Representing Earth remote sensing data as time series

Rodriges Zalipynis R.A.

Donetsk National Technical University,

rodriges@csm.donntu.edu.ua

Abstract

Rodriges Zalipynis R.A. "Representing Earth remote sensing data as time series" *To date, all remote sensing data are represented and stored as temporal sequences of separate "snapshots" – rasters or grids. This makes impossible to quickly obtain a time series of a variable values for the full available period for a region of a coordinate grid. Trend research – one of the most important topics in Earth science – becomes extremely complex and time consuming. This paper proposes an alternative data representation and corresponding storage technique. The data are represented as a collection of individual time series, one per each grid cell or raster pixel. New storage layout enables any time series to be always readily accessible. This approach considerably facilitates the application of existing time series techniques to remote sensing, climate reanalysis and similar data as well as provides new research and development opportunities not available before.*

Keywords: time series, Earth remote sensing, climate reanalysis, storage layout.

Introduction

Two closely related scientific products, Earth remote sensing data and climate reanalysis, contain numerous characteristics of vegetation, clouds, aerosols, trace gases as well as dozens of meteorological parameters [1, 2]. Being highly structured and quality assured, these data provide the most accurate and convenient way for exploring any place on the globe.

One of the most important topics in Earth science is analysis of a trend, aiming to determine long term changes. Time series are a powerful scientific tool for trend research due to the existence of many proven methods. For example, R [3] contains over 100 packages for forecasting, modeling, decomposition, filtering and other processing techniques dedicated to time series [4].

To date, all Earth remote sensing data of processing level 3 (according to NASA classification [5]) and reanalysis data are stored as regular latitude-longitude grids in numerous individual files with very different formats and compression algorithms. Their number often reaches several thousand files. This storage organization makes impossible a quick retrieval of a time series for a region.

The main idea of the proposed approach is to operate with time series that are readily available for each grid cell instead of individual grids. An appropriate time series storage format must enable their quick retrieval.

This reminds a transposed 3D matrix which dimensions are latitude, longitude, and time (fig. 1).

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.
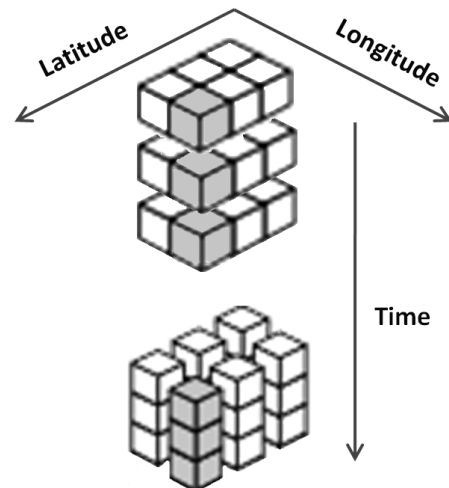

Figure 1. – Representation of Earth remote sensing, climate reanalysis and similar retrospective data by temporal sequence of grids (top) and a collection of time series available for each grid cell (bottom)

The R package was developed which makes time series directly available within R. A user needs only to call a single function providing the variable name and cell coordinates to immediately retrieve the time series for the specified cell and variable.

A time series can be converted to data types required by other R packages. It makes possible to benefit from the rich functionality available from R (over 4000 packages to date).

The case studies given in this paper demonstrate the benefits of operating with time instead of grids during data analysis. Some techniques presented here have been never applied to remote sensing data before. As case studies show, even simple techniques provide new insights and information important for environmental studies.

Climate reanalysis and Earth remote sensing data

During the history of meteorological observations, data were collected by different instruments and have different character like point measurements at ground stations and areal observations by satellites. In addition, they are irregularly distributed in time and space, stored in different formats, contain uncertainties, errors and gaps. It is extremely difficult to use that data to obtain a single picture of an atmospheric state [14, 15].

The purpose of a climate reanalysis is to derive the most precise retrospective picture of atmospheric states using all available data.

Earth remote sensing data are usually available for each radiometer separately. They as well undergo rigorous quality assurance and transformations. Beginning from processing level 3, Earth remote sensing data contain environmental parameters and are in the most convenient representation for global studies.

Both climate reanalysis and Earth remote sensing data of level 3 are represented as time series of global regular latitude-longitude grids to simplify their usage. The time step between successive grids and grid resolution are fixed during the whole period (fig. 2).

For example, AMIP/DOE Reanalysis 2 (R2) spans several decades, from 01 January 1979-present with 6 hour interval and contains over 80 variables. The grid resolution is $2.5° \times 2.5°$ [6].

The only technical difference between climate reanalysis and Earth remote sensing data is that a node of the grid in the former case (fig. 3) and a whole cell of the grid in the latter is associated with a variable value.

This paper will focus on Earth remote sensing level 3 and climate reanalysis data due to the same representation. Both will be referred shortly as remote sensing data further in the paper. By a variable the paper will assume a concrete indicator or parameter within the product (e.g. temperature, aerosol optical thickness, cloud top pressure, etc.).
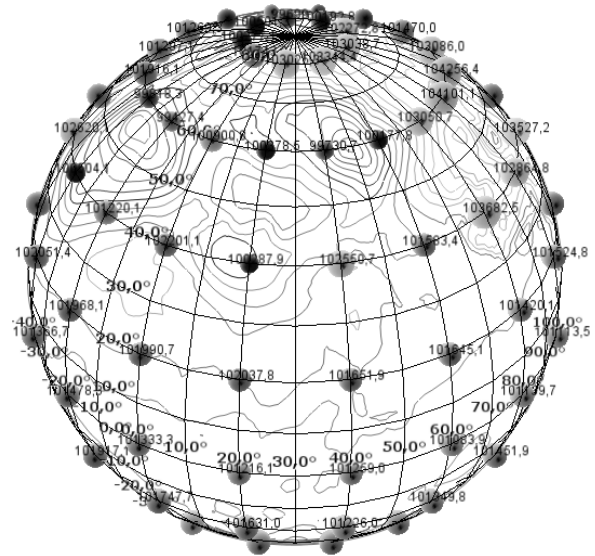

Figure 3. – Mean sea level pressure regular grid with isobars for 01/01/2003 00:00 UTC shown at 20° and 500 GPa intervals respectively. Circle sizes are proportional to pressure values. Built in 3D with Climate Wikience [7] and AMIP/DOE Reanalysis 2

The grids for a variable are stored in a sequence of separate files partitioned by time. Several subsequent grids may be placed in a single file. The number of grids depend on the period for which the data are available and product temporal resolution. File naming conventions are used to locate a file with the grid for a given time moment. Data are stored compressed. File formats, grouping schemes and naming rules depend on the product (table 1).

The period column in table 1 gives the time interval for which grids are placed in a single file and temporal resolution of the product.

