Exploration and visualization tool for analysing big data from a large measurements

Galib Hasan
galib.hasan@helsinki.fi

ACM Reference Format:
Galib Hasan. 2019. Exploration and visualization tool for analysing big data from a large measurements. In *Proceedings of DSNS'19: Data Science for Natural Sciences Seminar (DSNS'19).* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

Abstract

Keywords: SMEAR, MySQL, APIs, CSV, AJAX technology, Virtual globe, KML files

1 INTRODUCTION

Data exploration and visualization are the indispensable part of the scientific community. As massive source of data is collected and stored randomly, their effective use becomes more challenging. It is very important to analyze big data from a laboratory measurement, simulations, and research campaigns or from another source of measurement to reach in conclusions and proving the hypothesis. Many standard tools exist for data management and analysis from a large set of measurement to the scientific community. This report covers two specific (Smart-SMEAR and GOOGLE EARTH) data exploration and visualization tools for the measurement of atmospheric data.

2 SMART-SMEAR

The atmosphere is probably the most vulnerable major component of the earth system. Any irregularities of a single element of it might have a major impact on the living species. For, example atmospheric aerosol and trace gases have a potential influence on the climate change and hydrological cycle in our environment. Therefore, it is very important to observe the exact amount and types of trace gases present in the environment [1]. Smart SMEAR **(Station for measuring the forest Ecosystem-Atmosphere Relationships)** is an interface to visualize the data measured at the different SMEAR stations, Finland. The smart system is designed for data mining (aerosol data, meteorological parameters, and air mass back trajectories) and for the visualization of the measured data from the environment[1].

The SMEAR has FOUR stations and all those stations has scientific program to investigate aerosol and trace gas concentration[1].

DSNS'19, Spring 2019, Helsinki, Finland

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM.
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1. SMEAR I, Värriö forest, since 1991 to measure the pollution fall-out to Eastern Lapland from the mining industry in the Kola-peninsula in Russia.
2. SMEAR II, Hyytiälä forest, since 1994.
3. SMEAR III, Kumpula, Helsinki, since 2004 (Operated by Department of Physical Science, Chemistry, and Forest Ecology and Finnish meteorological Institute(FMI).
4. SMEAR IV Puijo in Kuopio, since 2009 (Aerosol Particles and clouds)

2.1 Database

The smart database of the software contains about 170 variables and all those data are saved in the MySQL database automatically and updated every two hours for meteorological, gas phase and some other aerosol measurements. The database is manually verified every half year with verified data is [1].

2.2 Application Programming Interfaces (APIs)

SMEAR has an API that can be used to access the SMEAR database. It is a relational database where time series data is stored in multiple tables. Every row of the table contains the timestamp of the measuring moment and the values of the measured variables at the given timestamp. The response data is found as CSV or TXT format. There is also metadata API which offers five endpoints (Variables, Events, Tags, Stations, and Tables) to access metadata, which describe variable stored in SMEAR database. Finally, trajectory API uses for accessing trajectory data for SMEAR II station Hyytiälä. Smart-SMEAR is php-based dynamic web page. The graphs are created with JPGraph (PHP Graph Creating Library). The maps are made using Google Maps API).

2.3 Smart-Search

Smart search is an interface for searching and downloading data from SMEAR database. User can see and download data from the smart search using a time interval of interest. The data are available with daily, weekly, monthly or hourly even with one-minute frequency. Smart-Search uses AJAX technology (Asynchronous JavaScript and XML) to communicate with the database. User can select data for their interests (gases, meteorological, radiation and aerosol data). The SMART Search connects to the database, checks the quality of the data and returns a table with downloadable variable and percentage. The graph comes with colors, where green meaning good quality data, gray medium-quality, and red low quality. The downloadable data is available in CSV and tab-separated file format [1].

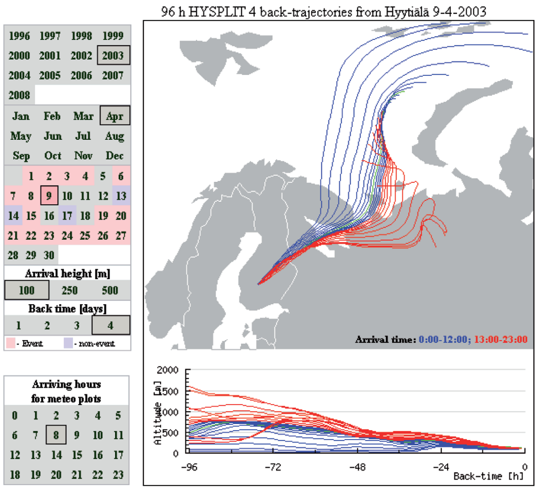

Figure 1: Selection panel and the back-trajectories map [1].

