

INTEGRATING GRID AND TRAJECTORY DATA VIA A WEB SERVICE: CASE STUDY OF ICEBERG MOVEMENT

Ha Duc NGUYEN ♣, Ulanbek D. TURDUKULOV ♥, Connie A. BLOK ♥

♣ International Centre for Advanced Research on Global Change (ICARGC), Vietnam National University Hanoi, Hanoi, VietNam. Email: nh14vn@gmail.com

♥ Department of Geo-Information Processing, Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, P.O.Box 6, 7500 AA, Enschede, the Netherlands

KEY WORDS: Spatio-temporal data integration, trajectory, OPeNDAP, IDV

ABSTRACT: The understanding about behaviors of moving objects is required in many areas of science and technology. However, moving objects are influenced by many factors. These factors hide significant knowledge for understanding and predicting object movement behaviors. Therefore, to study behaviors of moving objects, the relations between moving objects and influencing factors need to be identified. The development of tracking devices, satellite technology, grid exchange mechanism and web services bring unprecedented opportunity for researchers and scientists to discover these relations. Nevertheless, the difference in data formats (trajectories are recorded in two/three dimensional vector while environmental factors are often in multidimensional grid), spatio-temporal resolution and the large quantity of trajectory and grid data become obstacles for data integration.

To ease the integration process and facilitate users from different research fields and distinct data integration needs, we aim to build up a web service that provides the capability to integrate trajectory data with grid data from remote servers. Concerning the advantageous of data structure that includes time as a dimension, data availability and service supporting, we decide to concentrate on multidimensional grid data of the Earth Science community. The result of this research is a web service with an HTML user interface that allows users just upload their own trajectory data file, indicate grid data source links and receive the integrated data set. The integration mechanism is built up by using IDV and Jython script. It has capability to access grid data from OPeNDAP servers through OPeNDAP protocol then integrate grid and trajectory data. Some functions for interpolation and evaluating uncertainty of interpolated value are also supported. The server uses Python script as a middleware to get and respond requests from client side. Iceberg trajectory data and grid data of sea surface temperature, wind direction are used as a case study.

1. Introduction

Movement is a main component of many activities. In many areas of science and technology such as meteorology, biology, sociology, transportation engineering, it is important to be aware of movement behaviors (Dodge et al., 2008). Many researches have been studying behaviors of moving objects such as animal movements, transportation movements. With regard to the rapid development of inexpensive positioning technologies, trajectory data and other data related to dynamics of moving objects are obtained in quickly increasing quantities (Andrienko et al., 2007). However, moving objects are influenced by various factors that impact and constraint their movements (Dodge et al., 2008). For example, the movement of bird is affected by temperature and vegetation distribution (Underwood and Chapman, 1985) or the movement of airplane is affected by weather conditions (Ahlstrom, 2005). These influencing factors hide significant knowledge for understanding and predicting the movements. In order to get an understanding about movement behavior of dynamic objects, it is required to take a closer look at the movement itself. In other words, it is necessary to know not only geographical space and time (Butenuth et al., 2007) but also what and how external factors affect and limit the movement (URL.1). When the causes and consequences of object movement behaviors are determined, prediction models can be established. To evaluate and verify the quality of these models, the predicted data of movement behaviors and influencing factors need to be compared with actual data.

Due to the need of understanding behaviors of moving objects and their relationship of them with influencing factors, integrated data sets including these factors are needed. With the diversity of available data sources nowadays, data integration becomes a good and applicable solution to bring a clear link to answer user's questions.

As mentioned before, this research intends to use iceberg trajectories as a case study of integrating grid and trajectory data. Since icebergs play a significant role in navigation, communication networks and climate change studies, monitoring and tracking icebergs have been considered and performed by many researchers. *"Icebergs might be a huge threat if they go in the shipping routes, causing risks of collision with ship navigation. They may also have impact on fixed and floating oil installations, and risks of scour on sub-sea installations, such as pipelines and wellheads. The consequences of iceberg impact can be severe in terms of the structural integrity of installations, and the safety of personnel and environment"* (Veitch and Daley, 2000). Climatologists also use iceberg data to study climate change because of the potential sensitivity of icebergs. The process of calving of large iceberg can help to indicate the rapid change of climate, and have a potential influence on the global sea level increase (Kenneally and

Hughes, 2006). Movement or calving of icebergs are influenced by temperatures, wind, tides, radiation, seasonal variation, and many other parameters (Benn et al., 2007; Jansen et al., 2007). Thus, the movement and disintegration of icebergs cannot be seen as independent events. With lack of required functionality, it is difficult to integrate, analyze and understand the current movement patterns and hard to verify evolution models of iceberg.