For example, each file in R2 stores yearly data for one given variable: pres.sfc.1979.nc, pres.sfc.1980.nc, …, pres.sfc.2011.nc. Where "pres.sfc" denotes "surface pressure", 1979 is year, and ".nc" is file extension for NetCDF format.

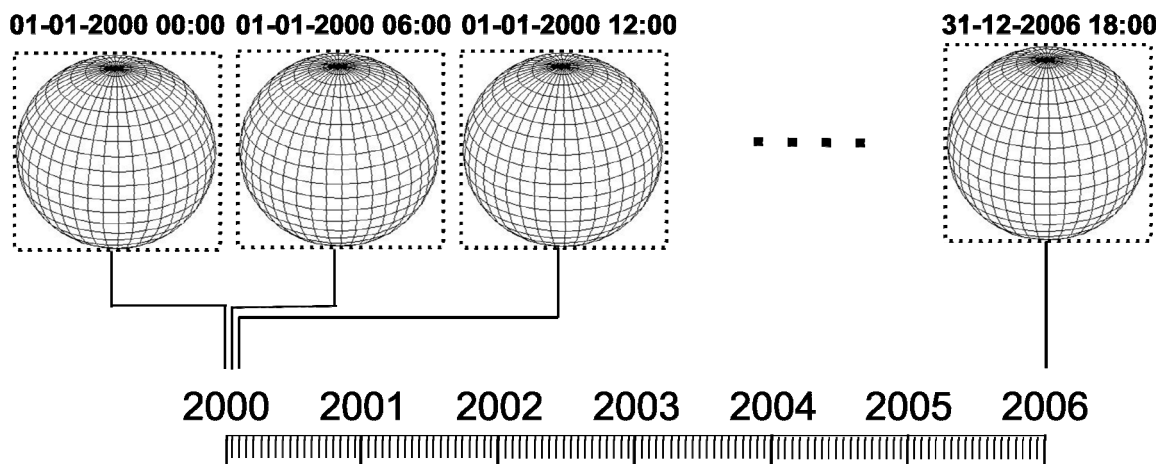Obviously, files have rather complex naming to be perceived by a human.

Figure 2. – Time series of global regular latitude-longitude grids with temporal resolution of 6 hours

Родрігес Заліпиніс Р.А.
Донецький національний технічний університет

Table 1. – Organization of variables in files

Product	Period	Datasets	Format	File name example
AMIP/DOE Reanalysis 2	Year, 6 hours	1	NetCDF	pres.sfc.1979.nc pres.sfc.1980.nc
MODIS L3 Atmosphere	Day, Day	>600	HDF4	MOD08_D3.A2000061.051.2010273210218.hdf MOD08_D3.A2000062.051.2010273161753.hdf
CFSR	Month, 1 hour	1	Grib2	ocnsst.l.gdas.198401.grb2 ocnsst.l.gdas.198402.grb2
MERRA	Day, 1..24 hrs	1..∞	HDF4	MERRA200.prod.assim.tavg1_2d_lnd_Nx.20000718.hdf
Aura satellite, OMI radiometer	Day, Day	14	HDF5	OMI-Aura_L3-OMSO2e_2004m1001_v003-011m0526t144250.he5

Each product has its own coordinate system and can have different spatial resolution along latitude and longitude (for example, $2/3° \times 1/2°$). Thus, grid layout may be different in distinct products

Besides the data, files store metadata. It includes actual data range, constants used for missing values and other important characteristics.

Generating a time series for a grid cell requires downloading all data files to the local hard drive and knowledge of many details of diverse file formats. This distracts a researcher from the primary scientific goals and delays results.

Generating, storing and accessing time series

Time series are generated from existing data in advance. Obtained time series are saved on a persistent storage for further reuse. Appropriate storage layout is used to enable immediate retrieval of a single time series for a given coordinate.

Generation stage takes a significant amount of time but it is required only once per variable. Retrieving all data for a given period, file by file, takes the main part of time.

Figure 4 gives the scheme of obtaining a time series for a single cell. On top is the time series of grids with 6 hour temporal resolution from the 1/Jan/2000 to 31/Dec/2006. Solid thin lines show the correspondence of each grid to the position on the timeline.

A larger grid on the left has the cell colored black. Its latitude and longitude coordinates are equal to *lat* and *lon* correspondingly. A time series for this cell is shown under the timeline.

Time series are built for each pair of latitude and longitude coordinates (each grid cell). The temporal resolution of a time series is equal to the original resolution of the time series of grids (6 hours in current case).
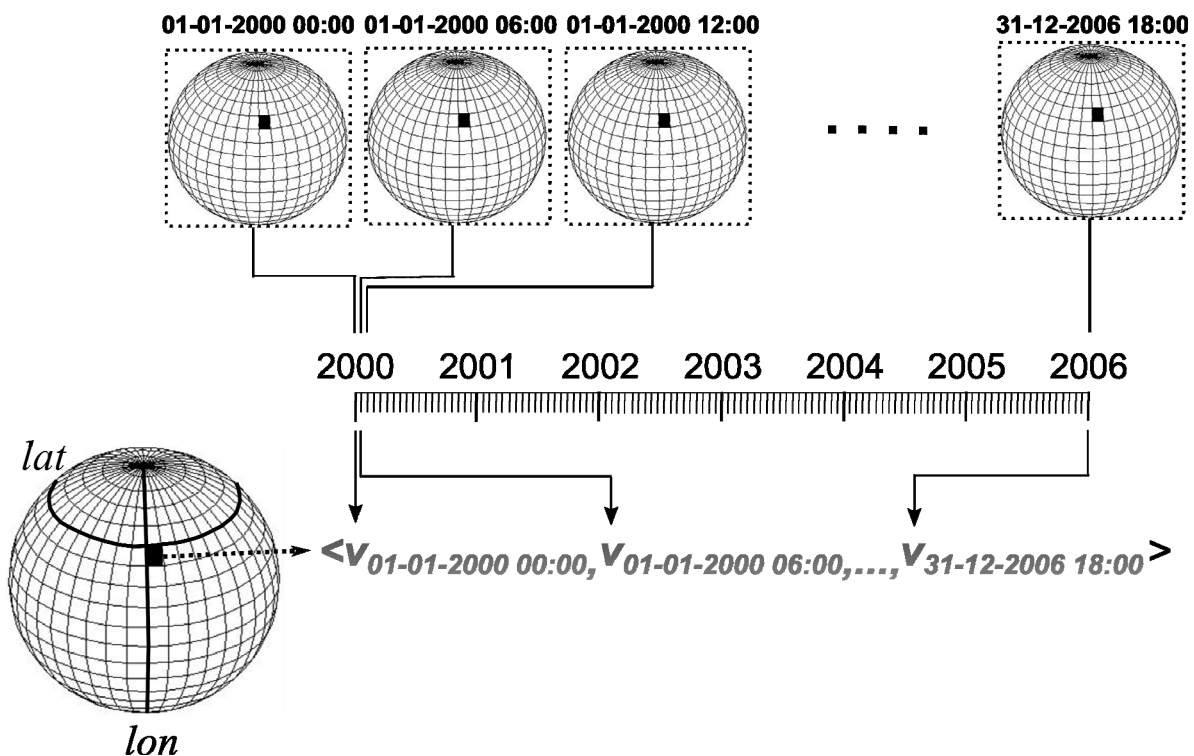

