Modern Information Technologies in Environmental Sciences: Climate Wikience

Ramon Antonio Rodriges Zalipynis, Olga A. Ivashchuk, Igor S. Konstantinov

Belgorod National Research University, Belgorod, Russian Federation

rodriges@wikience.org, ivaschuk@bsu.edu.ru, konstantinovi@mail.ru

Abstract—Climate Wikience is a desktop application for fast 3D visualization and analysis of retrospective climate reanalysis and Earth remote sensing data. For its several distinct features and certain tasks an analyst may prefer it to other tools. The features include rich collection of environmental variables readily available "out-of-the-box', one-click 3D visualization of any of them regardless of their storage formats and coordinate systems, tight R integration from which the data are accessible in native R data types. Large volumes of the data are stored on computer cluster by ChronosServer – new distributed file-based raster database. Only small portions of the data required for visualization and analysis are automatically delivered over the Internet to Climate Wikience in near real-time. The paper surveys the data available from Climate Wikience, its visualization capabilities and air pollution analysis as a case study. Climate Wikience is free for download at www.wikience.org

Keywords—climate reanalysis; Earth remote sensing data; 3D visualization; R connectivity.

I. Introduction

Climate reanalysis and Earth remote sensing data are becoming one of the most comprehensive resources for environmental studies. The data are attractive for powerful data assimilation and quality assurance techniques used for their production, global coverage and ever increasing spatial and temporal resolution.

Although the data are free, they have large volumes (the order of terabytes) stored in powerful but complex scientific file formats (like NetCDF, HDF, Grib) within different coordinate systems. Moreover, no single tool exists that seamlessly integrates distributed storage of large file-based data, near real-time access to them, 3D visualization and analysis.

Today, a time and effort distance must be overcome before the data could be actually used for research. This usually includes the search and download of files with the required data, development of computer programs reading the data from scientific file formats with complex naming. Besides supporting large collection of files on the workstation, the process is itself labor intensive, distracts from the initial research goals and delays results.

Climate Wikience hides the actual volumes of data from the user, the number of files, their formats, naming, coordinate systems, compression and other details allowing to concentrate on the data itself and start working with them as quickly as possible by easy to perceive visual means and powerful analytic tools like R.

There are a large number of tools tailored for working with gridded climate reanalysis and Earth remote sensing data. However, only few of them can be configured to organize data storage, access, visualization and analysis chain similar to Climate Wikience.

The THREDDS Data Server (TDS) [1] provides uniform data model over files of different formats and delivers data over various network protocols including OPeNDAP, WMS, WCS, and HTTP. However, it does not support large data volumes since deployed as war package inside a Tomcat or Glassfish web server.

GeoServer freely available at http://geoserver.org/ and Image Extension to commercial ArcGIS also directly support NetCDF files. However, to date, only ChronosServer [2] provides access to data distributed over nodes of a computer cluster.

A general purpose visualization tool ParaView [3] works with local, but not remote data. NOAA's Weather and Climate Toolkit (WCT) [4] has only 2D visualization with export to KML files for further view in Google Earth. The Integrated Data Viewer (IDV) [5] often delays the exploration process by downloading complete files of large volumes. NASA Giovanni [6] is 2D web plotting tool. None of them provides integration with R.

There are also powerful commercial products like MatLab [7] capable reading NetCDF files though still not having 3D globe visualization.

The overall Climate Wikience usage workflow is quite simple. Upon the application start, the user is presented the single tree view with available datasets ("Layers" tab, Fig. 1). The hierarchical representation helps to quickly check what's available and locate a required dataset.

It is sufficient to click the checkbox left to the dataset name to immediately trigger its 3D visualization. Clear and compact time slider shows the complete time interval for which the data for the selected dataset are available. A mouse drag is sufficient to precisely set the month, date and time for 3D visualization.

