Elements of software package for “cloud” analysis of the environment and climate change

I G Okladnikov\(^1, 2\), E P Gordov\(^1, 2\), A A Ryazanova\(^1\) and A G Titov\(^1, 2\)

\(^1\)Institute of Monitoring of Climatic and Ecological Systems SB RAS, Tomsk, Russia
\(^2\)Institute of Computational Technologies SB RAS, Tomsk branch, Tomsk, Russia

Abstract. The architecture and new basic components of an information and computing software package are presented. It is based on a software framework dedicated to carrying out scientific research related to statistical processing and analysis of spatial geophysical data archives obtained from observations and modelling. To create the package, experience in the development of some information-computational web GISs for processing of large amounts of spatial data has been used. The basic components of this complex are represented by several procedures for searching, sampling, and processing of spatial data arrays, as well as by elements of a graphical user interface. The flexible structure of the computing unit of this software complex provides a possibility of operative expansion of its functionality by using additional procedures for mathematical and statistical analysis, processing, and graphical representation of the results in the form of graphs, diagrams, and plots on maps of the territory of interest.

1. Introduction

The growth of environmental data to the petabyte level and the need to store, search, retrieve, process, visualize, and analyze such geospatial datasets cast doubt on the traditional approaches. This required the creation of new technologies for distributed access and data processing, and turned this area of information technologies into a new scientific field that has recently appeared: a data-intensive application domain, in which appropriate approaches and tools are developed [1-5]. Such tools should provide the functionality of searching for geographic information resources, data sampling and processing, as well as cartographic visualization services along with corresponding client applications [6]. The combination of these tools forms a specialized data processing infrastructure, which has many names nowadays: “virtual research environment”, “virtual research laboratory”, “scientific gateway”, “e-Science”, etc. [7, 8], and allows multidisciplinary distributed research teams to perform collaborative research [9].

This infrastructure is commonly based on the modern web and GIS technologies [10-12], which provides functionality for data browsing, navigation, scrolling, scaling and graphic overlay, as well as displaying map legends and corresponding meta-information in a window of a web browser. The advantages of web-oriented GIS technologies are obvious: theoretical independence from web browsers and operating systems, automatic software location, updating, combined use of geographically distributed data sources, and sharing of centralized data warehouses. Thus, the use of web-GIS technologies provides environmental scientists and decision-makers with reliable and understandable tools to study the effects of climate change using a familiar graphical user interface.

There are several computational and/or information web systems working with geophysical data linked to environmental science. Some of them deal with the analysis of real-time meteorological observational data [13]. Some of them, such as NASA’s GES-DISC Interactive Online Visualization And aNalysis Infrastructure (GIOVANNI), (http://daac.gsfc.nasa.gov/techlab/giovanni/) [14] are designed for analysis and visualization of satellite data. One of the most flexible web-GIS is a Regional Integrated Hydrological Monitoring System (RIMS) created at the University of New Hampshire, USA [15, 16]. As a fully integrated advanced system for online analysis of heterogeneous climate, hydrology, and remote sensing data, the RIMS was a successful attempt to realize a multi-functional GIS as a web application using MapServer...
The RIMS is widely used for various research projects [15, 17] including NEESPI (Northern Eurasian Earth Science Partnership Initiative, http://NEESPI.sr.unh.edu/). The IRI/LDEO Climate Data Library is a powerful and freely accessible online data repository and analysis tool that allows a user to view, analyze, and download climate-related data through a standard web browser (https://iridl.ldeo.columbia.edu/). The EVER-EST (European Virtual Environment for Research – Earth Science Themes: a solution) is a novel project aimed at creating a virtual research environment (VRE) focused on the requirements of the Earth science community allowing one to collaborate, share, analyze, and visualize data over the internet (https://ever-est.eu/).

This paper presents some of the new results of developing a software package “Climate” for distributed “cloud” processing and advanced statistical analysis of large environmental datasets [18, 19]. We describe some of the created new elements of the developed software package for the analysis of environmental and climate changes implemented in the previously proposed specialized environment for the development of thematic information and computing web GIS aimed at the processing of large amounts of spatial data [20].

2. Methods and results

The software package being considered is used to answer three kinds of user queries: plain, specific, and in-depth. Most user requests for climate and environmental data services are simple. The users are interested in the spatio-temporal dynamics and basic statistical characteristics of common climate variables (air temperature and pressure, total precipitation, snow cover, etc.). However, a special analysis (such as “what if?”) may be requested also. In this case, scientists prefer to perform their own research and use raw (or pre-processed) data accompanied by corresponding metadata. In-depth analysis is required only in rare cases. The first two kinds of queries can be processed online, and the third one is usually considered independently using dedicated software and on-demand.

An architecture that meets these requirements and outlines the main hardware and software components of the software package has been created. It includes geoportal, web client and web services that provide access to cloud computing resources. Such a software package is designed for distributed processing of large sets of climatic spatial data, and can be deployed on several interlinked (LAN, Internet) computing units. Each software complex consists of four typical blocks (Figure 1) [21]:

- structured archives of spatial datasets accompanied by metadata;
- server-side modular computing backend for data searching, sampling, processing, and visualization;
- geoportal providing communication with cartographic web services and running the computing core;
- client-side web-GIS client providing an interactive graphical interface to analysis and visualization tools.