Figure 4. – Obtaining a time series for a single grid cell

Родрігес Заліпиніс Р.А.
Донецький національний технічний університет

The procedure iteratively queries Chronos-Server [8] for the next grid and accumulates time series values taken from the obtained grid in a dedicated in-memory data structure. In case of grid absence for a given date, the special constant analogous to R "NA" is used to indicate the missing value. Thus, all time series will have NA for that date.

Also, not all grid cells of remote sensing data may have values, for example due to clouds. This means that some values in a time series may be also equal to NA.

A special time series storage format was designed. It allows storing time series not only for the global grid and the complete available period but also for a specified region and a time interval.

Time series data for a given region and time interval consist of an index and data file. Index file contains metadata. It is possible to have several pairs of index and data files for a single variable. This is convenient in case of generating time series in several steps by specifying different time intervals and/or regions for each step.

Figure 5 shows the data file structure for time series generated for a region. The boundaries of the region are: southern latitude equals to 30, northern latitude to 65. Analogously, western and eastern longitudes are 40 and 60 correspondingly. First, the time series are sorted on latitude coordinates, then on longitude coordinates and are stored in file sequentially one after another. Time series within the same file have equal length.
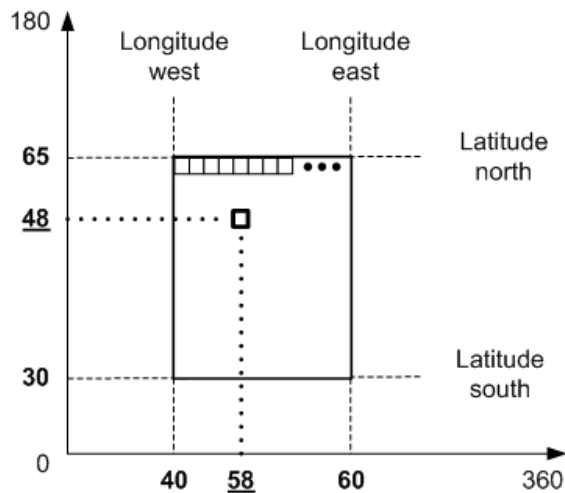

Figure 5. – The structure of time series data file

In order to read a time series for grid cell with coordinates (lat, lon), the position $P(lat, lon)$ of the first value of it is calculated as

$$P(lat, lon) = \quad ((lat - LAT_{south}) / LAT_{step}) \times L \times C + \\ ((lon - LON_{west}) / LON_{step}) \times C$$

where $L = (LON_{east} - LON_{west}) / LON_{step} + 1$ (the number of time series for a single latitude), $C = C_{tb} \times 4$ (data size in bytes of a single time series).

Since all values of a time series are located sequentially, they can be read by a single call to a file system. This feature, coupled with the absence of the need to open thousands of files to obtain an individual time series, yields the significant acceleration and the possibility to reuse time series generated once and always readily available further.

If there are several index and data file pairs for a variable, all of them are read. The retrieved continuous fragments of the time series obtained from the separate files are assembled into a larger single time series. The possibility of fragments having overlapping periods is also considered. In this case, only a single value is used for the same date.

Time series are directly available from R. Special R package was developed for this purpose. To load a time series into R, it is enough to call a single function providing the variable name and grid cell coordinates as arguments.

A grid cell is referenced by the latitude-longitude coordinates of its southwestern corner or lower-left corner when the map is rotated so that south is at the bottom of the map and the west is on its left. For example, reading a time series for ozone total column amount measured by OMI radiometer (Aura satellite) over a grid cell within Donetsk Region (Ukraine) (fig. 6).

```
t <- readTransposeData(w,
"OMI.omso2e.ColumnAmountO3",
     48.25, 37.75)
```

Figure 6. – Reading time series for grid cell with coordinates (48.25, 37.75)

In this case the coordinates are given with 0.25° step since OMI level 3 data global regular grid has 0.25°×0.25° spatial resolution. The function returns R data frame with three columns: unix time stamp (long type), date (string type) and the requested variable value (float) which in this case has temporal resolution equal to one day. There are 2820 values of the variable for the available period (01/Oct/2004 to 20/Jun/2012) (fig. 7).

	time_unix	date	value
1	1.096589e+12	2004-10-01	NA
2	1.096675e+12	2004-10-02	NA
…			
2818	1.339978e+12	2012-06-18	300.6
2819	1.340064e+12	2012-06-19	NA
2820	1.340150e+12	2012-06-20	307.7

Figure 7. – A fragment of a time series (R data frame) for total column ozone over grid cell with coordinates (48.25, 37.75)

The frame can be converted to any appropriate R data type required by other R package so that the functions from that package can be applied to the time series data. Thus, the direct availability of a time series within R equips a researcher with a large variety of analysis opportunities and tools.

Родрігес Заліпиніс Р.А.
Донецький національний технічний університет

Case studies

The paper concentrates on air quality remote sensing data as an example. The goal is to demonstrate the facilitation and acceleration of remote sensing data analysis as well as to give a starting point for suggesting many research options related to remote sensing time series data.

Today, for decision support environmental protection agencies mostly use air quality data that comes from ground stations. Stations deliver point measurements of pollutant concentrations.

Remote sensing data provide gas concentrations with high spatial and temporal resolution. However, these data are not widely used for decision support by the agencies due to high complexity of accessing to remote sensing data as well as prohibitively labor-intensive analysis costs.

The data used for case studies with their characteristics are given in table 2. The first column gives the variable name. Sulfur dioxide and aerosol were retrieved from measurements of Aura (OMI radiometer) and Terra (MODIS radiometer) satellites. Two different ozone data belong to both satellites to demonstrate intercomparison possibilities. Meteorological parameters, wind speed and direction were taken from the MERRA reanalysis.

All data are available at regular latitude-longitude grids. To give a feeling of the spatial resolution of the grids, the sizes of their cells are given in kilometers for ~ 49±4 northern latitude.

