third-party transfers, indexing should be implemented by hands. WSRF-based services can also pose some difficulties if we need to integrate them with other data-oriented software.

OGSA-DAI framework provides uniform interfaces to heterogeneous data. This framework makes possible to create high-level interfaces to data abstracting hiding details of data formats and representation schemas. Most of problems in OGSA-DAI are handled automatically, e.g. delegation, reliable transfer, data flow between different sources and sinks. OGSA-DAI containers are easily extendable and embeddable. Nevertheless, comparing to WSRF basic functionality implementation of OGSA-DAI extensions is more difficult. Moreover, OGSA-DAI require preliminary deployment of additional software components.

Task management: There are two possible approaches for task management. One of them is to use Grid portal (Figure 37) supporting different middleware platforms, such as GT4, gLite, etc. Grid portal is an integrated platform to end-users that enables access to Grid services and resources via standard Web browser. Grid portal solution is easy to deploy and maintain, but it does not provide application interface and scheduling capabilities.

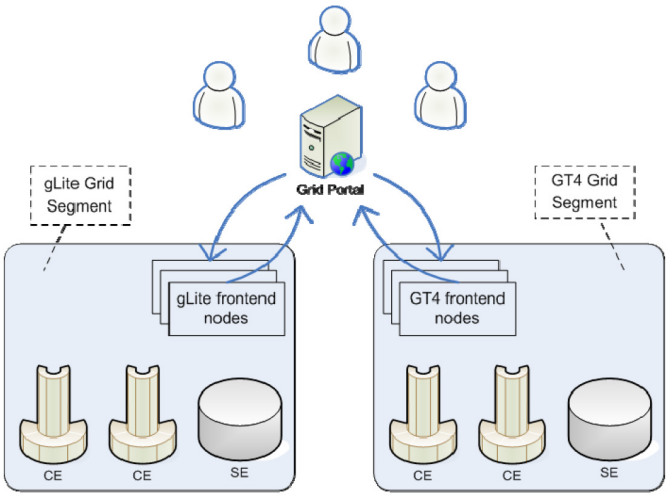

Figure 37. Portal approach to grid system integration

Another approach is to develop high-level Grid scheduler (Figure 38) that will support different middleware by providing some standard interfaces. Such metascheduler interacts with low-level schedulers (used in different Grid systems) enabling in such way system interoperability. Metascheduler approach is much more difficult to maintain comparing to portals; however, it provides API with advanced scheduling and load-balancing capabilities. At present, the most comprehensive implementation for the metascheduler is a GridWay system. The GridWay metascheduler is compatibility with both Globus and gLite middlewares. Starting from Globus Toolkit

v4.0.5 GridWay become standard part of its distribution. GridWay system provides comprehensive documentation for both users and developers that is an important point for implementing new features.

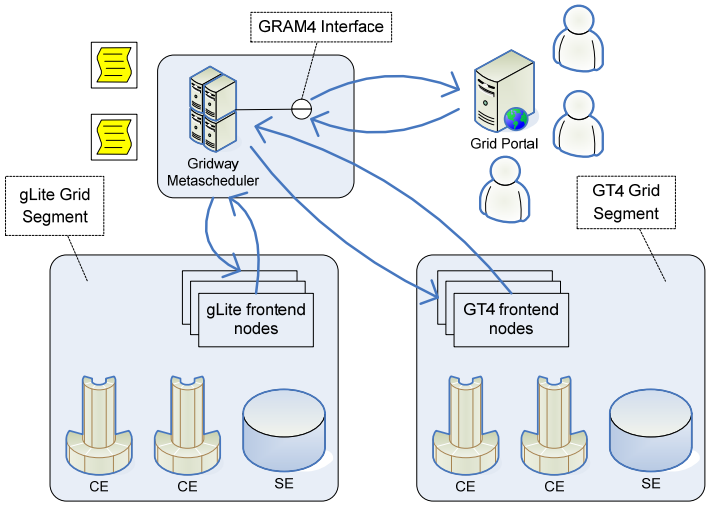

Figure 38. Metascheduler approach

In the next section, we show the examples of application of described approaches to integration of satellite monitoring systems and development of InterGrid environment.