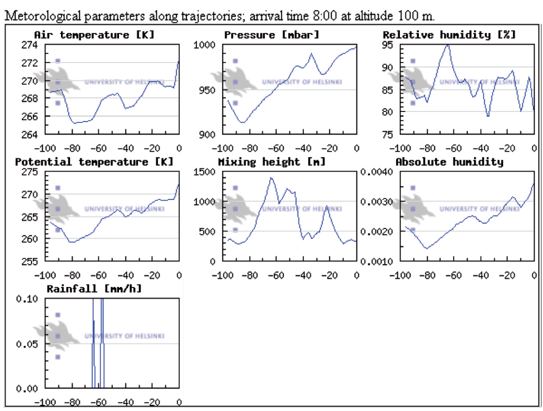

Figure 2: Meteorological parameters along trajectories [1].

In Smart-SMEAR system, the data visualizer has two parts, a selection panel, and graphs. In the selection panel, days with new-particle formation are presented on a red-colored background and non-events days on the blue-colored background. The graphs are generated based on the selection criteria and it changes automatically every time if any selection criteria are changed. **Figure 1** represents a back-trajectory for the one day. Here, each trajectory represents one hour. Trajectories with an arriving time from 00:00 to 12:00 are blue, those with an arriving time from 13:00 to 23:00 are red and one that is selected from the hour panel is green. **Figure 2** represents the meteorological parameters along with trajectories [1].

3 GOOGLE EARTH AS A VIRTUAL GLOBE TOOL FOR EARTH SCIENCE

Global environmental change has become one of the most talked topic issues among the scientific community. Modern technological instruments are needed to provide the measurement, monitoring, modeling, analysis and assessment tools for solving problems at global scales. Google earth is one of the best revolution in the earth system research to collect, manage and display global environmental data. It is a free program from Google that allows everyone to surf over a virtual globe and visualize high-resolution graphics and satellite image. It also allows viewing the landscape of many geographic areas from local to global scales in three-dimensional (3D) global representation. Efficient algorithms can handle huge volumes of mult-isource data for global scale information extraction and change monitoring, and spatial-temporal techniques can integrate observation data into process-based models. It is very exciting that anyone can clearly see his specific locations, areas even own home using google earth. Besides Google Earth, there are also a significant number of technologies developed in the field of Earth science such as NASA's World Wind, and ArcGIS Explorer, providing huge volumes of freely available images and 3D views of the Earth [2][3].

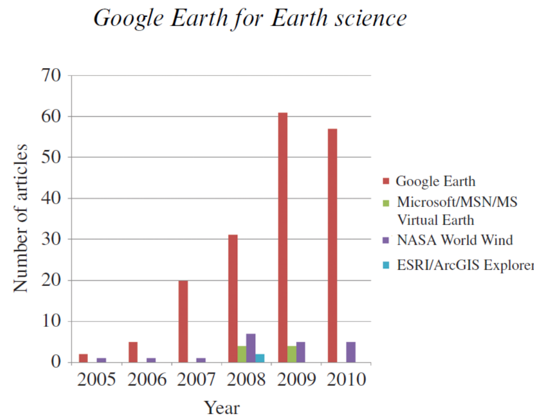

Figure 3: The number of published articles from 2005 to 2010 with reference to four virtual globes based on ISI Web of Science. The total number of published papers on 'Google Earth' is nearly six times of the total number of papers of the remaining three.

3.1 Virtual globe

A virtual globe is a 3D representation of the Earth with the ability to (1) explore in a virtual environment, (2) add users' own data and share them with others and (3) represent natural and fabricated features on the surface of the Earth. Buckminster Fuller, an American architect first proposed the idea of a virtual globe that computers could be used to model a virtual Earth and the concept of Gyroscope. Later his idea was conceptualized as 'Digital Earth' by Gore (1992) [2]. He predicted about digital future where people could interact with a computer-generated 3D virtual globe and access major amounts of scientific information in order to understand the Earth and its human activities. Virtual globes have opened the world of satellite images to the general public and allowed entertainment, education, and exploration of new findings. There are many virtual globes available nowadays, for example, Google Earth, NASA World Wind, and Skyline Globe. There was a literature survey on January 2011 for the period of 1 January 2005 to 31 December 2010 and showed Google Earth was used nearly six times higher than others programe in terms of publishing scientific journals **Figure 3**. There are three versions of Google Earth, namely Google Earth Free, Google Earth Pro, and Google Earth Enterprise. Here, Google Earth Pro is commercial purposes and can use by subscription-only. Google Earth Free is free to use and provide different measurement data to the customers with additional features in GIS and remote sensing (RS) data import. On the other hand, Google Earth Enterprise uses for the organizations that have satellite imagery or large quantities of geospatial data that need to be used in their own demand. Besides the innovative techniques of sphere visualization, there is another popular technique used in Google Earth is KML. Google Earth is the first program, which can view and graphically edit the KML files [2][3].

