ISSN 1561-8889. Космічна наука і технологія. 2009. Т. 15. № 2. С. 49–55.

UDC 549.903

N. Kussul[1], L. Hluchy[2], A. Shelestov[1], S. Skakun[1],
O. Kravchenko[1], M. Ilin[1], Yu. Gripich[1], A. Lavrenyuk[1]
[1] Space Research Institute NASU-NSAU
inform@ikd.kiev.ua
[2] Institute of Informatics SAV
hluchy.ui@savba.sk

DATA FUSION GRID SEGMENT

This paper presents a Grid infrastructure that is being developed at the Space Research Institute NASU-NSAU, and integrates the resources of several geographically distributed organizations. The use of Grid technologies 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. We use the Grid infrastructure for a number of applications that heavily rely on Earth observation data. The applications include: weather prediction, flood monitoring, biodiversity assessment, crop yield prediction, and Earth land parameters estimation.

INTRODUCTION

At present, global climate changes in the world made a rational land use, environmental monitoring, prediction of natural and technological disasters the tasks of a great importance. The basis for the solution of these problems is an integrated use of data of different nature: modelling data, in-situ measurements, and indirect observations such as airborne and space-borne remote sensing data.

Satellite observations have an advantage of acquiring data for large and hard-to-reach territories, and providing continuous and human-independent measurements. Such important applications like monitoring and predictions of floods, droughts, vegetation state assessment 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 [8]. Information on flood extent provided in the near real-time (NRT) can be also used for damage assessment and risk management, and can benefit to rescuers during the flooding [2]. Both microwave and optical data can provide means to detect drought conditions, estimate drought extent and assess the damage caused by the drought events [9, 17]. To assess vegetation health/stress, optical remote sensing data can be used to derive biophysical and biochemical variables. 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 [13].

It should be emphasized 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 is both important for floods and droughts applications as well as vegetation state assessment. That is, once we develop interfaces to access the required data and products, it can be used in a uniform way for different purposes and applications. Services and models that are common for different EO applications (e. g. flood monitoring and crop yield prediction) are shown in Fig. 1.

The EO domain is characterized by the large volumes of datasets that should be processed, catalogued, and archived [6, 16]. For example, GOME instrument onboard Envisat satellite generates near-

---

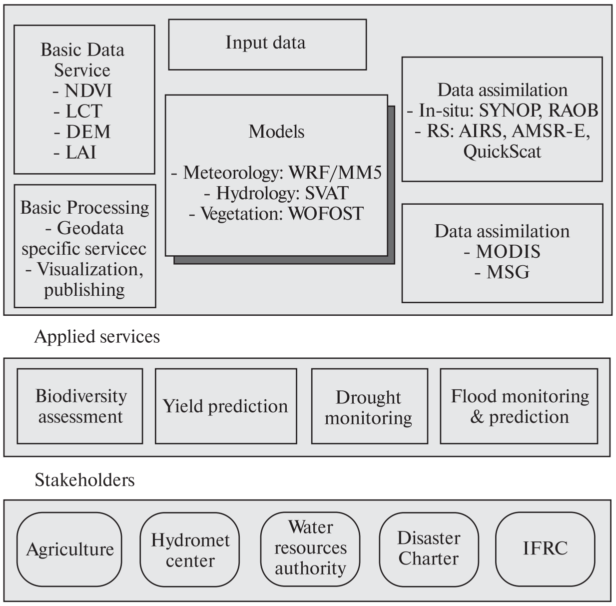

Fig. 1. Common services and models for a variety of applications

ly 400 Tb data per year [7]. The processing of satellite data is carried out not by the single application with monolithic code, but by the distributed applications. This process can be viewed as a complex workflow [3] that is composed of many tasks: geometric and radiometric calibration, filtration, reprojection, composites construction, classification, products development, post-processing, and visualization to the end-users. 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 at ESA and ESRIN [6]. This mosaic covers all Europe at 90 meters resolution, and corresponding products are automatically orthorectified using Digital Elevation Model (DEM). Dealing with EO data, we have to also consider security issues regarding satellite data policy, the need for processing in NRT for fast response within international programs and initiatives for disaster monitoring, in particular the International Charter «Space and Major Disasters» and the International Federation of Red Cross.

Hence, all these factors, such as the need for processing data in NRT, the need for managing large volumes of satellite data and derived products and providing a uniform access to them, lead to the use of Grid technologies [5, 6, 16]. In this case, a Grid environment is 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 [6].