Large Antarctic icebergs are identified and tracked by National Ice Centre using a variety of satellite sensors from 1976 till present. The dataset is offered freely by NIC through the web in Excel format and contains some information as iceberg name, time (registration and observation data), position (geographical coordinates as latitude and longitude), iceberg size and name of satellite (that observe a particular iceberg). The temporal resolution is variable from 1 to 5 days. Considering metocean data, data of metocean variables as sea-surface temperature, wind, currents can be accessed in grid format from multiple data sources. Grid format is generally thought as a set of numbers making up a two, three or multidimensional array of parameter values.

2. Data integration techniques

Continuous data such as metocean data (temperature, wind, humidity etc.), Normalised Difference Vegetation Index (NDVI), and so on are commonly stored in grid formats by both two main communities, namely Earth Science (ES) and Geo-Information Science (GIScience) communities. The development of grid data model for continuous data is considered and still being continue in many organizations in many organizations. However, this development in each organization is in different degree of maturity. ES community, especially Unidata organization - with over 20 years experience in this field, includes almost all grid data providers and contains the largest grid data resources. Moreover, concerning the data structure, Earth Science data structure includes time as a dimension. This is especially advantageous for the purpose of integration with trajectory data. Therefore, for this research, we decide to concentrate on multidimensional grid data of the ES community.

Earth Science and Geo-Information Science communities have the same objective of making spatial data freely and widely available through web services. However, they have different concept regarding the support of the services. The Open Geospatial Consortium (OGC) - a primary participant of Geo-Information Science community only provides the definition and standards for the architecture of the web services. OGC is not involved with client-server methodologies and is not concerned with the implementation of the service. Unidata provides a set of services including protocols, client/ server software, data management toolkits and libraries to build up client/server software (see Figure 2-1). Therefore, we opt Unidata architecture as the best model to build up an integration system of trajectory and grid data.