MODIS products of level 2 have spatial resolution of 5 km for ozone and 10 km for aerosol products. Level 3 MODIS products contain statistics for pixels that fall into 1.0°×1.0° cells of regular latitude-longitude grid. This paper takes maximal value among all pixels that fall into 1.0°×1.0° cell. This is due to assumption that if the cell covers a city or an industrial region, mainly anthropogenic factors may cause the maximal level of the aerosol reported for that cell.

Sulfur dioxide column measurements are taken for total columns within planetary boundary layer. Sulfur dioxide and ozone concentrations are given in Dobson Units (DU), 1 DU = 2.69×10^{16} molecules/cm^2. Aerosol optical thickness is dimensionless which values fall into -0.05..+5 interval.

Wind speed and direction are for the height of 50 meters above surface. Hourly data were aggregated to daily interval by selecting and averaging values from 9.00 to 12.00 of local Ukrainian time (UTC+02).

Time series were generated for each grid cell (for satellite data) and node (for reanalysis). All data taken are approximately until June 2012.

Table 2. – Data used for case studies

Variable	Radiometer or reanalysis	Grid resolution	Cell size for Ukraine, km	Temporal resolution	Starting date	Units
Total column ozone (O$_3$)	OMI	0.25°×0.25°	27.5 × 18	1 day	01.10.2004	DU
Sulfur dioxide (SO$_2$)						DU
Total column ozone (O$_3$)	MODIS	1.0°×1.0°	110 × 72	1 day	02.03.2000	DU
Aerosol optical thickness						-
Wind speed and direction	MERRA	1/2°×2/3°	55 × 48	1 hour	01.01.1979	m/s

```
1    library(RWikience)
2    w <- WikienceConnect()
3    t <- readTransposeData(w, "Modis L3 Atmosphere.
              Land.Optical_Depth_Land_And_Ocean.Maximum", 41, 12)
4    t_avg <- aggregate(t["value"], format(t["date"], "%Y-%m"),
                                      mean, na.rm = TRUE)
5    t_avg$date <- seq(min(t$date), max(t$date), length = nrow(t_avg))
6    plot(t_avg$date, t_avg$value, type="l")

7    ids <- which (t$value <= 2)
8    plot(as.factor(format(t$date[ids], "%m")), t$value[ids])

9    library(zoo)
10   t_avg.trim <- na.trim(t_avg)
11   t_avg.trim$value <- na.approx(t_avg$value)
12   t_avg.ts <- ts(t_avg.trim[,2], start=c(2000,3), freq=12)
13   plot(stl(t_avg.ts, s.window="periodic", robust=TRUE))

14   library(openair)
15   calendarPlot(t, "value", 2004)
```

Figure 8. – R code for aerosol optical thickness trend research over Rome, Italy

Родрігес Заліпиніс Р.А.
Донецький національний технічний університет

All code that is provided in the paper can be executed in RStudio [9] or RevolutionR [10] providing that Climate Wikience [7] is launched on the same computer and the required R packages are installed. The package 'RWikience' will be freely available soon. However, in release version the signatures of its functions might change.

Monthly aerosol optical thickness over Rome (Italy) with MODIS grid cell coordinates (41, 12) is presented on fig. 9. Monthly means used to avoid overplotting with too many daily values.

It was experimentally found that Rome cell has one of the lowest cloud cover rates among other cells in Europe (thus, containing much less missing values in its time series than other cells). In addition, it has very interesting trends.

Taken MODIS grid cell (referred simply as Rome, Italy) also covers a part of nearby Tyrrhenian Sea. Thus, the diagrams entitled as being for Rome may also reflect other natural patterns.

Some winter months still contain only missing values mainly due to clouds. This results in gaps in time series.

For drawing plot on fig. 4 the 'RWikience' package is loaded (fig. 8, line 1) and a connection to the running instance of Climate Wikience is established (line 2).

To obtain a time series, a hierarchical name of a required variable and grid cell coordinates must be supplied (line 3). All variables from remote sensing and reanalysis products are categorized in Climate Wikience and organized into a hierarchical namespace. This simplifies navigation in available variables and location of the required one.

Monthly means are loaded into R data frame at line 4. A trick (line 5) is required before using R 'plot' function (line 6) to visualize resulting data from 'aggregate' function.
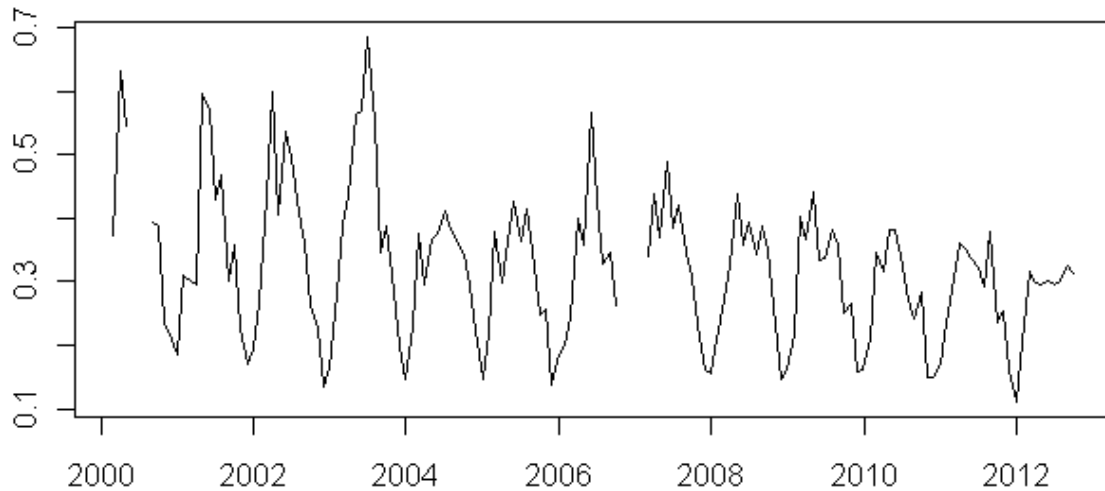

Figure 9. – Monthly aerosol optical thickness over Rome, Italy

An apparent feature of many remote sensing data variables is seasonality. It is possible to create a box plot for variable values in a time series for each month (considering daily values for months from all available years) to find out the typical seasonal character of a variable. Values higher than 2.0 were excluded before building the plot (fig. 10).

Although the fact of a seasonal cycle for a variable is well known, it is different in distinct places on Earth. Thus, having a time series for a cell or an area with precise coordinates, a researcher can have a better feeling of the seasonal pattern inherent exactly to the region under investigation.

As it can be seen from the plot (fig. 10), maximum aerosol concentrations over Rome are usually reached at summer months and minimum at winter months.

The plot on figure 10 was built simply with two additional lines of code (7 and 8).

A more sophisticated approach is to apply a seasonal-trend decomposition based on loess [11]. However, due to cloudy days during winters of 2000-2001 and 2007-2008, the time series does not have some values. Missing values can be linearly interpolated from neighborhood data (line 11).

