Virtual Research Environment in Fundamental and Applied Problems of Climatology

Evgeny P. Gordov[1,2], Igor G. Okladnikov[1,2], Alexander G. Titov[1,2], Anna A. Ryazanova[1]

[1] IMCES SB RAS, Tomsk, Russia, post@imces.ru
[2] ICT SB RAS, Novosibirsk, Russia, ict@ict.nsc.ru

Abstract. The description and new features of a developed virtual computing and information environment for analysis, assessment and prediction of consequences of global climate changes for ecosystems and climate in the selected region are presented.

Keywords: virtual research environment; big environmental datasets; climate change.

1   Introduction

Recent studies have shown that science is becoming increasingly global, multipolar and distributed [1]. This leads to the need of development of modern, open and universal research software products providing remote and shared access to various data archives, as well as their "cloud" processing and analysis by distributed multidisciplinary teams of scientists. Such products are currently usually called virtual research environments [2], although they have many other names: scientific gateways, cyberinfrastructure, etc. According to the work of L. Candela [3], the term virtual Research Environment (VRE) defines a software package with the following main characteristics: (i) it is a working environment accessible via the Internet, (ii) designed to meet the needs of the target community, (iii) providing it with the products and tools necessary to achieve given goals, and (iv) facilitate the exchange of research results.

In Earth sciences, such an approach to solving scientific problems has already become necessary: the corresponding tasks are essentially interdisciplinary; they are solved by geographically distributed research teams, and the volume of data sets reaches tens of petabytes (http://newsletter.copernicus.eu/article/data-volume). Current climate change, especially its extreme manifestations, have an increasing impact on economic, political and social processes [4-6]. A reliable analysis of these processes is important for developing adequate adaptation strategies and mitigating their negative effects (for example, for agriculture, forestry or planned infrastructure).

A reliable analysis of climate and environmental changes and society's reactions to them re-quire skills in working with bulky meteorological data sets, abilities to interact with powerful computing resources and complex numerical models, and knowledge of modern methods of statistical analysis for dealing with large model archives. These skills are not typical for specialists in the field of economic, political and social sciences dealing with phenomena and processes that are strongly influenced by climate change. Unfortunately, such skills are completely uncharacteristic for decision-makers, including those responsible for developing adaptation measures. Therefore, the development of VRE should be aimed at providing professionals and decision-makers with reliable tools for studying the economic, political and social consequences of climate change.

The article presents the new functionality of the thematic VRE to support climate research. This VRE is based on a specialized web-GIS "Climate" (http://sclimate.scert.ru), which is a cross-platform client-server software package with open source code that combines web and GIS functionality, supports cloud-based processing and analysis of geospatial climate data, and provides all its of functionality in the window of a modern Internet browser on a user's workstation. The developed VRE will provide multidisciplinary distributed research groups that are not experts in the field of information technology (climatologists, ecologists, biologists and decision-makers) with accessible and reliable online tools for quick analysis and visualization of multidimensional heterogeneous climate datasets obtained from various sources.

2   Thematic virtual research environment

The developed VRE is intended for processing, analysis and visualization of geospatial data sets in Earth sciences field. It is created on the basis of the web GIS "Climate" developed at IMCES SB RAS [7–9], which consists of three key components: a server-side computing core; server-side middleware, represented by a geoportal and a set of PHP controllers; and a web client based on AJAX technology and developed using a specialized JavaScript library containing widgets for a typical graphical user interface. To process geospatial data sets, a set of developed software

modules of the computing core is used. The results of processing geospatial data are presented on the map of the selected area with overlapping raster and vector cartographic layers, accompanied by the corresponding metadata. The architecture and functioning of the VRE are presented in detail by A. Bart [10].

The functionality of the VRE includes basic and comprehensive statistical analysis of data, and the elements of the geographic information system (GIS) give users an opportunity to combine and display results with a geographic reference on a selected cartographic basis. VRE pro-vides professionals and users without programming skills with reliable and convenient online tools for a comprehensive study of climate change and ecosystems in a single web user interface.