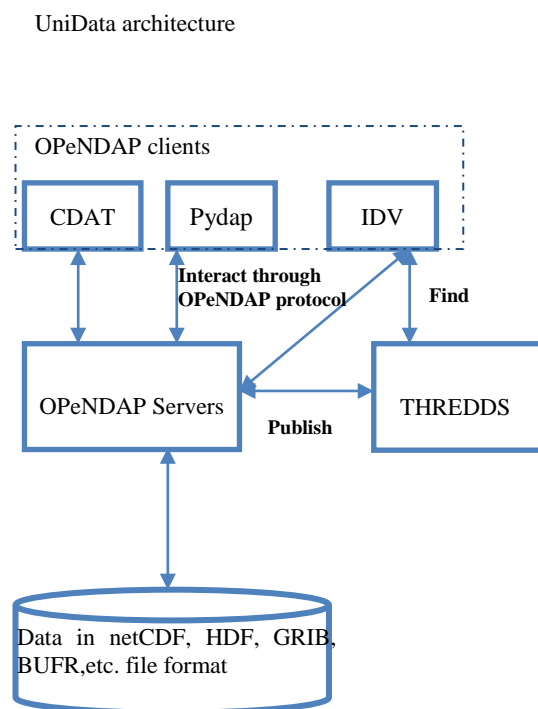


Figure 2.1: UniData architecture

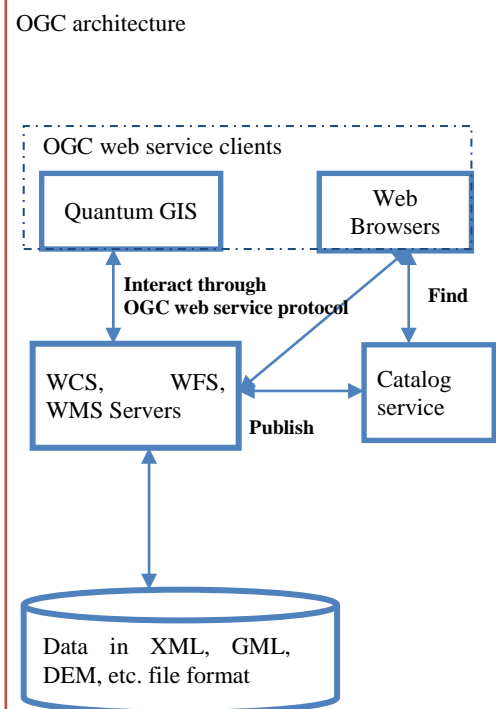


Figure 2.2: OGC architecture

Among protocols is provides by Unidata, the OPeNDAP stand for “Open-source Project for a Network Data Access

Protocol”, is the one of most popular protocol with many available OPeNDAP servers. The key issue in the distribution of science data is the incompatibility of data formats and web service interfaces. While the interoperability issue of web service interface is still being tried to handle, the problem relating to data formats has been addressed since the establishment of OPeNDAP. The practice in the scientific data analysis is that data are stored in large amounts in different file formats, and each of them could be read and written only by its own specialized Application Programming Interface (API). A part of them uses common standards, while others remain in less standard formats. Most of these formats are incompatible with others. Therefore, these issues become obstacles for scientist to freely share their data and impede correspondence with central data archives as well. OPeNDAP is a client-server protocol used for accessing and sharing of grid data. It allows users to access data from remote sources and then to store these data in the file format that users want. It also offers both a server software, to make data available on the remote servers, and a client software to access those data. The main objective of OPeNDAP is to have a computer program that can operate on data provided by different institutions and stored in various format on a remote computer. Therefore, we decide to use the OPeNDAP protocol and grid data from OPeNDAP servers.

One important is the selection of a suitable client. There are quite a lot of OPeNDAP clients such as IDV, Ferret, Pydap, GrADS, MATLAB... The selection of suitable client was based on the following criteria: Client should be extendable - should have scripting functionality, able to read trajectory data, possible to integrate into client/server architecture, support reading and analyzing grid data. Priority is given to open source, non commercial client software. IDV- a free and open source software satisfies all of these criteria and is also supported by many build-in functions (much more than others) therefore we decided to choose IDV as the best client software to use.

IDV allows users to create simple formulas by IDV Scripting Language or use the Jython language to execute calculations that are more complex. Jython is an implementation of Python programming language to run on Java environment.

3. Designing and implementing the prototype for data integration system

3.1 Desktop-based integration mechanism

For purpose of integrating data by using IDV, trajectory file needs to be stored in local directory and data source links need to be indicated. Jython script is used to get grid data from remote servers and integrate with trajectory data. The input trajectory data should be formed in a specific type. The trajectory dataset should be stored in text comma separated value (CSV) file format. To make IDV understand the trajectory dataset, the metadata has to be specified. One more requirement for trajectory data is that there must be latitude, longitude and time field in the data. Generally, grid data are heavy thus getting the whole grid dataset can lead to full memory problem in computer. The solution is that instead of getting the whole grid dataset at once, we get a subset of the grid corresponding to each trajectory record. The size of iceberg is large (only icebergs with area are bigger than 10 nautical miles are stored in database (URL2)). Therefore, an iceberg may overlay with a number of grid cell. Since we have information about approximate values of iceberg’s dimensions, but we do not have any information about iceberg’s orientation, we assume that all icebergs are square with side length is equal to the dimension length of the iceberg and the coordinates of the iceberg are extracted from its centre.

With large moving object such as iceberg, only the values of grid data at centre point may not be enough to give a relationship indication between influencing factors and moving object. Therefore, minimum, maximum and average values of grid data corresponding to the area of moving object are also acquired.

