Information-computational system for online analysis of georeferenced climatological data

© I.G. Okladnikov © E.P. Gordov © A.G. Titov © T.M.Shulgina
Institute of Monitoring of Climatic and Ecological Systems SB RAS, Tomsk
Tomsk branch of Institute of Computational Technologies SB RAS, Tomsk
Tomsk State University, Tomsk

oig@scert.ru gordov@scert.ru titov@scert.ru stm@scert.ru

Abstract

An Internet accessible information-computational system (ICS) is dedicated to visualization and analysis of geospatial gridded datasets in the area of Earth system science (local observations, remote sensing and modeling) and climatology in particular. This online ICS is based on a dedicated software framework consisting of three key parts: server-side computational back-end, server-side middleware represented by Geoserver software and PHP controllers, and JavaScript library for building typical components of web mapping client GUI developed using AJAX techniques. Huge multidimensional geospatial datasets are stored in a Network Common Data Form (NetCDF) and processed by a set of validated software modules included in the back-end. Results of a processing are represented by overlapped raster and vector layers placed over a georeferenced background. Data processing functionality allows performing basic and complex statistical analysis of climatological data whilst online geo-information system (GIS) instruments give a user an ability to combine and place georeferenced results over a chosen cartographical basis. It can provide specialists involved into multidisciplinary research projects, and even users without programming skills, with reliable and practical online instruments for integrated research of climate and ecosystems changes through a unified web-interface.

1 Introduction

Specifics and regional environmental applications of basic Earth system sciences make them multidisciplinary, involving into studies a number of nationally and internationally distributed research groups. Success of this cooperative work depends on ability of these groups to rapidly exchange data and knowledge, coordinating activities and optimizing usage of information-computational resources, services and applications. Georeferenced datasets are currently actively used in numerous applications including modeling, interpretation and forecast of climatic and ecosystem changes for various spatial and temporal scales [9]. Due to inherent heterogeneity of environmental datasets as well as their huge size which might constitute up to tens terabytes for a single dataset, at present, studies in the area of climate and environmental change require a special software support [5]. A dedicated online information-computational system for analysis of georeferenced climatological and meteorological data is presented in this work. It is based on Open Geospatial Consortium (OGC) standards and involves a number of modern solutions such as object-oriented programming model, modular composition, and JavaScript components based on GeoExt library [12], ExtJS Framework [23] and OpenLayers software [15]. This system is being developed in the framework of a big research project aimed to creation of a distributed hardware and software platform for monitoring and forecasting of regional climatic and ecological changes and supporting of continuous education. And the work is still not finished.

Proceedings of the XVII International Conference «Data Analytics and Management in Data Intensive Domains» (DAMDID/RCDL'2015), Obninsk, Russia, October 13 - 16, 2015

2 Related work

At present, only few web-based information systems devoted to handling of geophysical data are known: GIOVANNI – GES-DISC (Goddard Earth Sciences Data and Information Services Center) Interactive Online Visualization ANd aNalysis Infrastructure ([1], [8]); dissemination system based on Coupled Climate Model of Institute for Numerical Mathematics of the RAS (INM RAS, [3]); Climate explorer by Royal Netherlands Meteorological Institute (KNMI, [2]); distributed informational-analytical system for searching, processing and analysis of spatial data based on the combination of GIS and Web technologies currently being developed at Institute of Computational Technologies of the SB RAS (ICT SB RAS, [24]); An Integrated Mapping and Analysis System with Application to Siberia (RIMS, [19], [21]). Nevertheless, in spite of presence of some efforts made in Earth and Space Sciences Informatics research area there is still a lack of powerful online tools combining various capabilities to perform online processing, analysis and visualization of multidimensional heterogeneous data collections using unified web interface for integrated study of global and regional environmental changes.