The "Properties" tab contains available visualization options for the selected dataset. The form of name-value pairs allows to quickly check and adjust what's available.
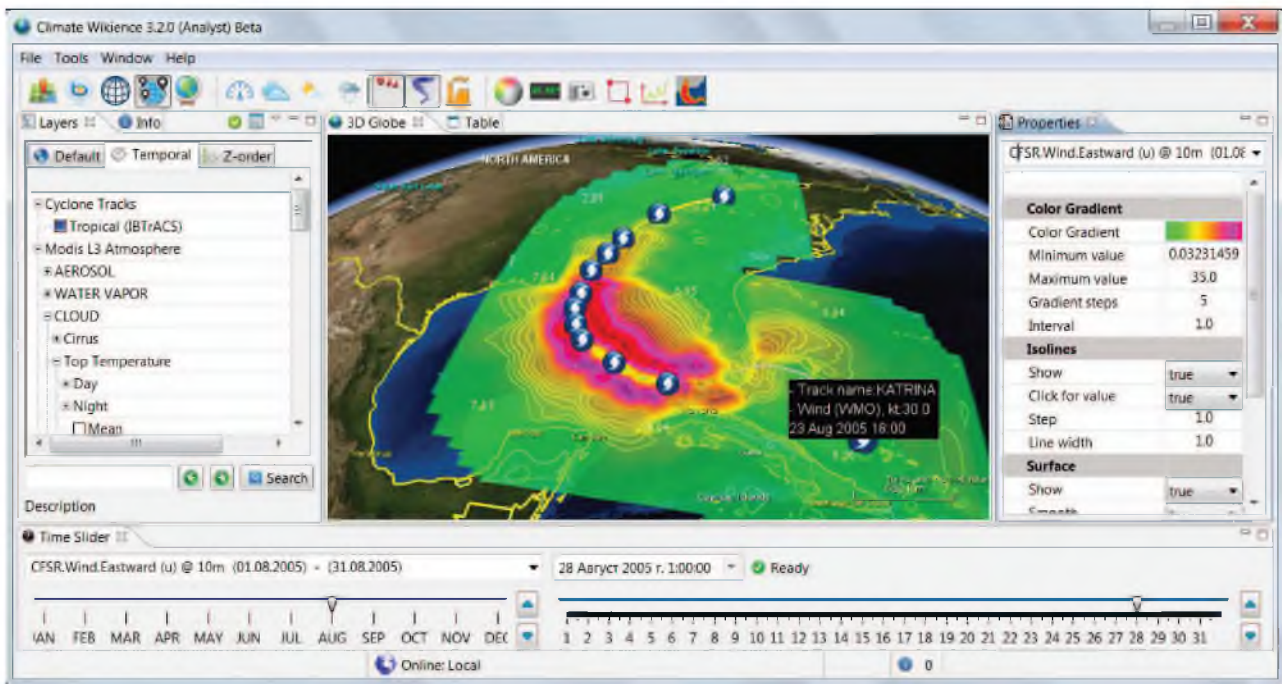
Figure 1. Climate Wikience interface with Layers tab, Properties tab and Time slider. The 3D globe shows hurricane Katrina track and its wind speed swath generated using CFSR climate reanalysis (best viewed in color).

II. Data and Visualization

Currently Climate Wikience concentrates on global gridded data with both regular and irregular grids. Climate reanalysis are already distributed in a gridded form. Only Earth remote sensing data of processing level 3 are gridded and, thus, provided by Climate Wikience.

Climate Wikience implements three ways to visualize gridded data: markers, isolines and surfaces. The details on these options are given further in this section.

A. Climate Reanalysis

Climate Wikience maintains a subset of variables taken from the climate reanalysis listed in Table I.