Frequently, a grid has missing cell values, therefore the interpolation methods are necessary for estimating missing values that lie in between the known cell values. There are many interpolation methods for weather data such as optimal inverse distance, cubic splining, kriging or cokriging etc. The key for selecting an interpolation method is the understanding about data. Based on the nature of particular data and application domain, the corresponding interpolation methods should be applied. However, for this research, we do not have the ambition to go deeply into interpolation issue and support users by all possible interpolation methods. Data fields such as temperature, wind speed or current speed, this prototype will offer only two methods for data interpolation. These two methods are: inverse distance weighting functions and interpolate from grid values of the record before and the record after the time of missing value. To evaluate the uncertainty of interpolated value, we built up a method that is based on two assumptions: firstly, measured values closer to the estimated value will have more influence than those are farther away and secondly, the more measured values are used to estimated, the more accuracy of the estimated value is.

3.2 Building up data integration web service

The previous section introduced data integration mechanism what is generated by using IDV and Jython script. To generate data integration mechanism, we assumed that trajectory data file, text files containing grid data source links, preferred data choices, interpolation method and time accuracy are already stored in local directory. However, the

objective of this research is to support diverse users from research fields and different data integration needs through web environment. Users should not be required about knowledge of IDV or IDV programming, they are only required for the understanding about data source location they need. Therefore, we build up a web service base on the functionality of IDV and Jython language (see Figure 3-1). To embed IDV on the server side, two problems are solved. The first problem is IDV software and scripts could be automatically evoked on a server when there are requirements from users. Another is IDV could acquire trajectory data file, grid data source links and options for data interpolation, time accuracy that are specified by users through the user interface.

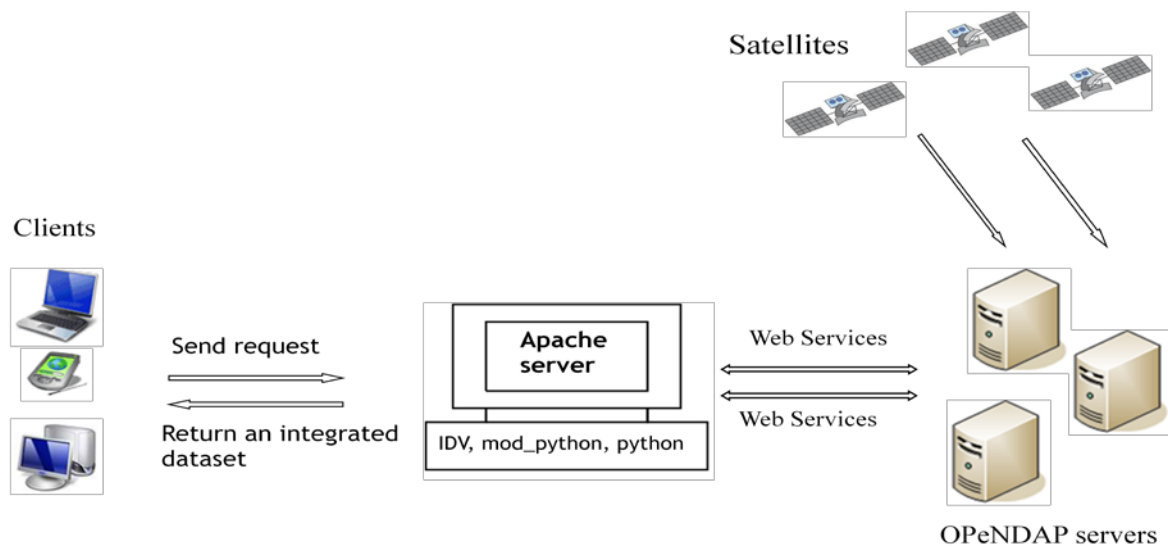


Figure 3.1: The conceptual model of data integration system via web service

The User Interface (UI) allows users to provide their input to the system and to get back the response from the server (output). The UI consists of HTML forms to acquire data source links, type of data (grid or image), and trajectory data file, preferred method for data interpolation and time accuracy (Figure 3-2) from users.