In this paper, we focus on the description of the Grid environment that is under development on the basis of the concept from Fig. 1 within the INTAS-CNES-NSAU project «Data Fusion Grid Infrastructure». We will also show the number of real-world applications that are solved with the use of the Gird infrastructure, in particular weather prediction, flood monitoring, biodiversity assessment, crop yield prediction, and Earth land parameters estimation.

APPLIED SERVICES WITHIN GRID

In this section we review applications that are solved using the resources of the Grid system.

Weather prediction. Weather forecast data is used in the core models of flood monitoring and crop state prediction applications in the Grid environment. The numeric weather prediction model WRF (Weather Research & Forecasting Model, http://www.wrf-model.org) was configured and adapted to the territory of Ukraine [10]. Currently, we routinely produce 72-hours forecasts every 6 hour with a spatial resolution of 10 km. With such configuration, the model runs approximately 6 hours on the Grid's SCIT-3 supercomputer at the Institute of Cyber netics (total 300 Intel Xeon 3.0 GHz cores) and produces approximately 5 Gb of output data. The visualization interface for the model is depicted in Fig. 2.

Flood monitoring. We developed a neural network approach to the flood extent extraction from satellite synthetic-aperture radar (SAR) imagery [12]. We developed a parallel version of our method that can be run on several computational nodes in the Grid. The use of the Grids allowed us to considerably reduce 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. The developed Web service is

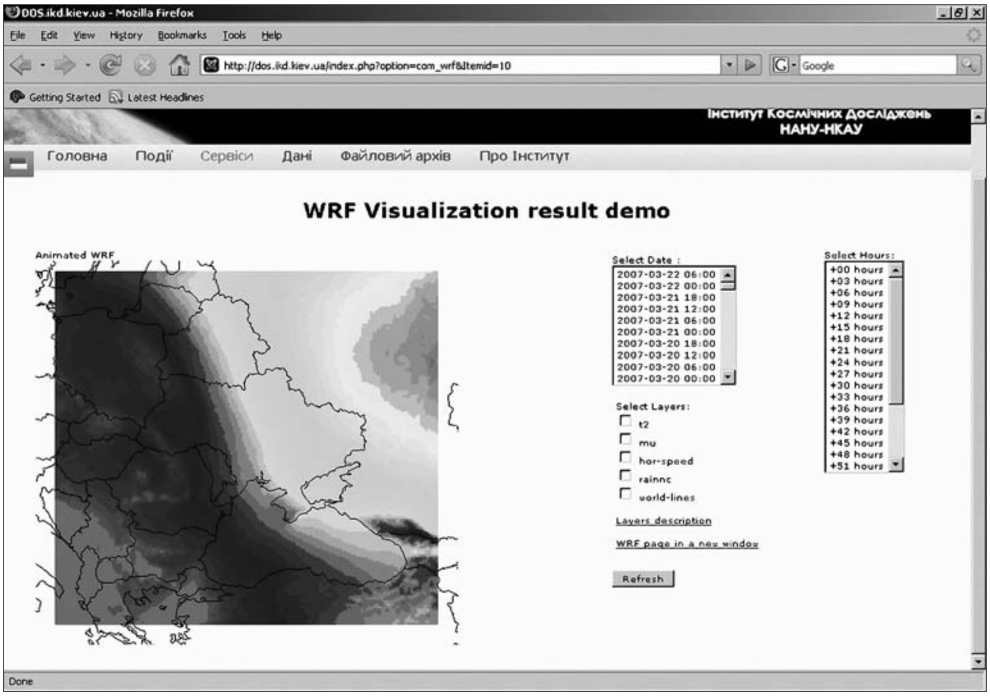

Fig. 2. Forecasts of the pressure using WRF model

accessible via Internet through the address http://floods.ikd.kiev.ua (Fig. 3).

Land biodiversity assessment. In collaboration with scientists from the Scientific Centre for Aerospace Research of the Earth of the National Academy of Sciences of Ukraine we have developed an approach for land biodiversity assessment and mapping using Earth observation data [15]. The proposed approach was developed for the Pre-Black Sea region, but, in general, can be extended to any other region. We developed a Web service for biodiversity monitoring that enables regular and operational acquisition of biodiversity estimates for the Pre-Black Sea region and allows one to track changes in its values. This, in turn, reveals negative changes in the environment of the given region and provides adequate information on biodiversity hot-spots. The developed Web service is accessible via Internet through the address http://biodiv.ikd.kiev.ua (Fig. 4).