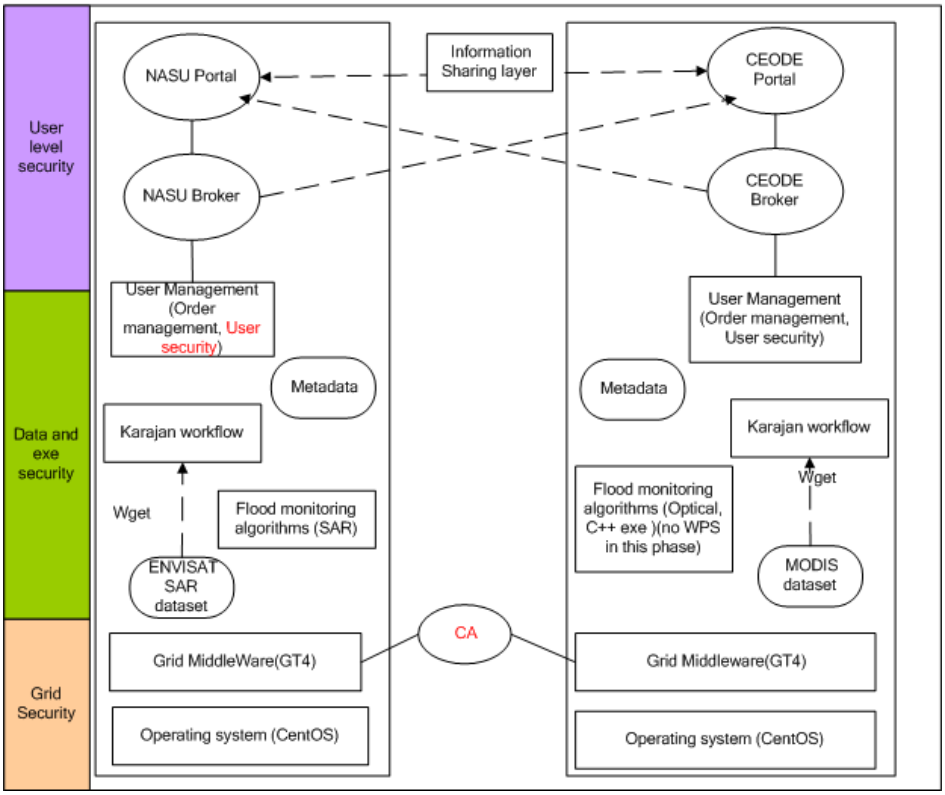

Figure 39. Three-level architecture of Wide Area Grid (WAG)

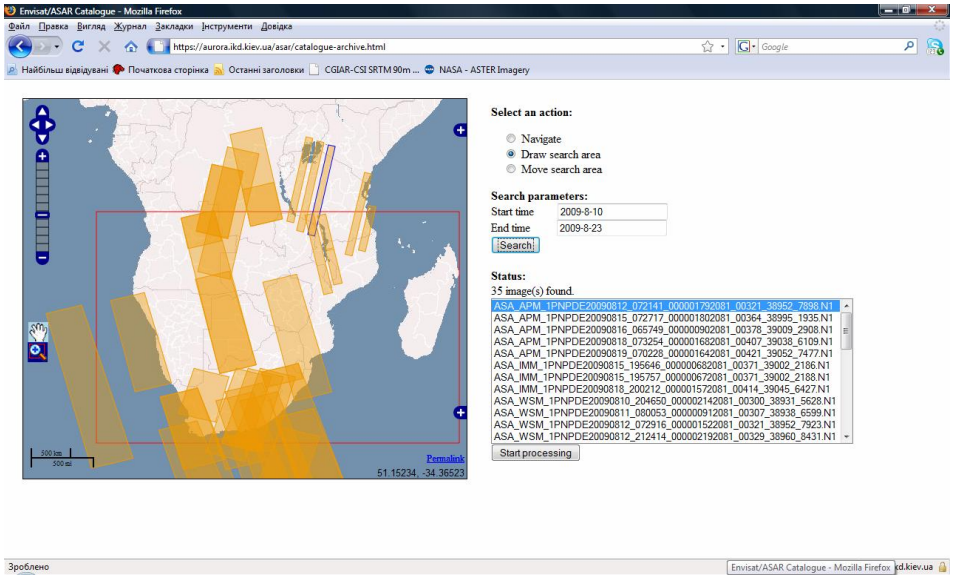

Figure 40. User interface of WAG geoinformation system

This approach is implemented within Wide Area Grid, developed within the project of CEOS by Space Research Institute NASU-NSAU, French Space Agency CNES, and Center of Earth Observation and Digital Earth. Structural scheme of its implementation is shown in Figure 39 and user interface – in Figure 40.