As it was noted earlier, time series data can be easily converted to types required by other packages to use their functions. The time series was converted into 'ts' type (line 12). A robust "stl" procedure with periodic time window was used (line 13) to construct plot on figure 11.

The monthly time series of aerosol optical thickness over Rome with linearly interpolated missing values is shown at the very top of figure 11. The seasonal component with yearly periodicity is shown next.

The most interesting component - the extracted trend - reveals obvious decline of aerosol concentration over Rome in 1.6 times (from about 0.40 to 0.25) during the last decade.

The remainder indicates outliers. Each event might be studied in detail and explained separately.

This insightful study can be easily applied to any city. Since query for a time series happens in a split of a second, the result is readily obtained.

Родрігес Заліпиніс Р.А.
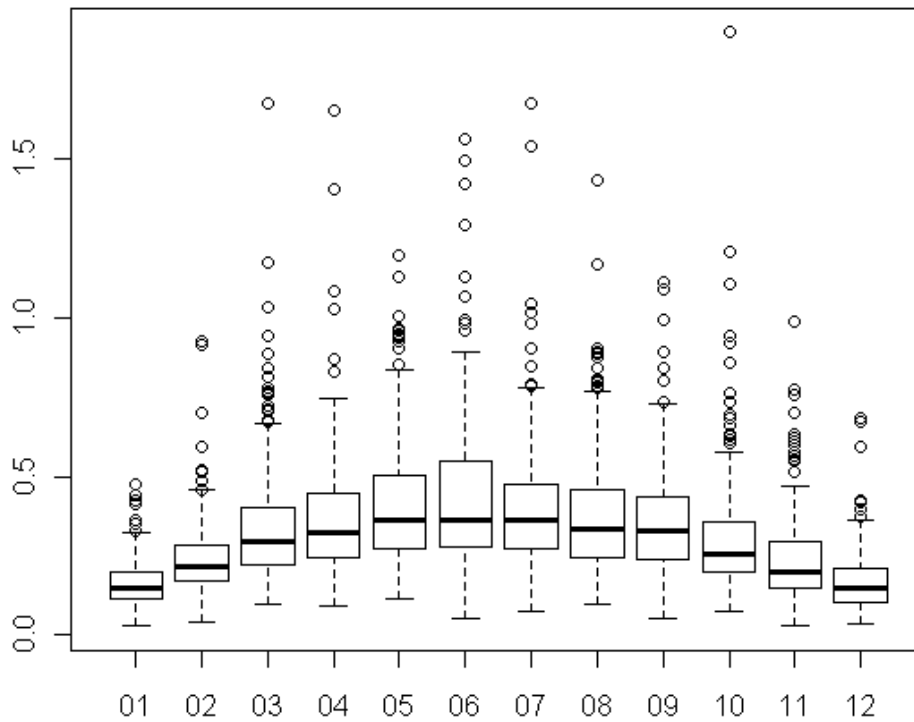Донецький національний технічний університет

Figure 10. – Box plot for each month, aerosol optical thickness, Rome, Italy
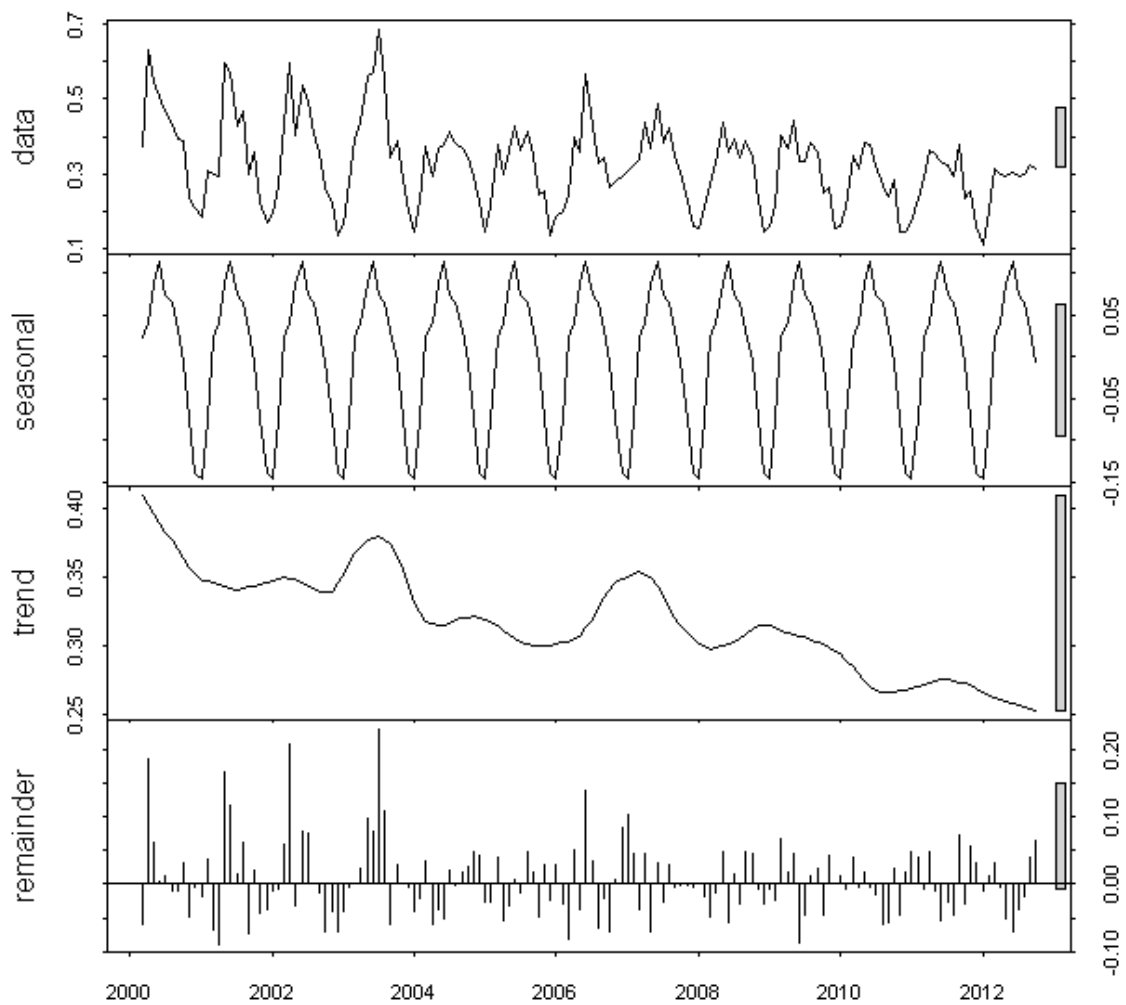