![Figure 1. General architecture of the developed software complex.](image)

The structured archives of spatial datasets are represented by heterogeneous modeling (reanalysis, meteorological models) and observation (weather stations, remote sensing) data for various spatio-temporal
domains in different file formats. After pre-processing they are stored in high-performance storage systems in accordance with the dedicated storage and processing model developed earlier [22] and accompanied by corresponding metadata. As the main file format for gridded spatial data, the Network Common Data Form (netCDF) was selected. The set of data available for processing can be easily expanded thanks to a dedicated metadata database.

According to this architecture, some key components of the software package were implemented. They provide searching, sampling, processing, and visualization of data stored in the data archives. Representative fragments of some spatial data archives were collected and pre-processed, and some methods for their basic processing were selected and implemented (for the purpose of demonstrating the functionality). Together, these modules form a server-side computing backend running on high-performance computational hardware and orchestrated by the geoportal. After that processing results are written by the core into files in Encapsulated PostScript (EPS), Geo Tagged Image File (GeoTIFF), ESRI Shapefile and NetCDF formats, and are available for download.

A specialized geoportal, also functioning on the server-side, serves as a link between the elements of the developed system, as well as between the system and the end-user. It implements the necessary basic functionality, such as user authentication, connection to databases, use of HTML templates, language localization, the content management system (CMS), and several other features. GeoServer software (http://geoserver.org/) conforming to OpenGIS standards (http://www.opengeospatial.org/standards) is proposed as a technological basis for presenting cartographic information on the Internet.

Access to the resources and interactive tools of the software package is carried out through a specialized web client (Figure 1). It works in a graphical web browser window, and provides a user familiar with geographic information systems (GIS) with a user-friendly and intuitive graphical interface. The web client interface allows you to select geophysical characteristics for analysis, specify a space-time domain, and also set the desired data processing method. After data processing the results are displayed in the graphical interface on the interactive map of the selected region in the form of cartographic layers. The user can additionally download the same results in NetCDF format, as well as through WMS (Web Mapping Service) and WFS (Web Feature Service) interfaces.

The interactive software tools implement both the basic (calculation of averages, WMO climate indices, trends) and advanced (joint analysis of two different georeferenced characteristics of climatic processes) functionality of spatial data analysis. The interface of the application performing analysis of two climatic characteristics is shown in Figure 2.

In the course of expanding the software package functionality, new interactive software tools for statistical analysis of spatial climatological data time series were also developed [23]. These tools are based on reliable algorithms from the packages “extRemes” [24], “quantreg” (https://cran.r-project.org/web/packages/quantreg/), and “copula” [25], which implement time-dependent extremum statistics, quantile regression and copula approach. These packages were written in the R language (https://www.r-project.org/) and provide a flexible API for integration with third-party software.

The results obtained show that the developed software complex and tools will be useful for decision-makers and specialists working in affiliated sciences, with focus on socio-economic and ecological impact assessment, adaptation strategies, science policy administration, and other climate-related activities. On this basis, they will get reliable climate-related characteristics required for studies of economic, political, and social consequences of global climate change at the regional level.
3. Conclusions

This work was aimed at developing a thematic software package for analyzing climatic changes by integrating interdisciplinary (geographic, climatic, meteorological) archives of observations, modeling, and remote sensing in the form of a thematic distributed information and computing system with GIS functionality and “cloud” processing capability. Such a computing complex is the next step in the development of applied information and telecommunication systems. It could provide the specialists in various fields of science with the unique functionalities for performing reliable analysis of heterogeneous geophysical data. Since only proven computational algorithms have been used, the results obtained are reliable. The system is accessible through the Internet and has an ability to work with data without special knowledge in programming and, thus, a wide range of researchers and decision-makers can concentrate on solving their specific problems.

The above-presented software package is based on a modular architecture and can be easily adapted to various sets of geophysical data and tasks from various areas of the Earth sciences. The open-source software used in its development makes its possible to deploy and non-commercially use the system on various software and hardware platforms running under the Linux family OS.

Acknowledgments

This work was performed under state budget theme no. AAAAA-A17-117013050037-0. The integration of new software tools for statistical analysis of time series of spatial climatological datasets was performed under ICT SB RAS state target program no. 0316-2018-0002.