Some examples of the successful application of the developed VRE in studies of the observed climate changes in Siberia and their consequences are discussed in several works [11−13].

3 Climate research at the regional level

The use of special statistical methods is required to describe extreme climatic phenomena. A correct statistical description of extreme precipitation and temperature can be obtained using extreme value statistics (EVS) [14]. The software implementation of EVS in the R language (the "extRemes" package) in the form of specialized software modules allowed to perform statistical analysis of the maximum values of selected meteorological values based on a non-stationary generalized distribution of extreme values. Values needed for risk assessment can be calculated using a distribution function that describes the likelihood that the value of an observed variable will exceed a certain level. These levels are often expressed as r_T return levels for a given return period T. The r_T value is defined as the level that, on average, exceeds each T, that is, with a probability of 1 / T.

After integrating the developed software modules into the web GIS "Climate", this new functionality was tested on the basis of 100-year return levels for the maximum precipitation in July based on ECMWF ERA-Interim data (Figure 1a) and APHRODITE JMA (Figure 1b) for South Siberian region (52.5-60 ° N, 75-95 ° E). Although the figures show similar behavior of the calculated characteristic in general, the JMA APHRODITE results (Figure 1b) demonstrate more details and higher values in individual regions. This is due to different spatial resolutions and different sources of the analyzed data sets.
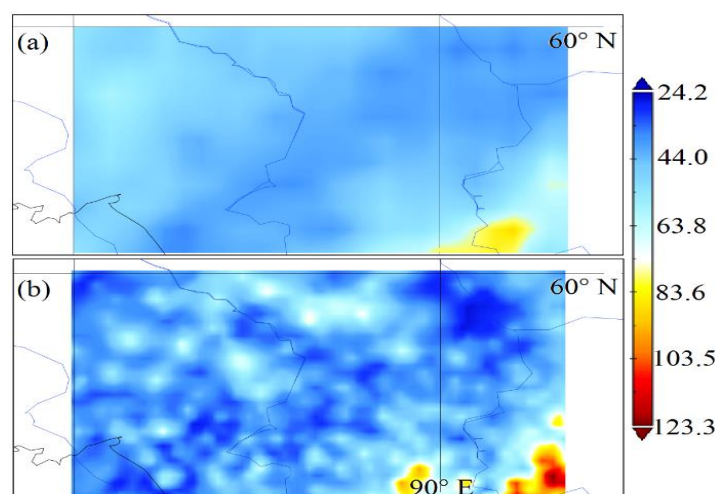

Figure 1. 100-year return levels for maximum precipitation in July for the South Siberian region:
a - ECMFW ERA-Interim data, horizontal grid 0.75x0.75, 1979-2007,
b - JMA APHRODITE data, horizontal grid 0.25x0.25, 1979-2007.

Trend analysis in meteorological observations is one of the most common approaches to climate change research. Quantile regression provides a well-defined statistical basis for estimating the rate of change, not only for the mean, as in a regular regression, but also in other parts of the data distribution function. If a random variable Y with a continuous cumulative distribution function FY (y) is given, the quantile function QY (τ) is determined from FY (y) as QY (τ) = FY-1 (τ). A quantile is defined as a value of QY (τ) such that P [Y ≤ QY (τ)] = τ, $0 \leq \tau \leq 1$. Then, given the conditional distribution of Y with X = x, the conditional quantile function QY | X (τ; x) checks P [Y ≤ QY | X (τ; x) | X = x] = τ. While the usual regression is based on the conditional function of the mean E [Y | X = x] and the minimization of corresponding residuals, the quantile regression is based on the conditional function of the quantile and the minimization of the sum of asymmetrically weighted absolute residuals $\sum_{i=1} \rho(\tau)|y_i - Q_{Y|X}(\tau; x = x_i)|$, where ρ is the function of the absolute value. The procedure for calculating quantile regression was implemented in the R language ("quantreg" package) [15] and integrated into the VRE. The range of quantile values of interest is set between 0 and 1.

