SPATIO-TEMPORAL INDEXING OF THE QUIKSCAT WIND DATA

Félix Rodríguez Rodríguez, Manuel Barrena García.

Department of Engineering of Informatic and Telematic Systems,
University of Extremadura, Cáceres, Spain.
{felixr,barrena}@unex.es

1. INTRODUCTION

Global wind data are increasingly used in many research fields, such as environmental sciences, biology, meteorology, and climate sciences. Satellite altimetry provides a unique opportunity to understand the dynamics of sea winds. Currently, the most widely data source used to work with global spatiotemporal wind information (e.g., an area covered by an evolving storm) is provided from the SeaWinds/QuikSCAT (QuikSCAT) satellite [1], an Earth-orbiting NASA scatterometer (a microwave radar). Mainly because it provides global coverage of ocean winds with an unprecedented view, and data are available at a sufficiently large spatial scale and high temporal resolution. A clear example is a recent work of great impact that use the spatiotemporal component of winds [2], where, as the authors say: ”Winds are one of the major agents driving movements of any air suspended object, from a particle to an albatros, particularly true in the marine environment, where winds are stronger and more constant than on land, and the apparent lack of barriers facilitate aerial movements”.

If we need to query about any isolated point without reading complete files, to select and query any particular wind area, to run complex queries, to make easier the update of existing data and the insertion of newly acquired data, to facilitate access to data at different resolution levels, to provide concurrency, buffering, data log, and access control, or to supply temporal evolution of data, it is extremely useful to maintain indexed all the global wind data provided by QuikSCAT, enhancing the retrieval process of the desired data by means of an efficient indexing platform that will serve the data requested by the user. A plethora of index structures have been proposed in literature [3, 4], but this type of highly variable data (wind data have a different value every time) needs a completely adapted index. Thus, we have developed a novel indexing method particularized to the QuikSCAT data since the satellite was launched on 1999.

2. MATERIALS AND METHODS

The SeaWinds on QuikSCAT mission daily provides data at a 25-km, with a typical accuracy of 1 m/s in speed, and 15° in direction (12.5-km resolution is also available). The data follows the NASA specific HDF4 format [5], laborious to handle if we want use them combined with other data sources, as they might be from the Terra satellite, or from the Shuttle Radar Topographic Mission (SRTM).

Wind data are daily downloaded via FTP from the QuikSCAT site (*ftp://podaac.jpl.nasa.gov/ocean_wind/quikscat/L2B*). After that, wind data are transformed into the appropriate index tuples to be massively loaded onto the index structure. Our index prototype combines both the Q-tree [6] and the LV-tree [7] to build a new spatiotemporal index structure well suited to the extremely variable data that QuikSCAT provides. The Q-tree is a hierarchical multi-purpose, dynamic, balanced and paginated (blocks of data are used) multidimensional index structure with concurrency and log file mechanisms. Index and data are clearly separated and every node is homogeneous in structure. All the nodes of the tree correspond to disk pages, and the structure is designed in such a way that the number of nodes visited during the search and insert processes is minimized. The Q-tree is used as the spatial index. The LV-tree, or Linear Version tree, is a hierarchical multi-version data structure, dynamic and paginated. The LV-tree is used as the temporal index. The index is partially persistent, *i.e.*, all versions can be read but only the most recent can be updated to create a new version. Combining the two previous structures, we build the new indexing platform, called Q-full-tree.

To massively load the daily wind data (bulk loading process), initially we follow the same method we have use for indexing the spatial SRTM data [8], but data pages are not saved onto the Q-tree. The Q-tree stores a spatial region identifier (a *key*)

This work was supported by ”Ministerio de Educación y Ciencia”, Spain (Project grant TIN2005-05939).

that fully accommodates all the data involved in each data page. Thus, both the data pages and their spatial region identifier are stored onto the LV-tree, updating the version of every data page inserted. The expensive process of data classification in all the spatial dimensions is avoided with the static and very orderly disposition of the QuikSCAT data files. In order to make faster the bulk loading process, we consider spatial regions of wind data that completely fill a data page with a single inserting operation. Figure 1 shows the general outline of the bulk loading process of wind data.