References

[1] Big Data Definition MIKE 2.0 http://mike2.openmethodology.org/wiki/Big_Data_Definition
[2] Kaczmarszyk D 16 February 2010 What is "Big Data?" ZDNet http://www.zdnet.com/article/what-is-big-data/
[3] Vance A 22 April 2010 Start-Up Goes After Big Data With Hadoop Helper New York Times Blog http://bits.blogs.nytimes.com/2010/04/22/start-up-goes-after-big-data-with-hadoop-helper/?dbk
[4] Kalinichenko L A et al 2016 Problems of access to data in Data Intensive Domain in Russia Informatika i ee primeneniya 10 3-23
[5] Hilbert M 2016 Big Data for Development: A Review of Promises and Challenges Development Policy Review 34 (1) 135-74 doi: http://doi.org/10.1111/dpr.12142
[6] Koshkarev A V 2008 Geoportal as a tool to control spatial data and services Spatial data 2 6-14
[7] Candela L, Castelli D and Pagano P 2013 Virtual Research Environments: An Overview and a Research Agenda Data Science Journal 12 GRDI75–GRDI81 DOI:10.2481/dsj.GRDI-013

[8] Bart A, Fazliev A, Gordov E, Okladnikov I, Privezentsev A and Titov A 2018 Virtual Research Environment for Regional Climatic Processes Analysis: Ontological Approach to Spatial Data Systematization Data Science Journal 17 14 DOI: 10.5334/dsj-2018-014

[9] Allan R J 2009 Virtual Research Environments: from Portals to Science Gateways (Chandos: Oxford Publishing)

[10] Yakubailik O E 2007 Geoformation geoportal J. Computational Technologies (Special Issue 3) 12 116-25

[11] Dragicevic S, Balram S and Lewis J 2000 The role of Web GIS tools in the environmental modeling and decision-making process 4th Int. Conf. on Integrating GIS and Environmental Modeling (GIS/EM4): Problems, Prospects and Research Needs (September 2-8, 2000, Banff, Alberta, Canada) http://www.srcosmos.gr/srcosmos/showpub.aspx?aa=5823

[12] Frans J M van der Wel 2005 Spatial data infrastructure for meteorological and climatic data Meteorol. Appl. 12 7-8

[13] Samet R and Tural S 2010 Web based real-time meteorological data analysis and mapping information system WSEAS Transactions on Information Science and Applications 7(9) 1115-25

[14] Berrick S W, Leptoukh G, Farley J D and Rui H 2005 GIOVANNI: A Web service workflow-based data visualization and analysis system IEEE Trans. Geosci. Remote Sens. 47(1) 106-13

[15] Shiklomanov A I 2012 Regional Integrated Monitoring System for the Pan-Arctic (RIMS) Proc. of the Northern Eurasian Earth Science Partnership Initiative (NEESPI) Regional Science Team Meeting devoted to the High Latitudes (June 2-6, 2012, Helsinki, Finland) pp 128-30

[16] Shiklomanov A, Prusevich A, Gordov E, Okladnikov I and Titov A 2016 Environmental science applications with Rapid Integrated Mapping and analysis System (RIMS) IOP Conf. Series: Earth and Environmental Science 48 012034 DOI: 10.1088/1755-1315/48/1/012034

[17] Welp L R, Randerson J T, Finlay J C, Davydov S P, Zimova G M, Davydova A I and Zimov S A 2005 A high-resolution time series of oxygen isotopes from the Kolyma River: Implications for the seasonal dynamics of discharge and basin-scale water use Geophys. Res. Lett. 32 L14401 DOI: 10.1029/2005GL022857

[18] Gordov E P, Okladnikov I G, Riazanova A A, Titov A G and Gordova Yu E 2018 A virtual thematic research environment for analysis of climate change and its consequences for Northern Eurasia: the current state of the art and perspectives IOP Conf. Series: Earth and Environmental Science 211 012079 DOI: 10.1088/1755-1315/211/1/012079

[19] Gordov E P, Okladnikov I G, Titov A G, Voropay N N, Riazanova A A and Lykosov V N 2018 Development of Information-computational Infrastructure for Modern Climatology Russian Meteorology and Hydrology 43(11) 722–8 DOI: 10.3103/S106837391811002X

[20] Titov A G and Okladnikov I G 2014 Architecture of web-GIS for climate monitoring based on geospatial data services Vestnik NGU. Series Information technologies 12(1) 79-88

[21] Okladnikov I G, Gordov E P, Riazanova A A and Titov A G 2018 Development of a “cloud” thematic software complex for analysis of climatic and environmental changes IOP Conf. Series: Earth and Environmental Science 211 012081 DOI: 10.1088/1755-1315/211/1/012081

[22] Okladnikov I G, Gordov E P and Titov A G 2016 Development of climate data storage and processing model IOP Conf. Series: Earth and Environmental Science 48 012033 DOI:10.1088/1755-1315/48/1/012033

[23] Ryazanova A A, Okladnikov I G and Gordov E P 2017 Integration of modern statistical tools for the analysis of climate extremes into the web-GIS “CLIMATE” IOP Conf. Series: Earth and Environmental Science 96 012014 DOI: 10.1088/1755-1315/96/1/012014

[24] Gilleland E and Katz R W 2016 extRemes 2.0: An Extreme Value Analysis Package in R Journal of Statistical Software 72(8) 39 DOI: 10.18637/jss.v072.i08

[25] Yan J 2007 Enjoy the Joy of Copulas: With a Package copula Journal of Statistical Software 21(4) 21 DOI: 10.18637/jss.v021.i04