Crop yield prediction. We implemented a time series analysis of vegetation index approach for yield prediction and vegetation state assessment [11]. As a basis, we use the enhanced vegetation index (EVI). Crop state estimation requires analysis of 217 Mb of data per run, and yield prediction requires approximately 4 Gb per run, which takes approximately 30 min in the Grid infrastructure. Estimation is started routinely for the next 16 days. Reanalysis and model real-time calibration requires nearly 20 Gb of historical data processing, and is started at least once per

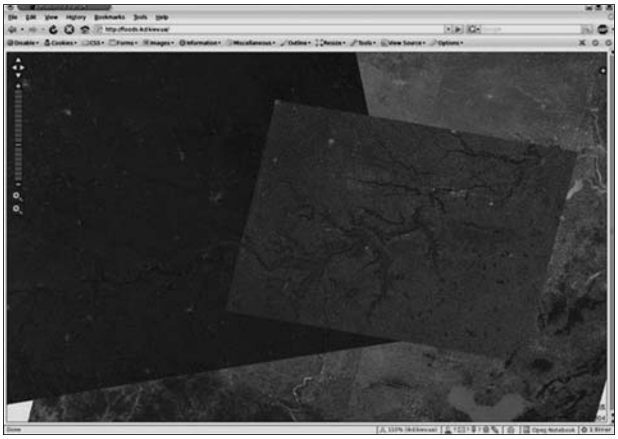

Fig. 3. Flood application within the Grid infrastructure. Flood event: River Huaihe, China, July, 2007. Data sources: Envisat/ ASAR (© ESA, 2007) and RADARSAT-1 (© CSA, 2007)

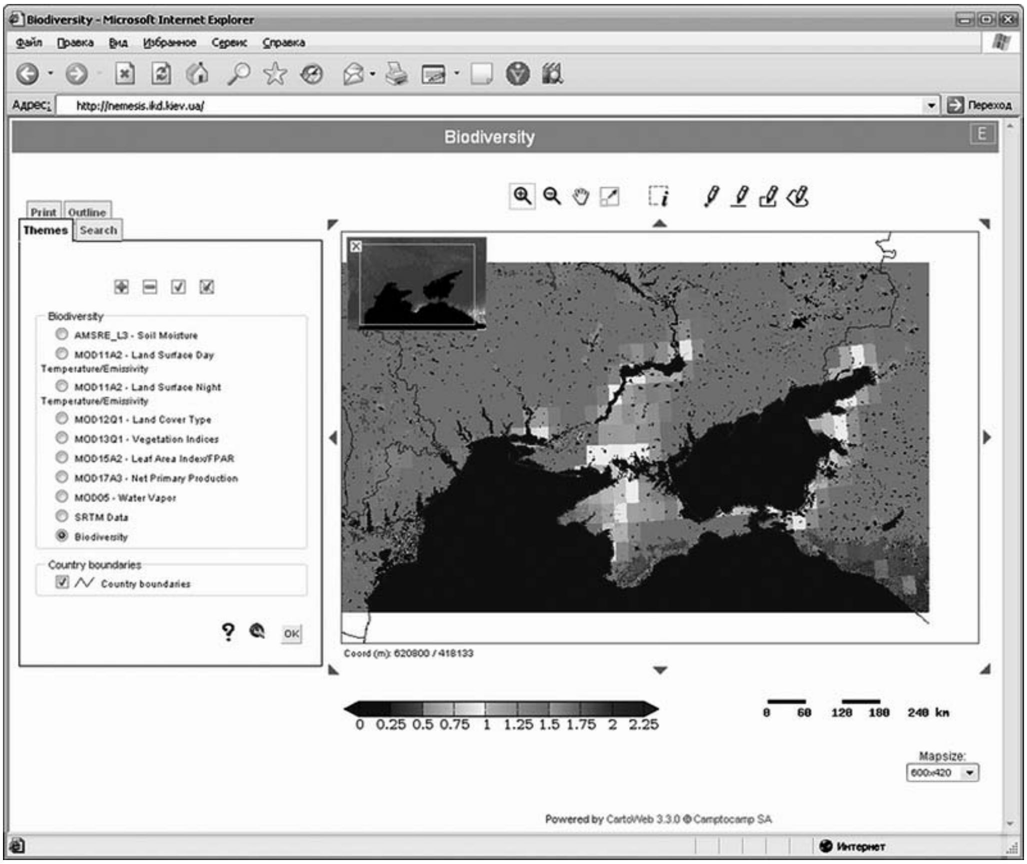

Fig. 4. Demonstration of Web service for biodiversity assessment using EO data products for the Pre-Black Sea region of Ukraine

harvest. The WOFOST model is used with the assimilation of Leaf Area Index (LAI) derived from satellite observations. The visualization interface for the developed services is shown in Fig. 5.