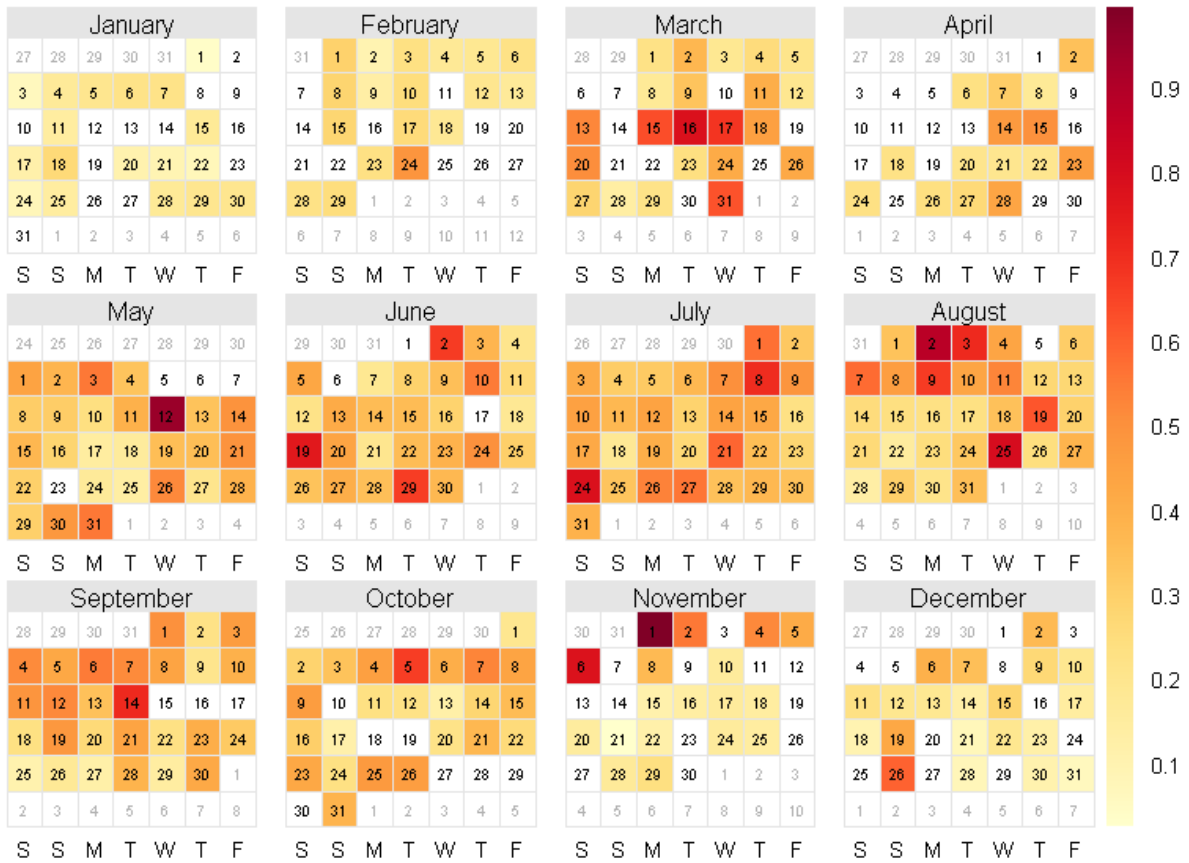Figure 11. – Seasonal-trend decomposition of aerosol optical thickness over Rome, Italy

Figure 12. – Daily aerosol optical thickness over Rome (Italy) during 2004 (Terra MODIS)

Some very effective functions are implemented in 'openair' package [12, 13]. They facilitate the detailed study of a variable dynamics.

For example, it is possible to plot aerosol optical thickness for each day in a calendar-like style (fig. 12). It was created by a single line 15 from figure 8. Days which are not shaded contain missing values.
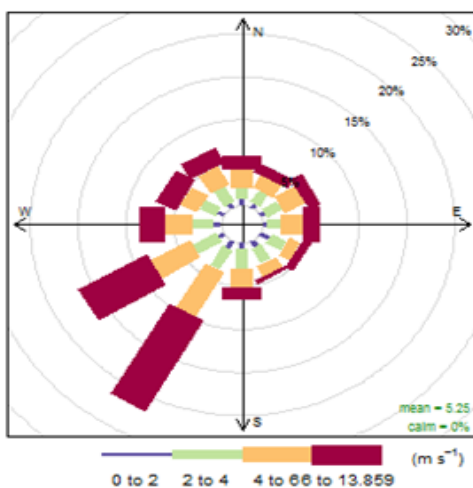

Figure 13. – Wind rose for a single MERRA node $(48^2/_3, 24)$ using daily data for 2000-2012

Recall that the figure 12 shows the variable value for a single grid cell (not averaged for several cells) enabling exact study of the desired region.

Many insights can be gained using hourly or daily time series of wind speed and direction. For example, a wind rose for any place on the globe can be built in a split of a second (fig. 13).

Even more can be achieved by fusing data from different products.

Time series of meteorological parameters and air pollutant for the same grid cell make possible to determine the dependencies of the latter on the former. This enables to detect the most liable to pollution areas by nearby anthropogenic objects as well as the pollutant spread distance from them.

Package 'openair' contains 'polarPlot' function to construct polar diagrams of the dependence of a pollutant from wind speed and direction.

The polar plot (fig. 14a) shows the SO_2 content in the total column of planetary boundary layer over the cell with coordinates (49.25, 24.25) depending on the wind speed and direction at the node with coordinates $(48^2/_3, 24)$ of the MERRA grid.

The plot was created using all data from the available period. Values of SO_2 content are averaged for the same wind speed and direction.
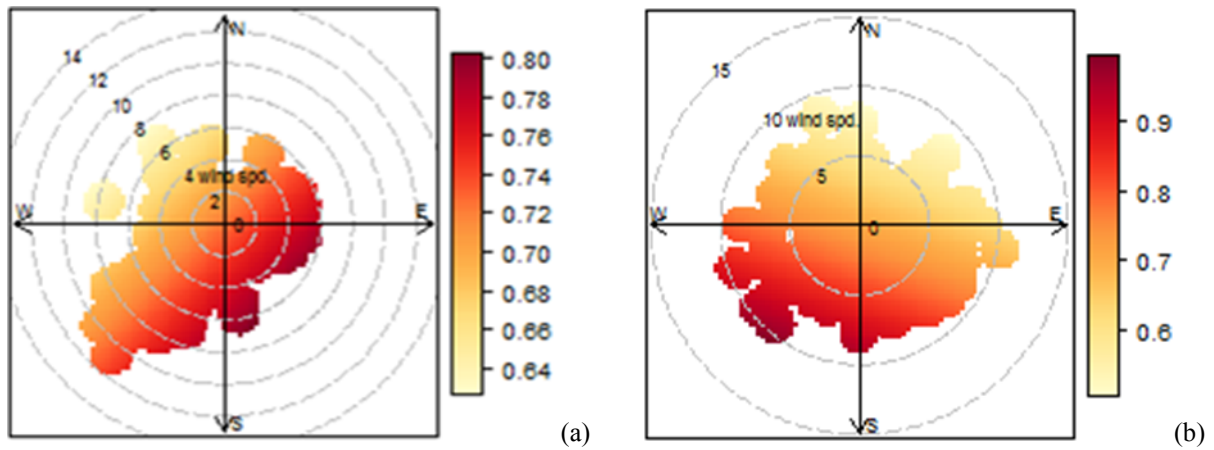
Родрігес Заліпиніс Р.А.
Донецький національний технічний університет