3 Results

Basic requirements to functional capabilities of the system were defined through meetings with specialists in the area (climatologists, ecologists and biologists from Russian and international institutions) and analysis of literature ([10], [13], [22]). These requirements also include key features of a typical thematic web GIS: Internet-access by means of a common graphical Internet-browser to all data processing and visualization tools; abilities to overlay processing results as layers on top of geo-referenced background with basic functions of desktop GIS, such as map navigation, synchronous overview of layers at different scales with capability to show/hide/rearrange them, associating specific colors in a legend with range of values, and inspecting data values at specific coordinates of several spatial data layers at once. To provide interoperability with other information systems requirements were expanded to include support for exporting of results in NetCDF files, geo-referenced Tagged Image File Format (geoTIFF, [7]), Open Geospatial Consortium (OGC) Keyhole Markup Language (KML) and Geography Markup Language (GML).

To satisfy these requirements the information-computational system was based on a dedicated software framework for developing thematic online ICS for analysis of dynamics of Earth system components [26]. This software framework includes:

1) A server-side computational back-end (core), providing data access and processing, and output of results;

2) A server-side middleware including PHP [16] controllers run by web portal and providing management of cartographical web services, computational core and components of graphical user interface (GUI);

3) A JavaScript library for developing components of client web GIS application GUI based on AJAX techniques.

Computational backend of the information-computation system is developed using high-level open-source GNU Data Language (GDL, [6]) and Python [20]. These are powerful programming languages providing reliable computational and visualization routines, as well as application programming interface (API) for NetCDF [14] library and PostgreSQL [18] database controller. Middleware implementing basic functionality such as user authentication, database connectivity, HyperText Markup Language (HTML) templates usage, language localization, content management system and so on is represented by a web portal and an integrated set of dedicated PHP modules (controllers). They are organized into the package structure and provide server-side API to interact with graphical user interface, computational backend, cartographical web services (WMS/WFS) and geospatial data storage. Client-side part of the ICS is developed using framework GeoExt, ExtJS and OpenLayers JavaScript libraries. It provides AJAX-based client-side API for implementation of