Earth land parameters estimation using optical remote sensing data. We developed a method for plant moisture estimation using optical remote sensing data. This information is valuable for vegetation stress assessment and drought monitoring, as well as for non-destructive moisture measurements at a field level. The proposed method consists in solving inverse problem for a canopy radiative transfer model. That is, given top of canopy reflectance in several bands in optical domain to find canopy parameters or parameters of canopy leaves, such as Equivalent Water Thickness (EWT) [13]. We use coupled leaf-canopy radiative transfer model that consists of PROSPECT [4] leaf optics model and SAILH canopy model. As input for inversion procedure, satellite measurements with atmospheric correction can be used (such as MODIS, product MOD09) or field measurements for inversion at leaf level.

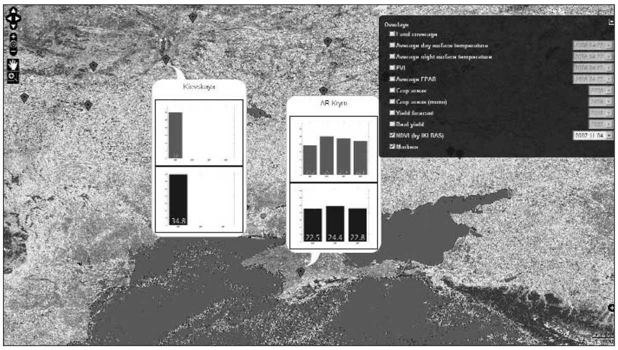

Fig. 5. Wheat yield forecast (Ukraine, 2008)

To solve the inverse problem a neural network was used, namely Mixture Density Network [1]. Such network can estimate an inverse operator and an error of inversion procedure for each given set of reflectance. This is done through estimation of a-posterior probability of canopy parameter given set of reflectance. In comparison with other methods, such as iterative optimization, our approach is much faster, while in comparison with traditional neural networks, we can estimate an error of inversion for each given input and not just an average one. Our method can be seen as a valuable alternative to lookup tables approach with a slightly different quality/performance ratio. An example of results of inversion for PROSPECT model is shown in Fig. 6. Reflectance and moisture data (Cw parameter in terms of PROSPECT model) is taken from LOPEX93 spectra database.

DESCRIPTION OF THE GRID INFRASTRUCTURE

The Grid infrastructure (that is being developed within INTAS-CNES-NSAU project «Data Fusion Grid Infrastructure») integrates the resources of geographically distributed organisations, in particular:

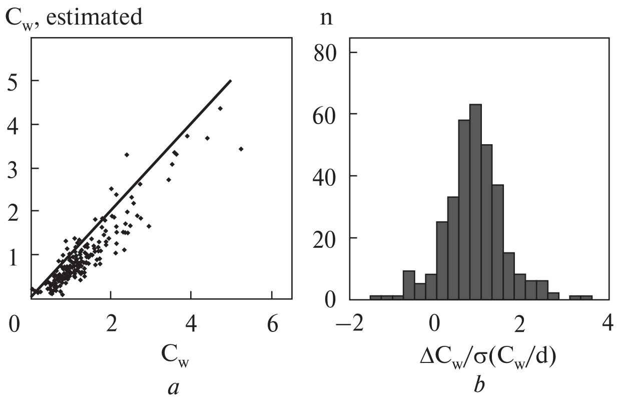

Fig. 6. Results of inversion of PROSPECT model: *a* — plot of estimated moisture parameter Cw vs. measured values, *b* — histogram of an error of estimation divided by estimation of a standard deviation

- Space Research Institute NASU-NSAU (Ukraine) with deployed computational and storage nodes based on Globus Toolkit 4 (htpp://www.globus.org) and gLite 3 (http://glite.web. cern.ch) middleware, access to geospatial data and Grid portal;
- Institute of Cybernetics of NASU (Ukraine) with deployed computational and storage nodes

