GIS solution for weather forecast data analysis

Krzysztof DRYPCEWSKI, Łukasz MARKIEWICZ*

Keywords: gis; numerical weather forecast; weather prediction; newr ems; rasdaman; data analysis

Abstract: In this paper authors present the GIS system for the analysis of the numerical weather prediction data. This kind of data has multidimensional character (three dimensions and time) and its analysis should consider all the available factors. Proposed GIS system consists of RASDAMAN application with implemented OLAP cube mechanism, which enables the user to process data in the spatial-time domain. It also simplifies the meteorological data division into specific geographic and temporal ranges.

1. Introduction

Geospatial data in a raster or vector form is commonly used in a wide range of computer systems. It fulfils a wide range of functions, from the representation of geographic regions, visualizing recent trends in population migration, land and sea navigation to logistic, marketing or biological research. Geographic Information Systems (GIS) provide multiple features, e.g. geospatial data storage, analysis and visualization that can be helpful in solving various problems. In this article authors propose GIS for the processing of weather forecast data and one of its possible applications. Described solution operates on the data received from the numerical weather prediction (NWP, performed by NEWR EMS software) and simplifies information segmentation into the specific geographic areas and temporal ranges.

2. System design

GIS can present and process various type of data, e.g. meteorological forecasts. This kind of data is predominantly three-dimensional (the images are two-dimensional, but all the values are typically calculated on various heights) and is depicted in the spatio-temporal domain. The main function of the proposed system (Fig. 1) is to

*Department of Geoinformatics, Faculty of Electronics, Telecommunications and Informatics, Gdansk University of Technology
provide the possibility to analyse specific fragments of the raster data from the NWP in the specific region, height and time. It uses general-purpose GIS applications to create dedicated system for meteorological research. All of the input data, obtained from the NEWR EMS software (which already runs on a three-computer cluster and is described in the third chapter) is stored in a relational database and is synchronized in fixed time intervals.

The aforementioned input can be referred as multidimensional, it carries 4-level information about weather variables values in relation to their geographic coordinates at specific heights and at a given time moments. Storing and processing this kind of data could become problematic. These tasks can be simplified using the OLAP (OnLine Analytical Processing) structure. It is dedicated solution, which consists of an array with any amount of dimensions - each of them represents different data parameter.

It is modelled as a cube divided into smaller fragments. This structure must be stored in a dedicated database, which additionally enables hierarchical arrangement of data.

In the presented work an OLAP cube was constructed from meteorological data received from numerical weather prediction process. Consecutive records created by NEWR EMS describe changes in the state of the atmosphere over the analysed domain. Each of the prediction variables is represented in the form of a two-dimensional array, where each cell stores a numerical value of given weather parameter (e.g. temperature, pressure, moisture). Those data sets are referred to as Multidimensional Discrete Data (MDD) and share common characteristics like potentially unlimited number of dimensions, spatial extent or cell semantics. Typical, popular database management systems (DBMS) do not provide the sufficient methods of storing MDD. Most of the available solutions focuses on two-dimensional (2D) instead of n-D data, therefore authors chose a specialized tool (Rasdaman) in accordance to GIS specification proposed in this article (Sarawagi and Stonebraker, 1994; Libkin et al., 1996;

Fig. 1. Proposed system architecture
This solution produces processed meteorological raster data, which can be visualized and shared by using various presentation mechanisms (e.g., WPS, WCS, WCPS etc.), which are not predefined and can be chosen according to the specific application of the system.

3. NEWR EMS Weather Prediction System

NEWR EMS (NEMS and Weather Research & Forecast Environmental Modelling System) is a software package that consists of various independent modules for the Unix-based systems (Rozumalski, 2013). These modules cooperate and perform different tasks during the NWP (numerical weather prediction) process. Various sets of algorithms are used to simulate the future states of the atmosphere using current and the past weather conditions. All software packages are coordinated by the editable scripts and configuration files located in the predefined domain, which is created in the DOMWIZ program before the process initialization.

Domain contains coordinates of the area selected for the weather forecasting along with the needed geographic and system parameters (e.g., resolution, map projection, weather simulation model). Every system run begins with the EMS_PREP routine, which downloads needed geospatial data and the current weather conditions (which are the input of our NWP process) over the selected region from the GFS (Global Forecast System) service. When the initialization is completed, the chosen weather model is started by the EMS_RUN routine. Each model, e.g., NMM, NEMS NMM-B or ARW, uses a given set of algorithms, which analyse current and predict future atmospheric variables’ values.

Each model has its application in terms of computational power requirements, output accuracy and the recommended prediction scale - global or regional. When the model run is completed the EMS_POST script processes the binary data produced earlier and creates various types of the output files (e.g., in GRIB or GRADS format, Fig. 2). They define the numerical values of all the weather variables in the specific time and geographic position (within the chosen domain). Output data can also be created for various pressure levels (heights) set in the configuration files.