TABLE I.   Climate Reanalysis Data

Short Name	Agency	Spatial Resolution	Period
NCEP/NCAR R2 [8]	NCEP	$2.5° \times 2.5°$	1979–current
ERA-Interim [9]	ECMWF	$1.5° \times 1.5°$	1979–current
MERRA [10]	NASA	$2/3° \times 1/2°$	1979–current
20CR [11]	NOAA	$2.0° \times 2.0°$	1869 – 2012
CFSR [12]	NCEP	$0.25° \times 0.25°$	1979 – 2009

The spatial resolution is given for the regular latitude/longitude grid. Some CFSR variables are on a Gaussian grid. In this case, ChronosServer interpolates the data onto regular $0.25° \times 0.25°$ grid using bicubic spline interpolation before it delivers data to Climate Wikience.

The common temporal resolution for climate reanalysis variables is 6 hours. Some variables, especially in MERRA and CFSR, have time step of 1 hour.

Climate Wikience draws markers at the grid nodes where variable data are located. Sizes and colors of markers are proportional to data values (Fig. 2). The number of visible markers depends on the zoom level of the 3D globe model to avoid clutter. Markers appear and disappear when zooming in and out respectively.
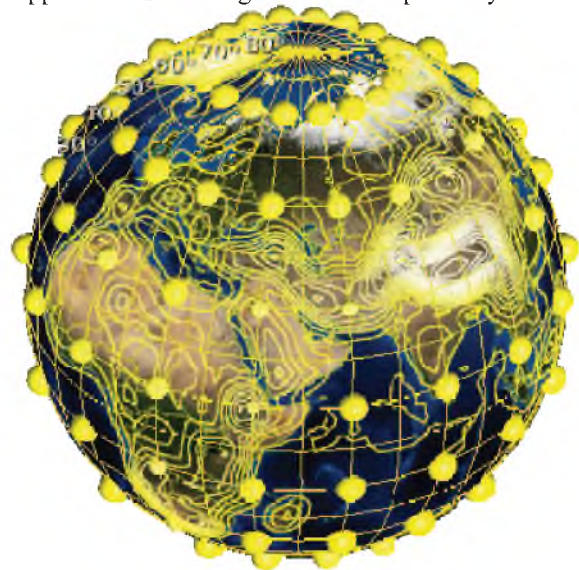

Figure 2. Markers and isobars for surface pressure on the 1st of August, 2000 from NCEP/NCAR Reanalysis 2. Markers are shown at $20° \times 20°$ grid. Isobars are plotted every 2500 Pa.

Isolines are built on-the-fly using new algorithm [13]. Unlike other isoline plotting algorithms, the proposed method represents isolines by polygons instead of separate segments. The polygons are guaranteed closed

and their number is almost 200 times less on usual than the number of segments for climate reanalysis grids. This significantly reduces the load on the visualization engine resulting in much faster performance.

Initially isolines are displayed without captions. A click on any of its points will place a caption with the data value used to build isoline (when "click for value" option is enabled at Properties tab).

The surface visualization is also possible for climate reanalysis data but we demonstrate it in the following subsection in order not to obscure the Fig. 2.

B. Air Pollution Earth Remote Sensing Data

Climate Wikience maintains the collection of air pollution data derived from measurements of remote sensing platforms (Table II).

TABLE II.   AIR POLLUTION DATA

Parameter	Satellite/ Instrument	Spatial Resolution	Start date
Sulfur dioxide (SO_2)	Aura/ OMI	0.25° × 0.25° 27.5 × 18 km	01 Oct. 2004
Nitrogen dioxide (NO_2)			
Ozone (O_3)			
Optical aerosol thickness (at 0.55 μm)	Terra/ MODIS	1.0° × 1.0° 110 × 72 km	02 Mar. 2000
Ozone (O_3)			
Carbon oxide (CO)	Terra/ MOPITT	1.0° × 1.0° 110 × 72 km	03 Mar. 2000
Carbon dioxide (CO_2)	Aqua/ AIRS	2.0° × 2.5° 220 × 144 km	01 Jan. 2004

The spatial resolution in kilometers is approximately given for the 48° latitude. All datasets are available daily since "Start date" till current.