To test the new VRE functionality, using ECMWF ERA 40 reanalysis data, January maximum temperature trends for the South Siberian region (50-65 ° N, 60-120 ° E) were calculated. The results obtained show (Figure 2) that the

maximum January temperature on the 0.05 quantile changed (both downward and upward) to a greater extent than on the 0.5 and 0.95 quantiles over almost the entire territory under consideration. At the same time, on the 0.95 quantile, the temperature changed to a lesser extent.
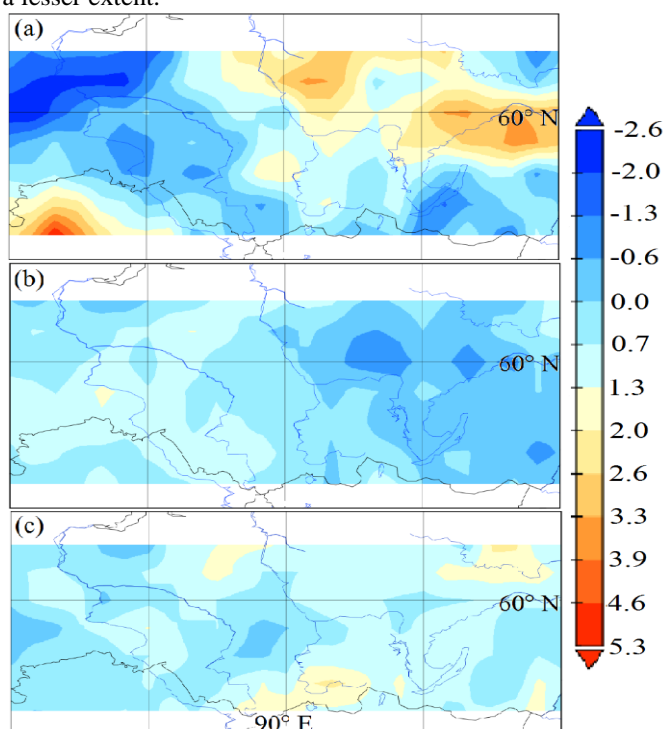

Figure 2. January maximum temperature trend based on ECMWF ERA 40 reanalysis data (horizontal grid 2.5x2.5, 1961-2002): a - quantile 0.05, b - quantile 0.5, c - quantile 0.95.

To calculate the Pearson correlation coefficient for average daily, monthly, seasonal, and annual values of meteorological variables a new software module was integrated into the software package. The considered variables are preliminarily reduced to the same temporal grid selected by a user. The correlation coefficient is then calculated for meteorological variables from the same or different data sets. If the variables are defined on different spatial grids, they are interpolated to the grid with a higher spatial resolution. In Figure 3 a special form for calculating the correlation coefficient between two parameters (air temperature at 2 m from the ERA-Interim reanalysis data sets and weather stations) is shown. If both weather station data (given on a scatter grid) and gridded data are used in a calculation, gridded data is interpolated into coordinates of weather stations.

4   Enhanced web portal functionality

The average annual temperature in the territory of the Russian Federation continues to grow. Current estimates show that extremely high temperatures and summer droughts caused by global warming will cause serious damage in some regions of Russia. They negatively affect crop yields and the state of forestry. The thawing of permafrost in the north will accelerate, and floods in some regions will intensify. To develop an effective adaptation strategy and measures to reduce the negative consequences of extreme climatic phenomena, accurate knowledge of the geography of extreme climatic phenomena, their frequency and intensity are necessary. Since the frequency of such events is small, to obtain the necessary information, it is necessary to analyze these phenomena using modern probabilistic-statistical methods and detailed meteorological information accumulated during the period of instrumental observations in the studied region.