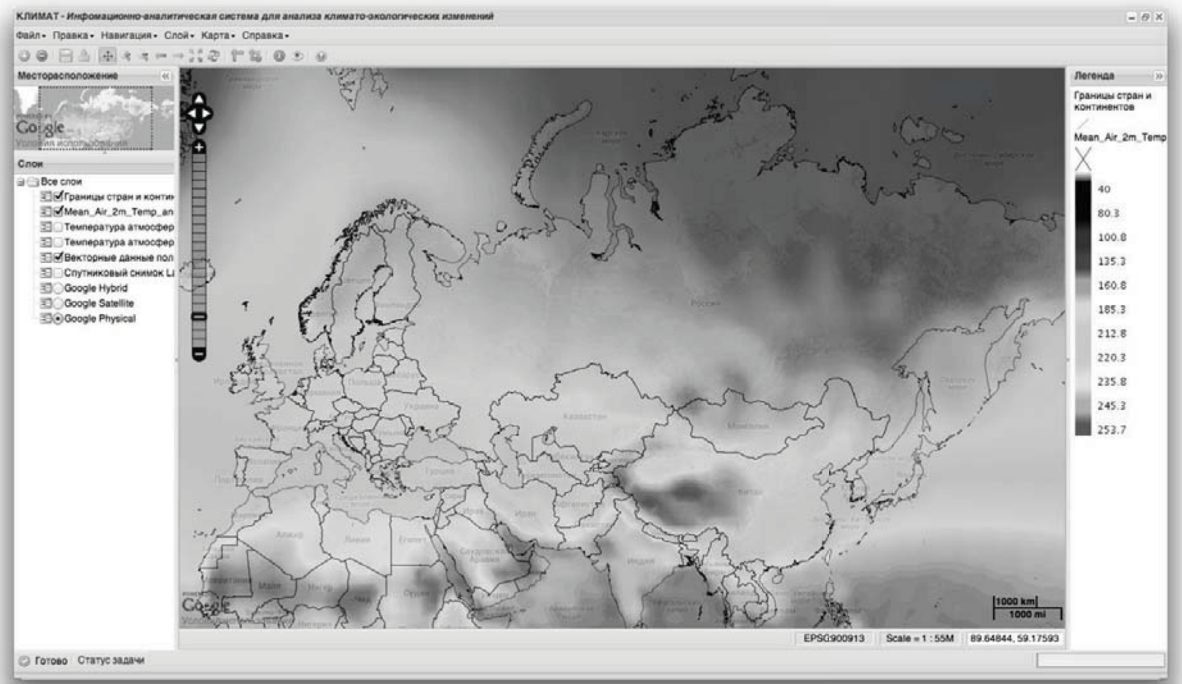

Fig.1. Main window of the system

corresponding PHP controllers as well as basic GUI elements of typical desktop application.

Several geophysical datasets are available for processing by the system: NCEP/NCAR Reanalysis II, JMA/CRIEPI JRA-25 Reanalysis, ECMWF ERA-40 Reanalysis, ECMWF ERA Interim Reanalysis, MRI/JMA APHRODITE's Water Resources Project Reanalysis, DWD Global Precipitation Climatology Centre's data, GMAO Modern Era-Retrospective analysis for Research and Applications, reanalysis of Monitoring atmospheric composition and climate (MACC) Collaborated Project, NOAA-CIRES Twentieth Century Global Reanalysis Version II, NCEP Climate Forecast System Reanalysis (CFSR), meteorological observational data for the territory of the former USSR for the 20th century, results of modeling by global and regional climatological models, and this list is extending. Flexibility of the system allows easy and fast expansion of number of datasets available for processing and visualization.

Reanalyses and modeling data are stored as NetCDF files organized in a hierarchy of a file system's directories:

```
/<data root directory>/
    <dataset name>/
        <spatial resolution>/
            <time resolution>/
                <other subfolders and files…>
```

Here, <data root directory> is an absolute path to a common root directory of all data archives, <dataset name> is a name of a subdirectory containing NetCDF files of a single data archive, <spatial resolution> is a name of a subdirectory containing files of a single spatial resolution, <time resolution> is a name of a subdirectory containing files of a single time resolution. Names of underlying subfolders as well as NetCDF files are not regulated and depend on a dataset. Each NetCDF file contains georeferenced multidimensional sets of climatological data and corresponding metadata (georeferencing and description) and spans some spatial and time ranges. Such data model allows to organize datasets on a storage systems. And a dedicated metadata base, developed as a part of the system, allows fast location of required data files according to a processing task.

Meteorological observational data for the territory of the former USSR for the 20th century are represented by time series of a several characteristics observed at meteorological stations. To provide fast location and extraction these data were converted and placed in a PostgreSQL database with PostGIS extension.

Also a functionality to run the Weather research and forecasting model (WRF, [28]) and "Planet simulator" model by Meteorological Institute of University of Hamburg [17] was implemented in the system. Due to many preset parameters, as well as limited temporal and spatial ranges set in the system these models have low computational power requirements and are used in educational process thus providing better understanding of basic climatological and meteorological processes by students and young scientists [11].

Spatially distributed results of a processing are represented by shaded color plots stored in geoTIFF files. Spatially distributed results in the form of contour plots, vector wind streams and local point observations are stored in ESRI Shapefiles [4]. 2-D plots and diagrams are stored in Encapsulated PostScript (EPS) format files. Additionally, files with raw values (CSV, XML, NetCDF) could be written to accompany GeoTIFF and EPS files.

The main advantage of the system is that it gives users an ability to perform mathematical and statistical data analysis of huge georeferenced datasets without downloading them, remotely through a window of a web browser installed on a common desktop computer connected to Internet [27]. In contrast to many scientific gateways and analytical web-systems existing nowadays (http://sciencegateways.org/), this system allows not only to do remote data analysis and visualization, but to represent results in a web interface as a stack of georeferenced layers on an interactive map implementing GIS features (Fig. 1 and 2), and to export them as OGC-compliant (in terms of metadata) layers using web mapping (WMS) and web feature (WFS) services, thanks to a tight fusion of web-, GIS- and cloud-computing technologies. At present, for the Earth system science area there is no similar information-computational system with analogous capabilities and data archive. The system's ability to export results as downloadable binary files and WMS/WFS-layers provides their interoperability with other online and offline analytical software including user's desktop GIS. Since results obtained are much smaller than raw datasets they can be downloaded, visualized and processed on a user's desktop computer in a reasonable time.