2.6 Implementation: lessons learned

✓ Integration of satellite monitoring systems

The first case study refers to the integration of satellite monitoring systems of NSAU (Ukraine) and IKI RAN (Russia). The overall architecture for integration of data provided by two organizations is depicted in Figure 41. The proposed approach is applied for the solution of problems for agriculture resources monitoring and crop yield prediction. Within integration NSAU provides WMS interfaces to NWP modeling data (using WRF model) [Kussul et al, 2008b], in-situ observations from meteorological ground stations in Ukraine, and land parameters (such as temperature, vegetation indices, soil moisture) derived from satellite observations from MODIS instrument onboard Terra satellite. IKI RAN provides WMS interfaces to operational land and disaster monitoring system. Both NSAU and IKI RAN provides user Web-interfaces to monitoring systems that support OGC WMS standards.

In order to provide user interface that will enable visualization of data from multiple sources we use open-source OpenLayers framework (http://www.openlayers.org). OpenLayers is "thick client" software based on JavaScript/AJAX and operational on client side. Main OpenLayers features also include: support for several WMS servers, support for different OGC standards (WMS, WFS), cache and tiling support to optimize visualization, support for of both raster and vector data. The provided data and products are accessible via Internet http://land.ikd.kiev.ua. The example of OpenLayers visualization of data from multiple sources is depicted in Figure 42.

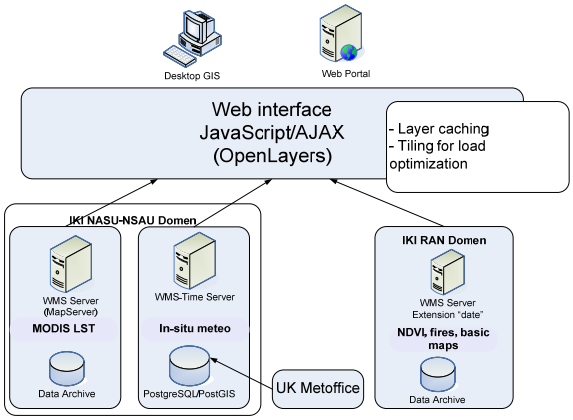

Figure 41. Architecture of satellite monitoring system integration

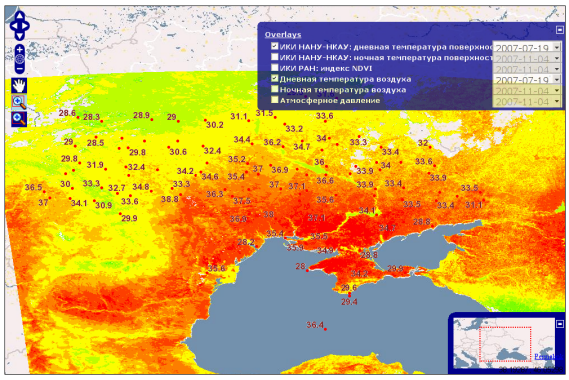

Figure 42. OpenLayers interface to multiple data

✓ InterGrid testbed development

The second case study refers to the development of InterGrid for environmental and natural disaster monitoring. InterGrid integrates Ukrainian Academician Grid (with Satellite data processing Grid segment) and CEODE Grid (Chinese Academy of Sciences) and is considered as a testbed for Wide Area Grid (WAG) implementation—a project initiated within CEOS Working Group on Information Systems and Services (WGISS).

The important application that is being solved within InterGrid environment is flood monitoring and prediction. This task requires adaptation and tuning of existing hydrological and hydraulic models for corresponding territories and the use of heterogeneous data stored at multiple sites. Flood monitoring and prediction requires the use of the following data sets: NWP modeling data (provided by Satellite data processing Grid segment), SAR imagery from Envisat/ASAR and ERS-2/SAR satellites (provided by ESA), products derived from optical and microwave satellite data such as soil moisture, precipitation, flood extent etc., in-situ observations from meteorological ground stations and digital elevation model (DEM).