The same types of visualization (markers, isolines, surfaces) as for climate reanalysis are applicable for the collection of air pollution data.

A surface for a variable on a $x{\times}y$ grid is drawn as the set of $x{\times}y$ grid cells. The color is constant throughout a cell. It is calculated depending on the variable value for the grid cell and the color scheme that can be adjusted at "Color gradient" section of Properties tab. Grids of Earth remote sensing data frequently contain cells with missing values. Climate Wikience will not draw those cells making their area transparent (Fig. 3).
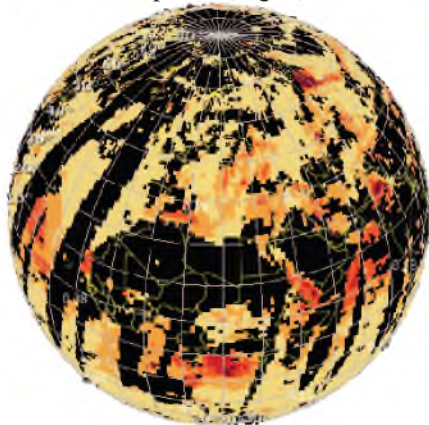

Figure 3.   Terra MODIS aerosol optical thickness on the 1st of August, 2005 (1.0° × 1.0°).

The surface interpolation may be optionally enabled. In this case, the cell area is filled with colors interpolated from the data values located at its neighboring grid nodes.

C. Terra MODIS Earth Remote Sensing Data

Over 700 variables from MODIS Atmosphere dataset [14] are accessible from Climate Wikience. They include extensive cloud properties: top temperature, pressure, height, optical thickness, effective emissivity, phase, and reflectance. The spatial and temporal characteristics of the dataset are the same as for Terra/MODIS ozone (Table II).

D. Time Series Data

A single data file of climate reanalysis or level 3 Earth remote sensing data stores global grids for a short time interval. The number of files for complete time interval often reaches several thousand. This makes impossible a quick retrieval of a time series for a region.

Today, Climate Wikience is the only system capable to quickly provide time series for a separate latitude/longitude grid cell [15]. ChronosServer generates time series in advance and saves them on a persistent storage for further reuse. Appropriate storage layout is exploited to enable immediate retrieval of a single time series for a given coordinate.

The availability of time series for each grid cell makes possible to carry out precise investigation of territories with areas comparable to the cell area, for example, cities or enterprise impact zones.

E. Tropical Cyclone Tracks

The data are from IBTrACS – International Best Track Archive for Climate Stewardship [16]. The dataset covers period since the 1st of January, 1848 till present.

The NOAA hurricane mapper [17] also visualizes tropical storm tracks, although in 2D. Climate Wikience also visualizes auxiliary data in addition to tracks. For example, wind speed from a climate reanalysis or cloud top pressure from Earth remote sensing collection.

Fig. 1 shows a wind speed "swath" generated by ChronosServer for hurricane Katrina (August, 2005) using CFSR wind speed for 10 meters above the surface. For each date during the hurricane, ChronosServer extracts 12°×12° region for eastward (u) and northward (v) wind components around a hurricane center, calculates wind speed, uses bicubic spline interpolation to derive 0.25° × 0.25° grid from Gaussian grid, overlays obtained regions and selects maximum wind speed for a grid point among intersecting regions.

III.   ANALYSIS

The R project [17] with its 5000 packages to date is used worldwide for statistics, modeling, spatial and time series analysis, sophisticated visualization and many other tasks. The R environment has its own programming language "R" and several development environments, e.g.

RStudio[1]. Anyone can create a package for R using C/C++/Java/Fortran/R or other programming language. The R was chosen to serve as analytic engine for Climate Wikience because it is extensible, free, popular and has large responsive community.