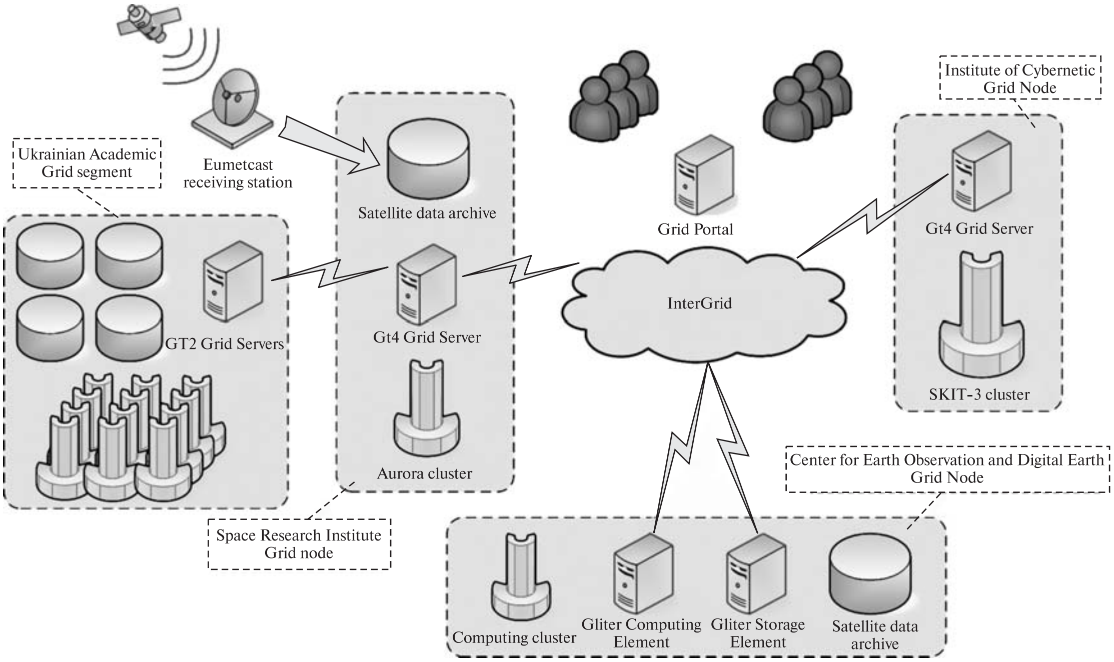

Fig. 7. Architecture of the Grid infrastructure

based on Globus Toolkit 4 middleware and access to computational resources (approximately 500 processors);

- Center for Earth Observation and Digital Earth of the Chinese Academy of Sciences (China) with deployed computational nodes based on gLite 3 middleware and access to geospatial data (approximately 16 processors).

It is worth mentioning that satellite data is 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 infrastructure. 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 organised via a high-level Grid portal that has been deployed using GridSphere framework [14]. Through the portal, users can access the required satellite data and submit jobs to the computing resources of the Grid in order to process satellite imagery. The workflow of the data processing steps in the Grid (such as transformation, calibration, orthorectification, classification etc.) is controlled by a Karajan engine (http://www.gridworkflow.org/snips gridworkflow/space/Karajan).

The existing architecture of the Grid is shown in Fig. 7.

CONCLUSION

In this paper, we presented a Grid infrastructure that is being developed at the Space Research Institute NASU-NSAU and integrates the resources of several geographically distributed organizations: the Space Research Institute NASU-NSAU, the Institute of Cybernetics NASU and the China's Center for Earth Observation and Digital Earth of CAS. The use of Grid technologies 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. Currently, we are using a Grid portal solution based on GridSphere framework to integrate Grid systems with different middleware, such as GT4 and gLite 3. In the future, we plan to implement a metascheduler approach based on a GridWay-like system.

We showed the use of the Grid infrastructure for a number of applications that heavily rely on Earth observation data. The applications include: weather prediction, flood monitoring, biodiversity assessment, crop yield prediction, and Earth land parameters estimation.