Currently, experimental version of the system is available for registered users on the web site: http://climate.climate.scert.ru/. It is hosted at IMCES SB RAS (Institute of Monitoring of Climatic and Ecological Systems, Siberian Branch of Russian Academy of Sciences, Tomsk, Russia) on a hardware that includes: high performance server HP Proliant DL585 G7 (four 12-core processors AMD Opteron 6172 and 32 Gb RAM) for the computational core, dedicated server (Intel Pentium 4 and 2 Gb RAM) for the web-portal, dedicated server (two Intel Xeon 5130 and 4 Gb RAM) for Geoserver software. All servers are interconnected using 1 Gb/s Ethernet. Such configuration is able to process up to 20 simultaneous user tasks.

4 Discussion

The information-computational system has proved its reliability and effectiveness during investigation of modern climatic changes in Siberia [25]. Surface air temperatures in Siberia for the last few decades were taken from ECMWF ERA Interim dataset, and precipitations for the same period were taken from JMA APHRODITE's project dataset tied to the system. Researchers have benefited the ability to easily perform analysis of huge multidimensional spatially distributed datasets, to visualize results obtained and investigate data in overlapping layers at particular geographical points. Analysis results obtained were found to be in good correlation with results obtained by another authors, and qualitatively precise spatial scales of the climate changes were revealed. Modular software architecture and continuous contact between researchers and developers allowed to timely implement required features in the ICS. Internet-accessible instruments and "cloud" data processing allowed distributed multidisciplinary group of researches to perform complex statistical analysis using familiar Internet browsers on their desktop computers without downloading huge amounts of geospatial data.

The decision to use framework GeoExt along with ExtJS and OpenLayers JavaScript libraries has been very beneficial for development of web GUI imitating desktop application. These libraries have provided AJAX-based client-side API for corresponding PHP controllers and reduced development overhead for construction basic GUI elements by using standard library component, thus producing reliable web application with fewer lines of code.

Although the system was developed for an analysis of global and regional climate changes and an assessment of an impact of these changes on ecosystems and human activities, it has found its purpose in an education process. The ICS was accompanied by electronic courses based on an educational environment MOODLE (Modular Object-Oriented Dynamic Learning Environment, https://moodle.org/), thus giving opportunity to use it to study high-grade and post-graduate students with profile specialties [11]. In 2012 high-grade students of chair of meteorology and climatology in Tomsk state university studied basics of climate change analysis using the web GIS and performed labs on courses "Regional climate change analysis" and "Future climate analysis".