The leading author of this paper developed R package called RWikience[2] to connect Climate Wikience and R. RWikience enables access to all available data in Climate Wikience directly from R. RWikience converts the retrieved information to native data types of the R language to facilitate its usage in other R packages.

RWikience package must be installed into R to enable the access to data provided by Climate Wikience.

The following subsections contain several case studies that provide evidence of how RWikience simplifies the retrieval and further analysis of data in contrast to working with thousands of files in scientific formats.

A. Time series analysis

Climate Wikience is able to provide time series of a variable for a separate grid cell (recall the subsection D of section II). Several lines of R code are sufficient to retrieve a time series using RWikience and readily perform its further analyses in R.

In the past, this was quite labor-intensive because this required ad-hoc programs to iterate over a voluminous collection of large number of files in diverse scientific formats, coordinate systems and complex naming.

Since we are able to operate with a time series for a single grid cell, it is possible to plot the aerosol optical thickness over a city measured by Terra MODIS (Fig. 4).

```
1   library(RWikience)
2   w <- WikienceConnect()
3   queryTimeSeries(w,
       "Modis L3 Atmosphere.AEROSOL.LAND AND
       OCEAN.Optical Depth.Maximum",
       "02 03 2000", "04 10 2012", 41, 12, 41, 12)
4   t <- getTimeSeries(w, 0)
5   t_avg <- aggregate(t["value"],
                    format(t["date"], "%Y-%m"),
                    mean, na.rm=TRUE)
6   t_avg$date <- seq(min(t$date), max(t$date),
                    length = nrow(t_avg))
7   plot(t_avg$date, t_avg$value, type="l")
```

Figure 4.   R code for retrieving and plotting aerosol optical thickness time series over Rome (Italy) during 02 Mar. 2000 – 04 Oct. 2012.

The standard R function "library" loads the RWikience package (line 1). Lines 2–4 use functions only from the loaded RWikience package (WikienceConnect, queryTimeSeries and getTimeSeries).

RWikience connects to Climate Wikience from R (line 2). We initiate time series retrieval for time interval 02 Mar. 2000 – 04 Oct. 2012 and region consisting of a single $1.0° \times 1.0°$ grid cell covering the city of Rome,

Italy (line 3). At this point, Climate Wikience queries ChronosServer to extract the required time series.

The resulting time series is loaded into R (line 4). The queryTimeSeries function allows specifying a geographical region that covers several grid cells. This results in several time series (one for each grid cell). Thus, the second argument to getTimeSeries function denotes the index of the time series in the resulting collection. In the current case we have only a single time series with index 0. The retrieval of a time series for any other variable is as straightforward: pass its name, time interval and bounding region to the queryTimeSeries function.

The getTimeSeries function returns a data frame – native R data type to facilitate usage of time series in R functions from other packages. The data frame contains two columns – date and value of time series for that date.

The next part of code (lines 5–7) uses standard R functions (not from RWikience) to calculate monthly means from daily values (lines 5–6) and plot the line chart of the monthly means (line 7 and Fig. 5). The aerosol optical thickness is unitless ranging from -0.05 (negative due to noise) to 5.0.
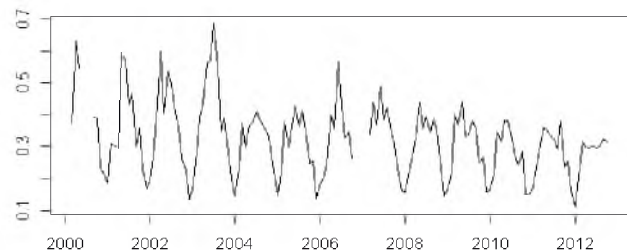

Figure 5.   Monthly aerosol optical thickness over Rome, Italy.

The line on Fig. 5 is not continuous since time series values are missing for several winter months mainly due to clouds (prevent sensing aerosol from space).