Data integration service

Please provide OPeNDAP grid data source links:

Data source 1:
 link:
 spatial resolution: lat long

Data source 2:
 link:
 spatial resolution: lat long

Data source 3:
 link:
 spatial resolution: lat long

Data source 4:
 link:
 spatial resolution: lat long

Data source 5:
 link:
 spatial resolution: lat long

Data source 6:
 link:
 spatial resolution: lat long

Data source 7:
 link:
 spatial resolution: lat long

Please select interpolation method:

Time accuracy at:

Time accuracy at:

Time accuracy at:

Time accuracy at:

Time accuracy at:

Time accuracy at:

Time accuracy at:

Figure 3-2: HTML form for putting data source link, spatial resolution, interpolation method and time accuracy

Python script was integrated with Apache by mod_python. It takes on three main roles: a) connecting client and server sides (or in other words is receiving input from users and sending output back); b) writing data and requirement from user to text files (these text files will be read by using IDV and Jython script); and c) evoking IDV and Jython script through command line.

After submitting the request through forms, users will retrieve the integrated dataset in a comma-separated values (CSV) form.

As an example, There were 289 iceberg trajectory records and three data sources links were used (sea surface temperature, U and V components of wind direction). These grid data are hosted by TPAC (Tasmanian Partnership for Advanced Computing) as below:

- Grid data source link of U component of wind in 1995:
dods://ngportal.sf.utas.edu.au:1/thredds/dodsC/library/ncep1/surface/uwnd.sig995/uwnd.sig995.1995.nc
- Grid data source link of V component of wind in 1995:
dods://ngportal.sf.utas.edu.au:1/thredds/dodsC/library/ncep1/surface/vwnd.sig995/vwnd.sig995.1995.nc
- Grid data source link of sea surface temperature in 1995:
dods://ngportal.sf.utas.edu.au:1/thredds/dodsC/library/auscom/forcing/sst_restore_1995.nc

This prototype was build on a personal computer (Intel Core 2 Duo CPU 2.26GHz) with 512 megabytes are used for Java memory. The processing time was about 680 seconds (11 minutes) in total. When applying iceberg trajectories and grid data (sea surface temperature, U and V components of wind direction), the server had to download data subset 867 times and the time consumption for downloading one data subset was 0.75 second. The method to get grid value at specified location and time was used 5780 times and the time consumption for each time was 0.0009 second.

To illustrate attributes sea surface temperature, wind direction of integrated dataset, we use Google Earth software to visualize iceberg trajectories. Icebergs are represented by arrow symbol. Wind direction attribute is represented by arrow's orientation and sea surface temperature attribute is represented by color saturation. According to the integrated dataset, values of sea surface temperature vary from - 53.620 to 1.713 C degree, we classify these data into five groups corresponding to five color saturations.