4. Rasdaman

The GIS application server core is implemented using Rasdaman (Baumann et al., 1999; Anon, 2013), which architecture corresponds with the established technological standards (Baumann, 2013a). It was designed using client-server model (Fig. 3) approach that allows query processing to be held completely on the server side. Rasdaman is the piece of software that aims to provide a database service for the MDD structures in a domain-independent manner. It is an array database management system (ADBMS). It extends standard relational database with the ability to store multidimensional raster data and is available in both open-source and commercially supported versions (Libkin et al., 1996; Marathe and Salem, 1999; Baumann et al., 2000; Anon, 2013).

User’s MDD data is processed by Rasdaman and stored in the underlying database, called base database. Open-source Rasdaman can be easily embedded in PostgreSQL...
PhD Interdisciplinary Journal

64

PhD Interdisciplinary Journal

PhD Interdisciplinary Journal

commercial version also in MySQL, Oracle, IBM Informix and DB2 databases). Alphamnumerical information that is not in an array form is stored outside Rasdaman, but reference between alphamumeric and MDD data can be established (Baumann, 2013a).

Rasdaman consists of several components that implement many of the system features. Description of the MMD is handled by Rasdaman definition language and the RasDL – tool for the creation of the data schema. User can define data types that will identify an unrestricted number of instances; no other information is required to access the database (Baumann, 2013a). Schema consists of three elements: the cell type definitions (similar to C++ primitive types), the MDD definitions (array with spatial domain over base type) and the collection type definitions (descriptions of the arrays of the same type).

Rasdaman query language (RasQL) enables the user to communicate with databases by usage of specific queries (Baumann, 2013b). RasQL syntax is similar to structured query language (SQL) and extends standards defined by the International Organization for Standardization (9075, 1992). It is based on an array algebra that incorporates MDD operators. As in the SQL, RasQL SELECT statement returns set of arrays. MDD operations can be used both in SELECT and WHERE queries (Baumann et al., 2000; Sarawagi and Stonebraker, 1994). Results are sent to client in a specified format. PNG, TIFF, JPEG types are supported. Spatial indexing, compression and tiling are also provided.

rView is an interface to Rasdaman for visualisation of MDD, i.e. multidimensional raster data and database server maintenance. Communication with DBMS is based on sending RasQL and RasDL queries. Received data can be visualised in the Graphical User Interface (GUI) provided in form of a chart, flat image or 3D view table (Baumann, 2013b).

The meteorological data are stored in a dedicated form (data files) inside the heterogenous store. At regular intervals, or on demand, the data are synchronized and replicated to the common multidimensional database. Rasdaman interprets multidimensional data as spatio-temporal cubes, where the first two dimensions are the pixel position and the third dimension represents time parameter. In this context a section

![Fig. 2. Predicted surface temperature (K) on 24.04.2012 15:00 (UTC+2)](image)

Fig. 2. Predicted surface temperature (K) on 24.04.2012 15:00 (UTC+2)
operation can be defined (Fig. 4). A section allows extracting lower-dimension layers (also called slices) from n-D array. Operation can be performed on each of the dimensions and is accomplished by specifying the slicing position. This operation reduces dimension of the data received from this operation by 1.

After the processing request is received, the working module analyses the MDD data and returns the results to the WCS form. Modules may partly replicate the data to ensure the higher responsiveness and efficiency of the system. For example, the presentation layers are also available in the form of WMS services.

5. Data processing & analysis

The following section contains the sample results of processing meteorological data by means of the proposed GIS system. After the spatio-temporal cube is constructed, the Graphical User Interface (GUI) is used for constructing sample queries, GUI is also responsible for establishing the data base connection. After the queries execution the results are returned to the user. The tool for results preview is presented in Fig. 5. The results window presents information about the obtained data (i.e. type, resolution,
size and spatial domain). The GUI is equipped with several specialized tools for data visualization, that allows the user to choose from the following activities: flat image, volumetric, orthosection, height field, chart, table, sound and string. The results are presented in the Fig. 5.

Cube analysis can be performed by using built-in Rasdaman features. Data set presented in the article is a time series of the meteorological three-dimensional images: time, longitude and latitude (height information is optional). Performing slice operation on temporal dimension results in the state of the atmosphere in the chosen time moments. More interesting results are received with section operation on geographical position dimensions. Fig. 6, 7 present slices in vertical and horizontal dimension that represent the temperature values on the height of two meters in time consecutively for the selected parallels or meridians.

6. Conclusions

In the article authors described the GIS system that enables the user to operate on the multidimensional data from the weather forecast processes. By using various implemented OLAP mechanisms, the analysis of the meteorological data in selected time and geographic domains is simplified and can be automated by configurable scripts. Operational testing using the real NWP output proved that this solution is efficient and provides significant assistance to the user during weather data processing and segmentation. Future work will focus on the presentation layer using GIS web
server, e.g. Geoserver and further testing using more specific meteorological data.

References

9075, ISO (1992), Database Language SQL.
Anon (2013), Rasdaman website.
URL: http://www.rasdaman.org/
Baumann, P. (2013b), rView the rasdaman DBMS Visual Frontend.
URL: http://strc.comet.ucar.edu/software/newrems

Fig. 7. Horizontal slice