B. Air pollution risk for Europe

This section contains more complicated analysis example using RWikience and R. RWikience is used to access time series of 7.5 years of daily nitrogen dioxide (NO_2) concentrations that are available for each $0.25° \times 0.25°$ grid cell across Europe. The time series are further processed in R to create the NO_2 pollution risk map for Europe that has both the highest spatial resolution and coverage to date. Remote sensing Aura OMI nitrogen dioxide data were used.

The proposed risk assessment method [19] derives the "typical" picture of air pollution over a given territory. Its need is motivated by the following observations.

First, daily satellite maps are highly cluttered due to gaps caused mainly by clouds. Second, the noise inherent to satellite measurements of trace gases distorts the overall pollution picture for a single day. Third, rare high pollution events must be filtered out and not contribute to "typical" air pollution picture. Forth, composite maps of several consequent daily maps or seasonal means of trace gas concentrations do not consider other periods of year.

Sometimes it is even impossible to select a period for creating a representative composite map [19].

The air pollution risk is defined as the probability of observing a pollutant concentration in a given interval over the territory under investigation [19].

The risk is calculated for each grid cell as $R(a, b)=S(a, b) / T$, where $S(a, b)$ is the number of days with pollutant concentration in $[a..b]$ interval and T is the number of days for which data are not missing (clouds or other reasons).

The air pollution risk map for nitrogen dioxide daily Aura OMI measurements between 01 Oct. 2004 and 20 June 2012 is shown on Fig. 6. The 0.2 DU..0.4 DU (Dobson Units) interval was experimentally selected as "moderate air pollution risk" for NO_2 [19].

The highest risks of air pollution by NO_2 in Europe are observed over Germany, Belgium, Netherlands and southern part of the North Sea as well as over large cities.

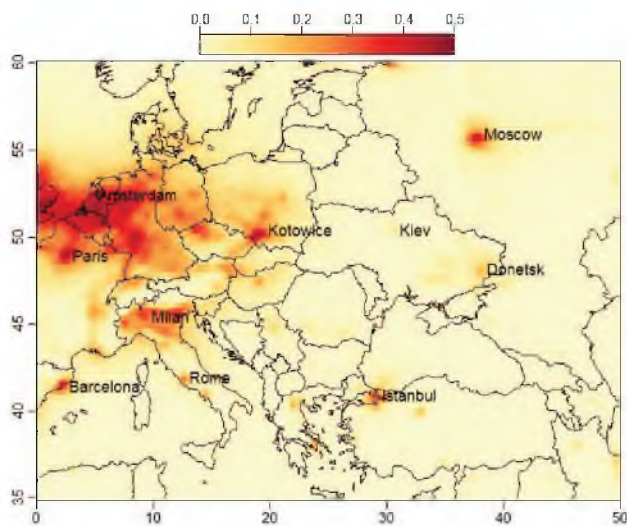The complete R code that uses RWikience and creates the risk map at Fig. 6 can be found at http://gis-lab.info/qa/climate-wikience.html.

Figure 6.   Map of moderate air pollution risk by nitrogen dioxide over the Europe with spatial resolution 0.25°×0.25° ($\approx$ 27,5×18 km).

IV.   CONCLUSIONS AND FUTURE WORK

Climate Wikience can be used in the operation of intelligent systems for biotechnosphere monitoring on the territories of any hierarchical scale (global, federal, regional, local). To accomplish this, Climate Wikience needs to be integrated with the decision support system endowed with data mining functions that help to transform monitoring data into rational management decisions to reduce and prevent ecologic risks.

These functions are implemented through the dynamic construction of the necessary mathematical and situational models: short, medium and long-term forecasting of the biotechnosphere dynamics, integrated assessment of the current and forecast environmental conditions, and formation of alternative management scenarios to assess their performance [20, 21].

REFERENCES

[1]   The THREDDS Data Server (TDS), http://www.unidata.ucar.edu/software/thredds/current/tds/TDS.html (last accessed 7 Feb. 2015).