3.2 Applications of Google Earth

There are several applications of Google Earth based on its demand to a wide variety of scientific discipline. Its applications are focused on the large-scale phenomenon in the atmosphere, cryosphere, ecosystems, energy, geology, natural disaster, and many other fields. The application of Google Earth can be over-viewed into a few general categories, for example, visualization, data collection, validation, data integration, modelling, data exploration, and decision support [4],[5]. Visualization is the main function of Google Earth. This function likely to be used for the numerous types of task depending on which data are to be visualized. Another important function is data collection which explicitly requires field trips equipped with GPS units or other positioning techniques. This process is limited by the lack of GPS units or lack of positioning signals in some locations. Images with global coverage favor the data collection process. In most cases, those images provide useful information, but for some cases, image processing is required. Unlike data collection, Google Earth images are also widely used for the validation process. Data integration is another application of Google Earth, which actually refer to the integration of heterogeneous georeferenced 1D/2D/3D/4D data in local computers or in the fly from distributed sources. In case of Google Earth, these data are usually used in KML format. Another significant use of Google Earth is modelling, where two types of modelling need to be considered, one is constructing static 3D models, another is modelling of dynamic phenomenon. Besides, Google Earths are widely used for the exploration of data and decision support.

3.3 Limitations of Google Earth

Although Google earth has become much more popular among the environmental scientist but still there exist several limitations. Inconsistent quality of images, insufficient capability for quantitative measurement and lack of analytical functionalities are the main limitations of Google Earth. The inconsistency of remotely sensed image quality in Google Earth impedes consistent analysis. Sometimes the positions of images are not accurate on Google Earth. High resolution of Google Earth covers an area about 20-30 percent of the world, meaning that most areas in the world suffer from poor coverage of high-resolution imagery. Precise quantitative measurement and analysis are lost on Google Earth. Basically, image from Google Earth uses for visualization purposes and it is not perfect for analysis purposes because various multi-spectral combinations of remotely sensed images are not supported, no temporal flexibility of image acquisition, large inconsistent radiometric distortion. It is also not possible to measure the topographic parameters quantitatively using Google Earth. Analytical parameters are limited in Google Earth, it supports only a small portion of what a full GIS software package does in applications. The representation and tessellation of the Earth's surface in Google Earth contain inadequate operations. Google Earth uses geographic coordinates (latitude or longitude) as the internal coordinate system based on WGS84 datum. WGS84 model is a sphere-based geographic coordinate system which is called 'WGS 1984 Major Auxilary Sphere'. Using this model the earth shape will likely introduce a horizontal and vertical location error of images .

4 CONCLUSION

Visualization of complicated data set from the set of large measurement has become easy to handle using modern visualization tools. Still, the visualization algorithm is developing and tends to get it more and more user-friendly. Smart-SMEAR and Google Earth are considered to be one of the useful and simple visualization tools which provide easy hands-on and insights to a complicated larger data set. They actually have shown to be effective in giving the basis for different case studies and different data sets. We hope these tools will be further developed by including more data into it and by expanding its current capacity to a state of modern technologies.

5 REFERENCE

[1] H. Junninen et al., "Smart-SMEAR: On-line data exploration and visualization tool for SMEAR stations," Boreal Environ. Res., vol. 14, no. 4, pp. 447–457, 2009.

[2] L. Yu and P. Gong, "Google Earth as a virtual globe tool for Earth science applications at the global scale: progress and perspectives," Int. J. Remote Sens., vol. 33, no. 12, pp. 3966–3986, 2012.

[3] A. Chen, G. G. Leptoukh, S. J. Kempler, and L. Di, "Visualization of Earth Science Data Using Google Earth," Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci., vol. 34, no. 1, 2009.

[4] M. F. Goodchild, "The use cases of digital earth," Int. J. Digit. Earth, vol. 1, no. 1, pp. 31–42, 2008.

[5] A. S. Stensgaard et al., "Virtual globes and geospatial health: The potential of new tools in the management and control of vector-borne diseases," Geospat. Health, vol. 3, no. 2, pp. 127–141, 2009.