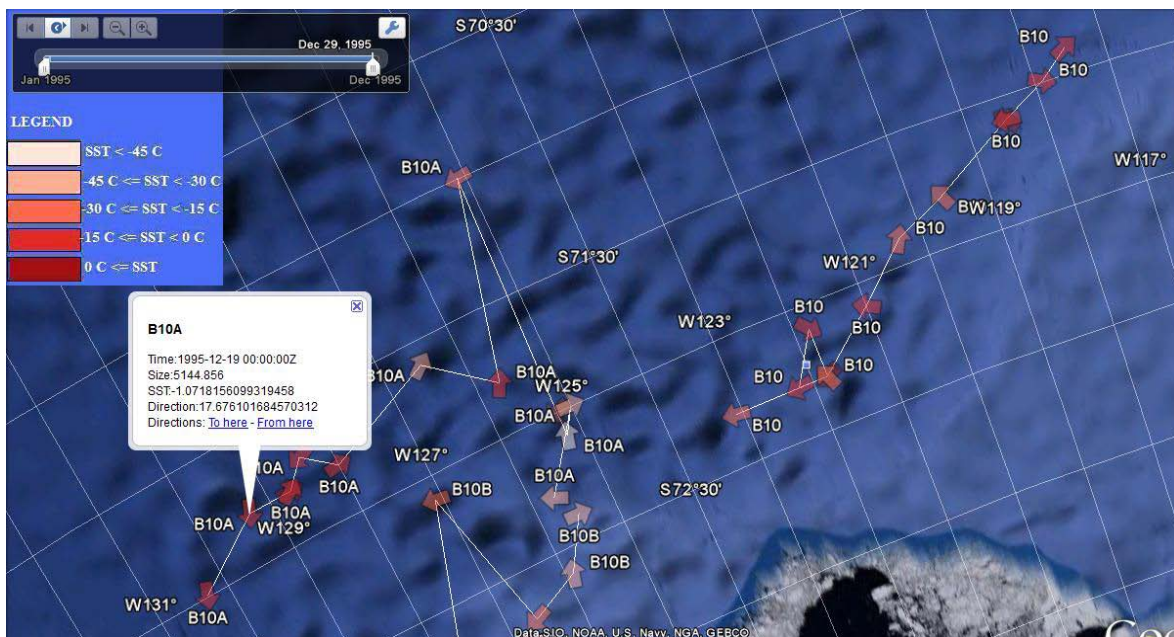


Figure 3-3: Visualizing integrated dataset on Google Earth

4. Recommendations

There is an issue relating to the performance of system needs to be concerned. It is the time consumption for processing of integration mechanism on server. While was designing the prototype, two scenarios were raised. The first scenario, the whole grid data are obtained and integration, interpolation processes will be done with this data grid, the advantage of this scenario is the system have to connect to server and download grid data only once. The disadvantage is that it requires a powerful server with large memory and the process of integration or interpolation will have to work with a big grid, this will consumes more time for processing (comparing to process with small data grid). The second scenario, instead of obtain the whole data grid, only subset data (which based on assumed area of iceberg and record time) is achieved. The advantage and disadvantage of this scenario are opposite with the first one.

It does not require much memory of server, however it needs to download data many times (depends on the number of trajectory records). To select the better one, both these scenarios should be test on a real server computer. The criteria should be the time consumption and also the stability of server in case many users request service at the same time.

Data interoperability is always an important issue. Considering the trend in both Earth science and Geo-Information science communities, they are moving closer to each other on the issues of standardization. It is a promising trend for this prototype to extend its capability to support user integrating many other types of data from both these communities.

Acknowledgments The authors wish to thank Mr Valentijn Venus for his advice and giving some IDV script

References

- DODGE, S., WEIBEL, R. & LAUTENSCHUTZ, A.-K. 2008. Towards a Taxonomy of Movement Patterns. *Information Visualization*, 7.
- ANDRIENKO, G., ANDRIENKO, N. & WROBEL, S. 2007. Visual analytics tools for analysis of movement data. *ACM SIGKDD Explorations*, 9.
- UNDERWOOD, A. J. & CHAPMAN, M. G. 1985. Multifactorial analyses of directions of movement of animals. *Journal of Experimental Marine Biology and Ecology*, 91, 17-43.
- AHLSTROM, U. 2005. Work domain analysis for air traffic controller weather displays. *Journal of Safety Research*, 36, 159-169.
- BUTENUTH, M., GÖSSELN, G. V., TIEDGE, M., HEIPKE, C., LIPECK, U. & SESTER, M. 2007. Integration of heterogeneous geospatial data in a federated database. *ISPRS Journal of Photogrammetry and Remote Sensing*, 62, 328-346.
- VEITCH, B. & DALEY, C. 2000. Iceberg Evolution Modeling. National Research Council of Canada. OERC Report - 2000 - 002
- KENNEALLY, J. P. & HUGHES, T. J. 2006. Calving giant icebergs: old principles, new applications. *Antarctic Science*, 18, 10.
- BENN, D. I., WARREN, C. R. & MOTTRAM, R. H. 2007. Calving processes and the dynamics of calving glaciers. *Earth-Science Reviews*, 82, 143-179.
- JANSEN, D., SCHODLOK, M. & RACK, W. 2007. Basal melting of A-38B: A physical model constrained by satellite observations. *Remote Sensing of Environment*, 111, 195-203.

URLs

1. Patterns-of-Movement (2008) [cited 2009 15/08]; Available from:

<http://movementpatterns.pbworks.com/Patterns-of-Movement>

2. National Ice Centre- Mission Statement [cited 2009 08/09]; Available from:

<http://www.natice.noaa.gov/index.htm>