To provide regional decision-makers with information necessary for their target activities, for the Siberian region a set of relevant climatic characteristics was calculated using the web GIS "Climate". On its basis, an open archive was created and integrated into the VRE for the subsequent analysis of current climate changes by decision-makers in the region. This archive provides quantitative reference materials for assessing future climate and environmental risks and adapting regional development policies to these risks (Figure 4).

The archive presents various characteristics of air temperature and precipitation recommended by WMO (World Meteorological Organization) for the analysis of extreme climatic phenomena. These characteristics provide information on the maximum/minimum values of temperature and precipitation, information on the frequency and duration of various extreme events, determine the number of days when the temperature or precipitation exceeds a certain threshold (abnormal heat / cold wave, abnormal precipitation, etc.).
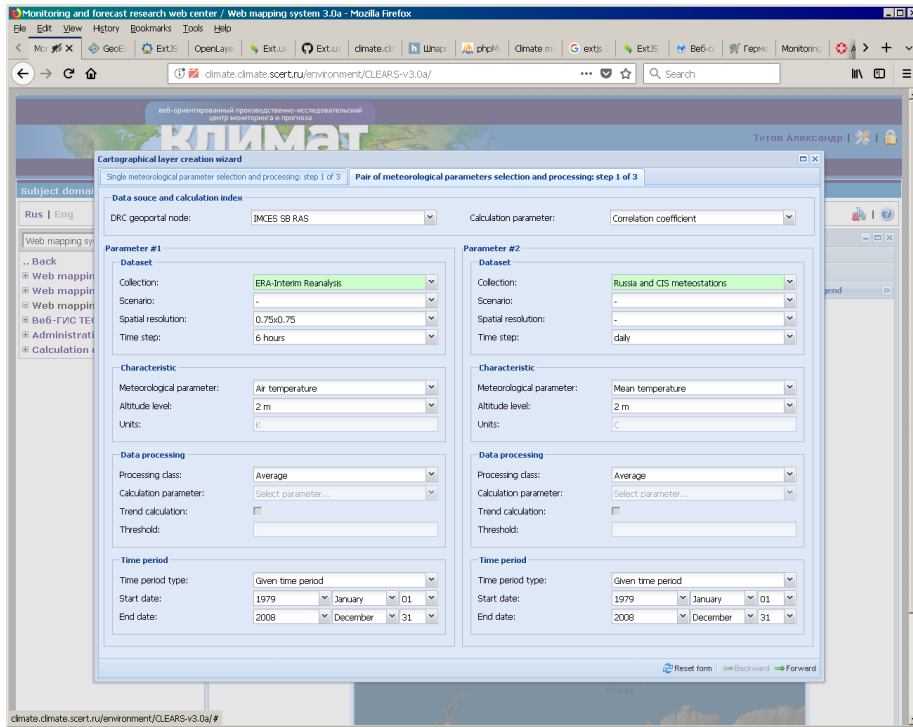
Figure 3. Selection of a pair of meteorological parameters and procedures for their joint analysis.

The archive also has links for downloading files in various formats (netCDF, GeoTIFF, WMS and WFS) with calculated characteristics. If there is a need for further work, users can download the required files and use them in a third-party software.

Figure 4. A web page with links to sets of relevant climate characteristics, calculated for the Siberian region.