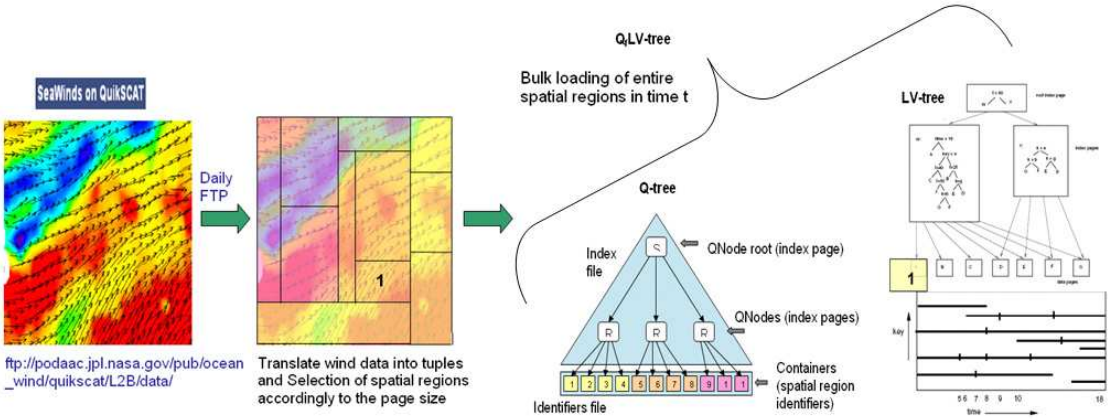

Fig. 1. General outline of the bulk loading process of QuikSCAT data onto the Q-full-tree. It shows a partial daily bulk loading of one wind data page. A spatial region is selected (identifier=1) according to a page size previously established. The spatial region identifier and the data page are inserted onto the Q-full-tree.

3. CONCLUSIONS

A completely adapted index structure for the QuikSCAT wind data is presented: the Q-full-tree. The spatiotemporal index integrates two existing structures: *(i)* the Q-tree, as the spatial index, and *(ii)*, the LV-tree, as the temporal index. Wind data are daily bulk loading onto the Q-full-tree, facilitating spatial queries (point and window/range), temporal queries (time slice, and interval time) and spatiotemporal queries by combining the spatial and temporal queries. The index provides a unique and useful way to link wind data with other, or be embedded in an open GIS as it has been done with the SRTM data [8].

4. REFERENCES

[1] Phisical Oceanography, DAAC, "SeaWinds on QuikSCAT," http://podaac.jpl.nasa.gov/DATA_CATALOG/quikscatinfo.html.

[2] A. M. Felicísimo, J. Muñoz, and J. González-Solís, "Ocean Surface Winds Drive Dynamics of Transoceanic Aerial Movements," *Plos ONE*, vol. 3, no. 8, pp. e2928, ago 2008, doi: 10.1371/journal.pone.0002928.

[3] M. F. Mokbel, T. M. Ghanem, and W. G. Aref, "Spatio-temporal access methods," *IEEE Data Eng. Bull.*, vol. 26, no. 2, pp. 40–49, 2003.

[4] M. Vassilakopoulos and A. Corral, *Encyclopedia of Database Technologies and Applications*, chapter Indexing of Moving Objects and other Spatio-Temporal Data, pp. 652–657, IDEA Group Publishing, 2005.

[5] The HDF Group, "Hierarchical Data Format," http://www.hdfgroup.org/products/hdf4.

[6] E. Jurado and M. Barrena, "Efficient similarity search in features spaces with the Q-tree," in *Proceedings of the International Conference on Advances in Databases and Information Systems, ABDIS*, 2002, pp. 177–190.

[7] B. Salzberg, D. Lomet, M. Barrena, and L. Jiang, "The BT-Forest: A branched and temporal access method," 1999, http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.38.2332.

[8] F. R. Rodríguez, J. J. Hernández, and M. Barrena, "A tool to query and visualize the complete SRTM data set indexed by the Q-tree in an open GIS," in *IEEE International Geoscience and Remote Sensing Symposium, IGARSS*, Boston, MA, USA, 2008, vol. 3, pp. III 1410 – III 1413, d.o.i.:10.1109/IGARSS.2008.4779625.