Acknowledgments*. This work is supported by the 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), and by the 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» (N 4928).*

REFERENCES

1. *Bishop C.* Pattern Recognition and Machine Learning. — New York: Springer, 2006. — 740 p.
2. *Corbley K. P.* Radar Imagery Proves Valuable in Managing and Analyzing Floods Red River flood demonstrates operational capabilities // Earth Observation Magazine. — 1999. — **8**, N 10. — http://www.eomonline.com/Common/Archives/1999dec/99dec_corbley.html.
3. DEGREE — Earth Science White Paper on Grids. — http://www.eu-degree.eu/DEGREE/internal-section/wp6/DEGREE-D6.1.2_v2.8.pdf.
4. *Feret J. B., François C., Asner G. P., et al.* PROSPECT-4 and 5: Advances in the leaf optical properties model separating photosynthetic pigments // Remote Sens. Environ. — 2008. — **112**, N 6. — P. 3030–3043.
5. *Foster I.* The Grid: A New Infrastructure for 21st Century Science // Phys. Today. — 2002. — **55**, N 2. — P. 42–47.
6. *Fusco L., Cossu R., Retscher C.* Open Grid Services for Envisat and Earth Observation Applications // High performance computing in remote sensing / Eds A. J. Plaza, C.-I. Chang. — New York: Taylor & Francis Group, 2007. — P. 237–280.
7. *Fusco L., Goncalves P., Linford J., et al.* Putting Earth-Observation on the Grid // ESA Bull. — 2003. — **114**. — P. 86–91.
8. *Horritt M. S.* A methodology for the validation of uncertain flood inundation models // J. Hydrology. — 2006. — **326**. — P. 153–165.
9. *Kogan F., Stark R., Gitelson A., et al.* Derivation of Pasture Biomass in Mongolia from AVHRR-based Vegetation

Health Indices // Int. J. Remote Sens. — 2004. — **25**, N 14. — P. 2889–2896.

10. *Kravchenko O., Kussul N.* Simulation cascade of NWP and land surface model for drought monitoring // European Geosciences Union General Assembly. — Vienna, Austria, 2008. — http://www.cosis.net/abstracts/EGU2008/00582/EGU2008-A-00582.pdf.
11. *Kussul N., Ilin M., Skakun S., Lavrenuk A.* Crop state estimation and winter wheat yield prediction using remote sensing data // Int. Book Ser. — 2008. — **3**. — P. 103–109. — (Decision Making and Business Intelligence, Strategies and Techniques / Eds K. Markov, K. Ivanova, I. Mitov).
12. *Kussul N., Shelestov A., Skakun S., Kravchenko O.* Data Assimilation Technique for Flood Monitoring and Prediction // Int. J. Inform.Theory and Applications. — 2008. — **15**, N 1. — P. 76–84.
13. *Liang S.* Quantitative Remote Sensing of Land Surfaces. — New Jersey: Wiley, Inc., 2004. — 534 p.
14. *Novotny J., Russell M., Wehrens O.* GridSphere: An Advanced Portal Framework // Proc. of the 30th EUROMICRO Conf. — 2004. — P. 412–419.
15. *Popov M., Kussul N., Stankevich S., et al.* Web Service for Biodiversity Estimation Using Remote Sensing Data // Int. J. Digital Earth. — 2008. — **1**, N 4. — P. 367–376.
16. *Shelestov A., Kussul N., Skakun S.* Grid Technologies in Monitoring Systems Based on Satellite Data // J. Automation and Inf. Sci. — **38**, N 3. — P. 69–80.
17. *Wagner W., Pathe C., Sabel D., et al.* Experimental 1 km soil moisture products from ENVISAT ASAR for Southern Africa // ENVISAT & ERS Symp. — Montreux, Switzerland, 2007.

Received February 20, 2009

Н. Куссуль, Л. Глухи, А. Шелестов, С. Скакун,
О. Кравченко, М. Ільїн, Ю. Грипич, А. Лавренюк

РОЗРОБЛЕННЯ НАУКОВО-ТЕХНОЛОГІЧНИХ ОСНОВ ІНТЕГРАЦІЇ ДАНИХ НА БАЗІ GRID

Описано Grid-інфраструктуру, яка розробляється в Інституті космічних досліджень Національної академії наук та Національного космічного агентства України. Grid-інфраструктура інтегрує обчислювальні та інформаційні ресурси географічно розподілених організацій. Використання Grid-технологій обумовлене необхідністю виконувати обчислення в режимі, наближеному до режиму реального часу, для моніторингу надзвичайних ситуацій та необхідністю управління великими об'ємами даних. Розроблена Grid-інфраструктура використовується для розв'язання низки прикладних задач з використанням даних спостереження Землі, зокрема чисельного прогнозування погоди, моніторингу повеней, оцінки видового біорізноманіття, прогнозування урожайності та оцінки параметрів земної поверхні.