5   Conclusion

Tasks considered in Earth sciences and, in particular, climatology, are usually associated with the use of large archives of geo-referenced data. For their effective analysis and obtaining useful information, the integration of web and GIS technologies, data, processing and visualization tools into distributed, accessible via the Internet, thematic software systems is required. This paper presents the development of a thematic VRE for analysis of meteorological and climatic processes. It is a free, cross-platform, composite open-source software package that allows performing "cloud" analysis of climate data in a window of an Internet browser. The flexible modular architecture of the VRE makes it easy to add new computing nodes, data storage systems, and provides a reliable computing infrastructure for regional studies of climate change based on modern web and GIS technologies. This thematic VRE will provide multidisciplinary distributed teams of researchers which are not experts in the field of information technologies (climatologists, ecologists, biologists and decision-makers) with easily accessible reliable online tools for reliable analysis and visualization of multidimensional heterogeneous sets of climate data obtained from various sources. The ability to obtain analysis results in GeoTIFF, ESRI Shapefile, Encapsulated PostScript, CSV, XML, netCDF, float GeoTIFF formats opens up the possibility for researchers to perform further data analysis using their own software.

References

[1] Llewellyn Smith C., Borysiewicz L., Casselton L., Conway G., Hassan M., Leach M. Knowledge, Networks and Nations: Global Scientific Collaboration in the 21st Century. UK: The Royal Society, 2011.

[2] Carusi A., Reimer T. Virtual Research Environment Collaborative Landscape Study. JISC, 2010.

[3] Candela L., Castelli D., Pagano P. Virtual Research Environments: An Overview and a Research Agenda // Data Science Journal. 2013. Vol. 12. GRDI75–GRDI81. DOI: http://doi.org/10.2481/dsj.GRDI-013.

[4] IPCC. Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change. Cambridge, UK: Cambridge University Press, 2012.

[5] IPCC. Fifth Assessment Report 'Climate Change 2013'. Cambridge, UK: Cambridge University Press, 2013.

[6] Sillmann J., Donat M.G., Fyfe J.C., Zwiers F.W. Observed and simulated temperature extremes during the recent warming hiatus // Environmental Research Letters. 2014. Vol. 9(6). P. 064023.

[7] Okladnikov I.G., Gordov E.P., Titov A.G., Shulgina T.M. Information-computational System for Online Analysis of Georeferenced Climatological Data. // XVII International Conference on Data Analytics and Management in Data Intensive Domains (DAMDID/RCDL 2015). Selected papers / CEUR Workshop Proceedings. 2015. Vol. 1536. P. 76-80.

[8] Okladnikov I.G., Gordov E.P., Titov A.G. Development of climate data storage and processing model // IOP Conference Series: Earth and Environmental Science. 2016. V. 48. P. 012030. DOI: https://doi.org/10.1088/1755-1315/48/1/012030.

[9] Gordov E., Shiklomanov A., Okladnikov I., Prusevich A., Titov A. Development of Distributed Research Center for analysis of regional climatic and environmental changes // IOP Conf. Series: Earth and Environmental Science. 2016. V. 48. P. 012033. DOI: http://dx.doi.org/10.1088/1755-1315/48/1/012033.

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

[11] Riazanova A.A., Voropay N.N., Okladnikov I.G., Gordov E.P. Development of computational module of regional aridity for web-GIS 'Climate' // IOP Conference Series: Earth and Environmental Science. 2016. Vol. 48. P. 012032. DOI: https://doi.org/10.1088/1755-1315/48/1/012032.

[12] Ryazanova A.A., Voropay N.N. Droughts and Excessive Moisture Events in Southern Siberia in the Late XXth - Early XXIst Centuries // IOP Conference Series: Earth and Environmental Science. 2017. Vol. 96. P. 012015. DOI: https://doi.org/10.1088/1755-1315/96/1/012015.

[13] Shulgina T.M., Gordov E.P., Genina E.Yu. Dynamics of climatic characteristics influencing vegetation in Siberia // Environmental Research Letters. 2011. Vol. 6. P. 045210.

[14] Coles S.G. An Introduction to Statistical Modelling of Extreme Values. London: Springer, 2001.

[15] Koenker R., Portnoy S., Tian P., Zeileis A., Grosjean P., Ripley B.D. Package "quantreg". The Comprehensive R Archive Network (CRAN), 2017. https://cran.r-project.org/web/packages/quantreg/quantreg.pdf.