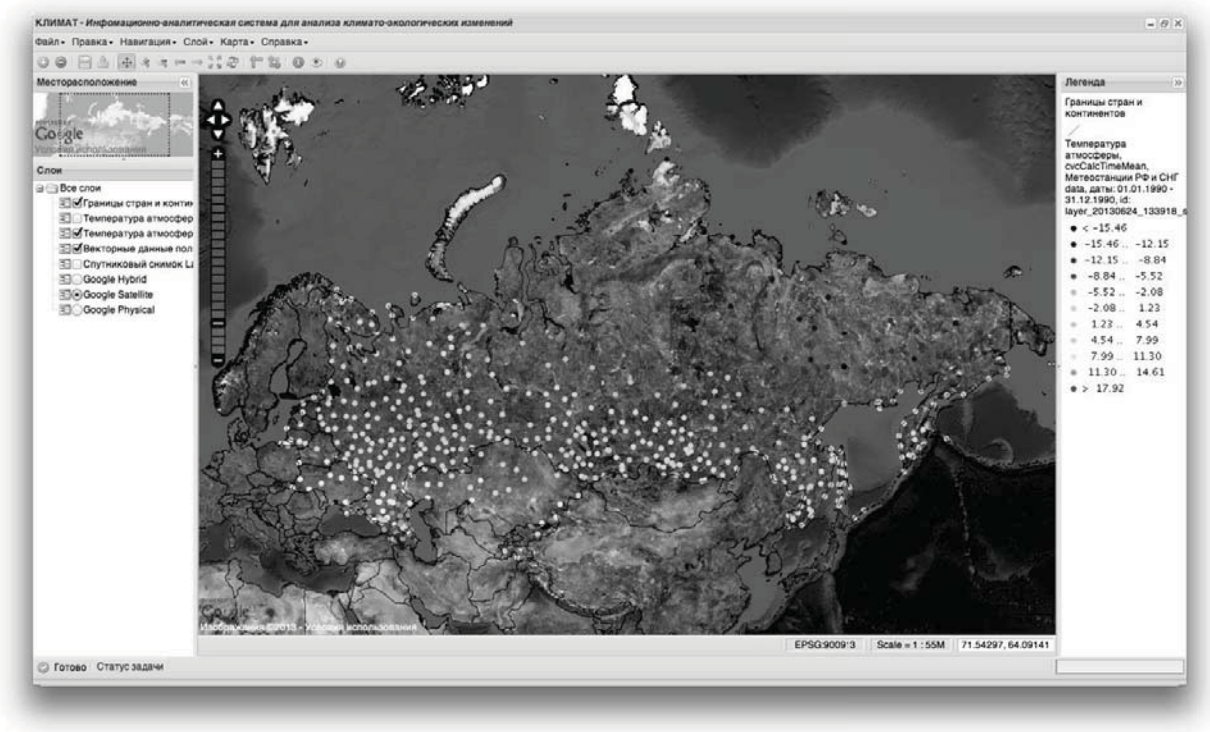

Fig. 2. Visualization of meteorological stations data analysis

The ICS is the first thematic web GIS for climatological research based on the dedicated software framework designed by the authors [27], and its still under hard development. Progress of extending its functionality depends on needs of partners within the climatological and ecological communities. Modular composition of the system allows an easy addition of new online processing and visualization tools as well as new GUI elements implementing new requirements of users.

4 Conclusion

The web information-computational system for geophysical data analysis provides specialists involved into multidisciplinary research projects with reliable and practical instruments for complex analysis of climate and ecosystems changes on global and regional scales. Using it even unskilled user without specific knowledge of software development data standards can perform computational processing and visualization of large meteorological, climatological and satellite monitoring datasets through unified web-interface in a common graphical web-browser.

This work is partially supported by SB RAS Basic research project VIII.80.2.1 and Russian Foundation for Basic Research (RFBR grants 13-05-12034 and 14-05-00502).

References

[1] Berrick, S.W., Leptoukh, G., Farley, J.D., and Rui, H., 2009. *Giovanni: A Web service workflow-based data visualization and analysis system*. IEEE Trans. Geosci. Remote Sens., 47(1), 106-113.

[2] Climate Explorer, http://climexp.knmi.nl/

[3] Coupled Climate Model of the INM RAS, http://ksv.inm.ras.ru/GCM_DATA_PLOTTING/GCM_INM_DATA.html

[4] ESRI Shapefile Technical Description. http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf

[5] Felice Frankel, Rosalind Reid. Big data: Distilling meaning from data // Nature. Vol. 455. N. 7209. P. 30.

[6] GDL – GNU Data Language. http://gnudatalanguage.sourceforge.net/

[7] GeoTIFF. http://trac.osgeo.org/geotiff/

[8] GIOVANNI, http://daac.gsfc.nasa.gov/techlab/giovanni/

[9] Gordov E.P., Lykosov V.N., Krupchatnikov V.N., Okladnikov I.G., Titov A.G., Shulgina T.M. Computational and information technologies for monitoring and modeling of climate changes and their consequences. – Novosibirsk: Nauka, Siberian branch, 2013. – 195 p.