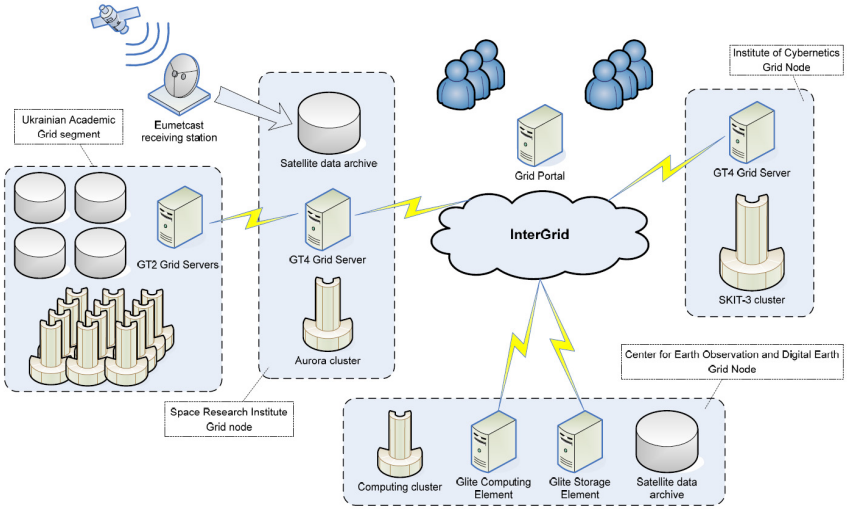

Figure 43. InterGrid architecture

The process of model adaptation can be viewed as a complex workflow and requires the solution of optimization problems (so called parametric study). Satellite data processing and products generation tasks also represent complex workflow and require intensive computations. All these factors lead to the need of using computational and informational resources of different organizations and their resources into joint InterGrid infrastructure. The architecture of proposed InterGrid is depicted in Figure 43.

GridFTP was chosen to provide data transfer between Grid systems. In order to enable interoperability between different middleware (for example, Satellite data processing Grid segment is using GT4; CEODE Grid is using gLite 3.x; Ukrainian Academician Grid is based on NorduGrid) we developed Grid portal that is based on GridSphere portal framework (http:// www.gridsphere.org). The developed Grid portal allows users to transfer data between different nodes and submit jobs on computational resources of the InterGrid environment. The portal also provides facilities to monitor statistics of the resources such as CPU load, memory usage, etc. The further works on providing interoperability between different middleware are directed to the development of metascheduler using GridWay system. In the nearest future, we are intended to provide integration with ESA's EO Grid-on-Demand infrastructure.

The system is used within the UN-SPIDER project for flood monitoring and prediction (Figure 44).

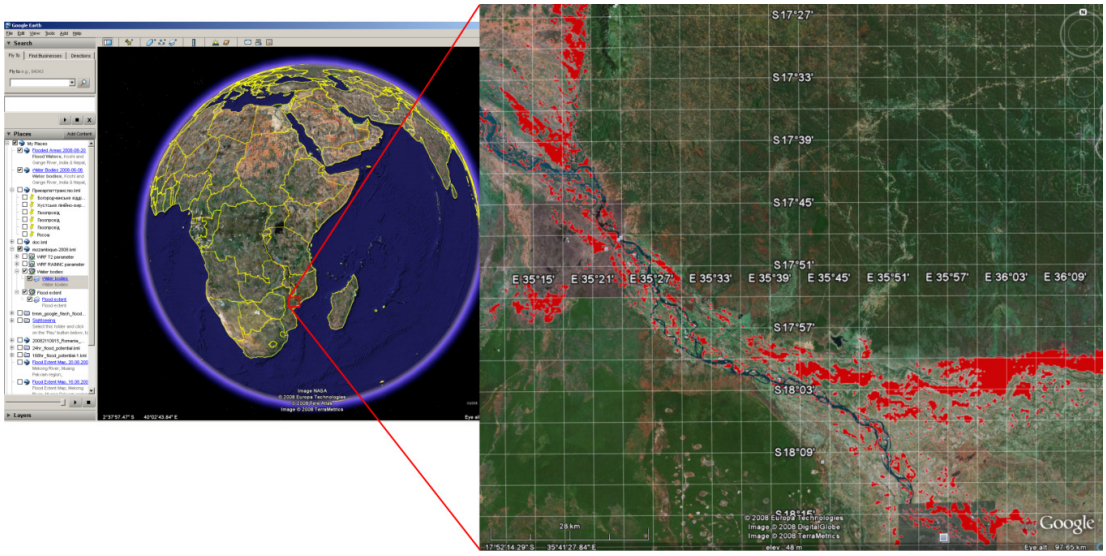

Figure 44. Global UN-SPIDER flood monitoring and risk assessment system

In conclusion, we need to point, that the Ukrainian segment is implemented under the standards of GEOSS. We use intelligent data processing technique for geographically distributed information, and this allows us to provide visual data mining and risk assessment for large-scale disasters. We studied different approaches to system integration allowing uniting different national risk assessment systems into common international infrastructure, for example, UN-SPIDER.