Figure 14. – Influence of (a) Burshtynska and (b) Slovyanska thermal power stations on SO_2 air pollution levels at nearby cells depending on wind speed and direction

The center of the pollution polar plot represents a wind speed of zero, which increases radially outward. There is evidence that maximal SO_2 values are observed during southeastern wind with speed 2-4 m/s while minimal SO_2 values during northwestern wind.

The cell with coordinates (49.25, 24.25) is on the northwest from Burstynska thermal power station - one of the major SO_2 emitters in Ukraine. Its coordinates approximately are (49.20, 24.66).

The same plot for MERRA node with coordinates (49, 38) and SO_2 grid cell with coordinates (49, 38) is shown on figure 14b.

The cell (49, 38) is on the northeast from Slovyanska thermal power station with coordinates (48.87, 37.76) approximately.

Although this visualization technique was developed before this publication, in this paper it is applied to climate reanalysis and remote sensing data for the first time.
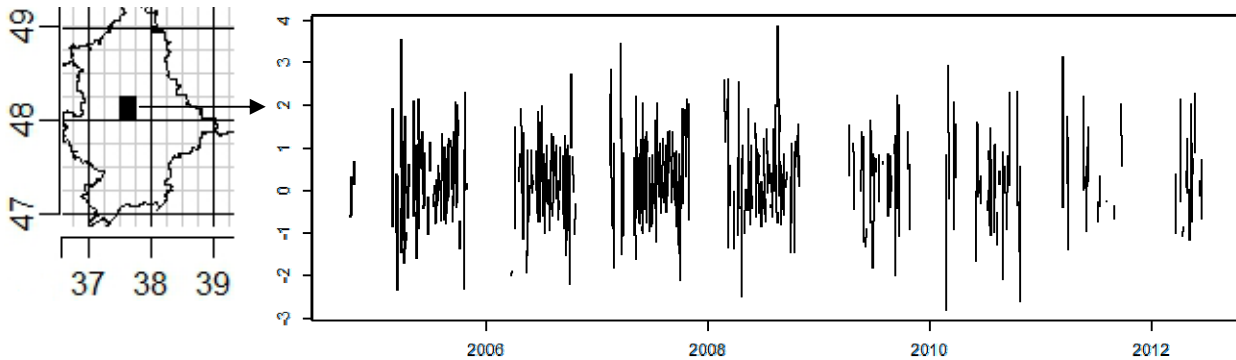

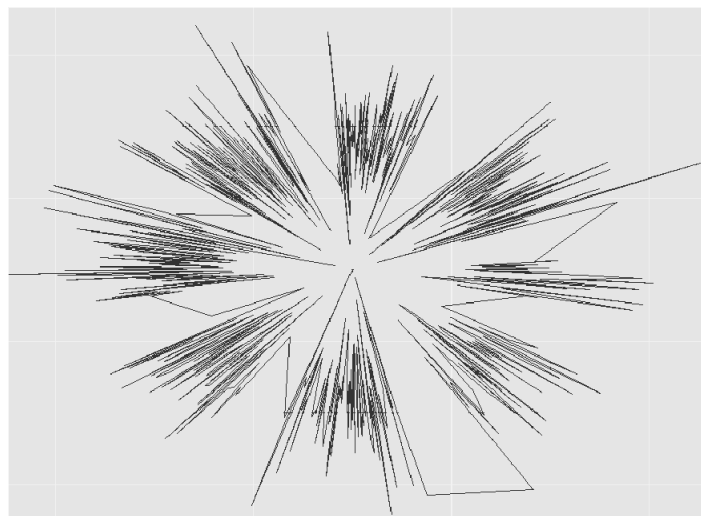Figure 15. – Time series of SO_2 concentration for Donetsk, Ukraine (48, 37.5) during 2004-2012

Figure 16. – Representing time series from fig. 15 in radial form (radius – value, angle – time)

Родрігес Заліпиніс Р.А.
Донецький національний технічний університет

Other packages as well offer a large variety of exploratory data analysis techniques [14, 15].

The techniques used so far allow to investigate either spatial or temporal features of data but not both of them simultaneously.

It is also possible to plot a time series in radial, so called "star" form. Figure 15 shows a single time series for a cell inside Donetsk region while figure 16 represents it in radial form.

To show the dynamics of a variable for a whole region, at the center of each cell within that region stars of their time series are plotted.

This provides a good feeling of data character, allows comparing different subareas, notice data peculiarities and deficiencies at a glance.

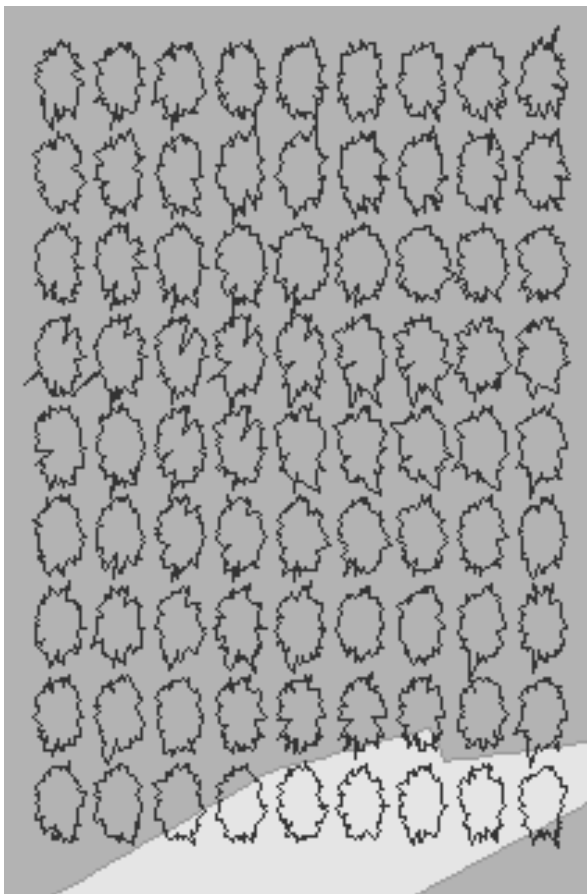This is only possible if we have a time series for each grid cell as proposed in this paper.