[10] Gordov, E., K. Bryant, O. Bulygina, I. Csiszar, J. Eberle, S. Fritz, I. Gerasimov, R. Gerlach, S. Hese, F. Kraxner, R. B. Lammers, G. Leptoukh, T. Loboda, I. McCallum, M. Obersteiner, I. Okladnikov, J. Pan, A. A. Prusevich, V. Razuvaev, P. Romanov, H. Rui, D. Schepaschenko, C. Schmullius, S. Shen, A. I. Shiklomanov, T. Shulgina, A. Shvidenko, A. Titov, 2013. Development of Information-Computational Infrastructure for Environmental research in Siberia as a baseline component of the Northern Eurasia Earth Science Partnership Initiative (NEESPI) Studies / Regional Environmental Changes in Siberia and Their Global Consequences // Series: Springer Environmental Science and Engineering. Ed.: Groisman, Pavel Ya., Gutman, Garik. Vol. XII, 2013. P. 19-55.

[11] Gordova Yu.E., Genina E.Yu, Gorbatenko V.P., Gordov E.P., Kuzhevskaya I.V., Martyanova Yu.V., Okladnikov I.G., Titov A.G., Shulgina T.M., Barashkova N.K. Supporting educational process in modern climatology by web-GIS platform "Climate" // Open and remote education. Tomsk. 2013. N.1 (49). P. 14 – 19.

[12] JavaScript Toolkit for Rich Web Mapping Applications. http://www.geoext.org

[13] Kadochnikov, A.A., Popov, V.G., Tokarev A.V., Yakubailik, O.E., 2008. Implementation of Internet GIS Portal for Environment and Natural Resources Monitoring Tasks. *J. of Siberian Federal University. Engineering & Technologies*, 1(4), 375-384.

[14] Network Common Data Form. http://www.unidata.ucar.edu/software/netcdf/

[15] OpenLayers 3. http://openlayers.org

[16] PHP: Hypertext preprocessor. http://php.net/

[17] Planet Simulator model. http://www.mi.uni-hamburg.de/index.php?id=216

[18] PostgreSQL. http://www.postgresql.org/

[19] Proussevitch, A.A., Vörösmarty, C.J., Glidden, S., Fekete, B.M., Green,P., Lammers, R.B. (in review 2009). Global-RIMS: A Global Rapid Integrated Mapping System for Hydrology Visualization and Analysis. Computers and the Geosciences.

[20] Python. http://python.org

[21] RIMS, http://RIMS.unh.edu/

[22] Romañach, S.S., M. McKelvy, K. Suir, C. Conzelmann, 2015. EverVIEW: A visualization platform for hydrologic and Earth science gridded data // Computers & Geosciences 76 (2015), P. 88–95.

[23] Sencha Ext JS. The most comprehensive MVC/MVVM JavaScript framework for building feature-rich cross-platform web applications. http://www.sencha.com/products/extjs

[24] Shokin, Y.I., Zhizhimov, O.L., Pestunov I.A., Sinyavsky Y.N., Smirnov V.V., 2007. Distributed information-analytical system for searching, processing and analysis of spatial data. *Computational Technologies,* 12, Special issue, 108-115.

[25] Shulgina, T.M., Genina, E.Yu., Gordov, E.P. Dynamics of climatic characteristics influencing vegetation in Siberia. // *Environ. Res. Lett*., 2011, 6(4). doi: 10.1088/1748-9326/6/4/045210.

[26] T.M. Shulgina, E.P. Gordov, I.G. Okladnikov, A.G., Titov, E.Yu. Genina, N.P. Gorbatenko, I.V. Kuzhevskaya, A.S. Akhmetshina. Software complex for a regional climate change analysis. // Vestnik NGU. Series: Information technologies. 2013. Vol. 11. Issue 1. P. 124 – 131.

[27] Titov, A.G., Gordov, E.P., Okladnikov, I.G. // Hardware-software platform "Climate" as a basis for local spatial data infrastructure geoportal. Vestnik NGU. Series: Information technologies. 2012. Vol. 10. Issue 4. P. 104 – 111.

[28] Weather Research and Forecasting Model. http://www.wrf-model.org/