[2]   R. A. Rodriges Zalipynis, "ChronosServer: real-time access to 'native' multi-terabyte retrospective data warehouse by thousands of concurrent clients," Informatics, Cybernetics and Computer Engineering, vol. 14 (188), 2011, pp. 151–161.

[3]   ParaView, http://www.paraview.org/ (last accessed 7 Feb. 2015).

[4]   NOAA's Weather and Climate Toolkit (WCT), http://www.ncdc.noaa.gov/wct/ (last accessed 7 Feb. 2015).

[5]   The Integrated Data Viewer (IDV), http://www.unidata.ucar.edu/software/idv/ (last accessed 7 Feb. 2015).

[6]   Giovanni – GES DISC Information Services Center http://disc.sci.gsfc.nasa.gov/giovanni/overview/index.html (last accessed 7 Feb. 2015).

[7]   MATLAB, http://www.mathworks.com/products/matlab/ (last accessed 7 Feb. 2015).

[8]   M. Kanamitsu, et al., "NCEP–DOE AMIP-II Reanalysis (R-2)," Bulletin of American Meteorological Society, no. 83(11), 2002, pp. 1631-1643.

[9]   D. P. Dee, et al., "The ERA-Interim reanalysis: configuration and performance of the data assimilation system," Quarterly Journal of the Royal Meteorological Society, no. 137 (656), 2011, pp. 553-597.

[10]   M. Rienecker, et al., "MERRA: NASA's modern-era retrospective analysis for research and applications," Journal of Climate, no. 24, 2011, pp. 3624-3648.

[11]   G. P. Compo et al., "The twentieth century reanalysis project," Quarterly Journal of the Royal Meteorological Society, vol. 137, issue 654, 2011, pp. 1-28.

[12]   S. Saha, et al., "The NCEP climate forecast system reanalysis," Bulletin of American Meteorological Society, no. 91, 2010, pp. 1015-1057.

[13]   R. A. Rodriges Zalipynis, "Efficient isolines construction method for visualization of gridded georeferenced data," Problems of Simulation and Computer Aided Design of Dynamic Systems, no. 10 (197), 2011, pp. 111-123.

[14]   MODIS Atmosphere Daily Global Product, http://modis-atmos.gsfc.nasa.gov/MOD08_D3/format_c51.html (last accessed 8 Feb. 2015).

[15]   R. A. Rodriges Zalipynis, "Representing Earth remote sensing data as time series," Journal of System Analysis and Information Technology in Environmental and Social Sciences, no. 1(2) – 2(3), 2012, pp. 135-145.

[16]   IBTrACS database, http://www.ncdc.noaa.gov/oa/ibtracs/ (last accessed on 5 Feb., 2015).

[17]   NOAA Hurricane Mapper, http://coast.noaa.gov/hurricanes (last accessed on 5 Feb., 2015).

[18]   The R Project for Statistical Computing, http://www.r-project.org/ (last accessed on 5 Feb., 2015).

[19]   R. A. Rodriges Zalipynis, "Ecologic assessment of air pollution by nitrogen dioxide over the territory of Europe using Earth remote sensing data," Journal of Informatics, Cybernetics and Computing Machinery, no. 1(19), 2014, pp. 126-130.

[20]   O. A. Ivashchuk, O. D. Ivashchuk, "Automation and intellectualization to control the ecological situation in the urbanized territories," in Proceedings of the IEEE 7th International Conference on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS'2013), Berlin, Germany, Sept. 12–14, 2013, vol. 2, pp. 814-820.

[21]   I. S. Konstantinov, O. A. Ivashchuk, "Approaches to creating environment safety automation control system of the industrial complex," // in Proceedings of the IEEE 7th International Conference on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS'2013), Berlin, Germany, Sept. 12–14, 2013, vol. 2, pp. 828–832.