Figure 17. – Time series of SO_2 concentrations during 2004-2012 for each grid cell (18×27.5 km) for Donetsk region (fig. 15), Ukraine

Empowered by the RWikience package, it is also quite easy to carry out quick intercomparison of the values of the same variable from different remote sensing products. It is very important as it gives the idea of how well the algorithm or radiometer performs in different situations. It may also help to choose the product that suits best for a particular purpose.

Since ozone measurements from MODIS and OMI radiometer have different spatial resolution, they must be compared within a common area.

For example, the region with most northwestern (48, 36) and most southeastern (47, 37) corner can be taken since it contains a single cell for MODIS grid ($1.0° \times 1.0°$) that fully covers 16 cells of OMI grid ($0.25° \times 0.25°$) (fig. 15).

First, all time series for grid cells within the selected region must be averaged to obtain a single time series. For MODIS there is only 1 cell (no averaging is required) while for OMI there are 16 cells (from which a single time is obtained).

While averaging several time series into a single one, a special consideration must be taken on dates for which one or several series contain missing values. In this case, an average for that date is calculated based only on valid values.

Figure 18 gives a scatter plot for two resulting time series. The MODIS values are depicted along abscissa while the OMI values are along ordinate. Only dates for which both time series do not have missing values are considered.
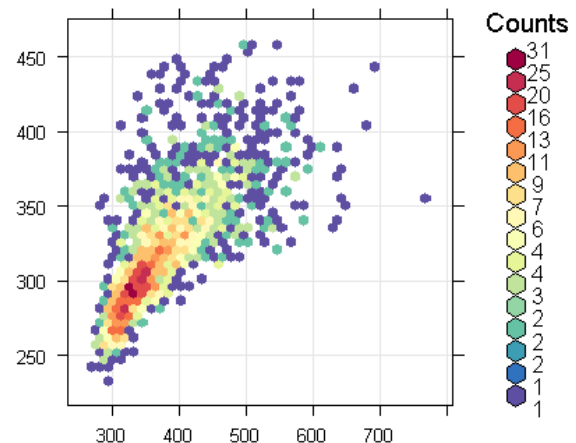

Figure 18. – MODIS versus OMI total column ozone for a subarea of Donetsk region, Ukraine

There is a noticeable agreement between different satellites in spite of MODIS product used reports maximum ozone value for a cell while OMI gives an average.

Related work

To date, the only system capable to generate time series for Earth remote sensing data is Giovanni [16]. Its main disadvantages are:

• it generates a time series upon user request which requires a considerable amount of time;

• only a single resultant time series for a selected region can be ordered making impossible to carry out precise investigation of territories which areas are comparable to an area of a single cell;

• the output is presented as a file of a given format (text, binary, etc.) for manual download via

Родрігес Заліпиніс Р.А.
Донецький національний технічний університет

a web interface. This makes a researcher to distract on managing many files (in case of studying several territories) on a destination personal computer and increases data to result time;

• only a subset of variables from only NASA datasets are provided and most of which are available at reduced time/space resolution.

Representation of Earth remote sensing data as time series partially resembles column-wise storage layout used in database systems. While at the first sight it is analogous to row-wise approach, significant performance benefits could be obtained for certain workload patterns by utilizing dedicated column-oriented processing techniques [17].

Conclusions

The paper proposes an alternative view of the retrospective georefernced data. They are represented as a collection of time series instead of temporal sequences of grids.

Time series have the same temporal resolution as original sequence of grids and are available for each grid cell. This preserves the finest possible spatial and temporal analysis precision.

Corresponding storage technique and R package have been developed to enable immediate retrieval of a time series for any grid cell and its direct availability from R analysis environment. The R package enables to benefit from over 4000 packages available from R environment to date.

The proposed alternative look at data representation and their organization greatly facilitate the application of existing time series analysis techniques to Earth remote sensing data. It is far not a trivial task without having time series for each grid cell and quick access to them.

The proposed representation may also alter the perception of the data by a researcher and may stimulate the emergence of Earth remote sensing data use cases not considered before.

The author believes that this work will suggest further exploration of the potential and opportunities that give the representation of Earth remote sensing and similar data as time series.

Acknowledgements

References

1. MODIS data products // El. resource. URL: http://modis.gsfc.nasa.gov/data/dataprod/dataproducts.php
2. MERRA data holdings // El. resource. URL http://disc.sci.gsfc.nasa.gov/daac-bin/DataHoldings.pl?LOOKUPID_List=MAT1NXLND
3. R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/
4. CRAN Task View: Time Series Analysis // El. resource. URL: http://cran.r-project.org/web/views/TimeSeries.html
5. Processing Levels | EOSDIS - Earth Data Website // El. resource. URL: http://earthdata.nasa.gov/data/references/processing-levels
6. AMIP/DOE Reanalysis 2 // El. resource. URL: http://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis2.html
7. Climate Wikience // El. resource. URL: wikience.donntu.edu.ua
8. Rodriges Zalipynis R.A. ChronosServer: real-time access to "native" multi-terabyte retrospective data warehouse by thousands of concurrent clients. Informatics, cybernetics and computer engineering, pp. 151–161. Vol. 14 (188) – Donetsk: DonNTU – 2011. – 391 p.
9. RStudio // El. resource. URL: http://www.rstudio.com/
10. Revolution analytics // El. resource. URL: http://www.revolutionanalytics.com/downloads/
11. Cleveland R, Cleveland W, McRee JE. Seasonal-trend decomposition procedure based on LOESS. J. Offic. Stat., 1990; 6:3–73.
12. Carslaw, D.C. and K. Ropkins openair --- an R package for air quality data analysis. Environmental Modelling & Software. Volume 27-28, 52-61, 2012.
13. David Carslaw and Karl Ropkins openair: Open-source tools for the analysis of air pollution data. R package version 0.5-21, 2012.
14. Visualize complex data with subplots // El. resource. URL: http://blog.revolutionanalytics.com/2012/09/visualize-complex-data-with-subplots.html
15. Garrett Grolemund, Hadley Wickham Package ggsubplot // El. resource. URL: http://cran.r-project.org/web/packages/ggsubplot/index.html
16. Giovanni // El. resource. URL: http://disc.sci.gsfc.nasa.gov/giovanni/overview/index.html
17. D. J. Abadi, P. A. Boncz, and S. Harizopoulos. Column-Oriented Database Systems. VLDB, 2(2), 2009.

Родрігес Заліпиніс Р.А.
Донецький національний технічний університет