Advances in Engineering Software 45 (2012) 360–369

Contents lists available at SciVerse ScienceDirect

Advances in Engineering Software

journal homepage: www.elsevier.com/locate/advengsoft

A GIS-based management and publication framework for data handling of numerical model results

J.J. Yu [a,b], X.S. Qin [a,*], L.C. Larsen [b], O. Larsen [b], A. Jayasooriya [b], X.L. Shen [b]

[a] School of Civil & Environmental Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798, Singapore
[b] DHI-NTU Water & Environment Research Centre and Education Hub, DHI Water & Environment (S) Pte. Ltd., 200 Pandan Loop Pantech 21, #08-03, Singapore 128388, Singapore

ARTICLE INFO

Article history:
Received 1 June 2011
Received in revised form 15 September 2011
Accepted 16 October 2011
Available online 15 November 2011

Keywords:
Time series
Numerical model results
GIS
Web map service
Web service
Data atlas

ABSTRACT

There has been an increasing awareness that large-scale datasets of earth-system model results would be produced and accumulated rapidly within the next few decades with extensive usage of numerical models. The successful exploration of these data for scientific research and engineering applications requires the capability of efficient storage, retrieval, integration and visualization of these large datasets. Robust publication methods are also needed to enable data exchange through the web to increase global research collaborations. In this study, a comprehensive GIS-based data management and publication framework (GMPF) was developed and applied to the use, representation, and analysis of hydrodynamic modeling results of Hamburg Port, Germany. Linking with GIS, a desktop- and a web-based model-results data atlas were developed to provide an interactive environment for various cartography and time series plots. The implementation of the framework was based on modular software design, open specifications and reuse of open source projects. The study results showed that the project life-cycle was shortened and development efforts were reduced through applying GMPF. The developed system could be integrated into commercial hydrodynamics software platforms to enhance the capability of model-results management and publication.

1. Introduction

Numerical models have been widely used to describe complex water resources and environmental phenomena (e.g. flood inundation, contaminant transport and wave propagation) under various conditions and scales. Over the past decades, tremendous datasets of model results have been generated and accumulated in many scientific and engineering research works. In order to share, understand, examine and compare various modeling conditions and outputs easily and interactively, there is a need for building the capability to access, organize, integrate, visualize and analyze these data [1]. Moreover, the published results from local research projects have the potential to advance the global research collaborations. Therefore, it is also desired to have a robust method to publish the model results that can be easily discovered and utilized for visualization and analysis in distributed networks.

Geographical information system (GIS), with its capability to display, manipulate, and analyze geographically referenced information, provides an integrated user-interactive environment to compile data into a manageable system [2]. The advancement of the spatial database techniques improves the capability of GIS to handle large datasets efficiently. Over the past decades, it has become an integrated platform for data storage, manipulation, retrieval, visualization and analysis, in support of water resources modeling and management studies [3–5]. Liang and Molkenthin [6] developed a virtual GIS-based hydrodynamic model for Tamshui River. Poveda et al. [7] developed an interactive digital atlas of Colombia's hydro-climatology. Ng et al. [8] presented the integration of a GIS and a complex three-dimensional hydrodynamic, sediment and heavy metal transport numerical model, and demonstrated its applicability in a study case in Pearl River Estuary. In recent years, WebGIS has provided the solution to publish and disseminate the spatial related data in the internet [9]. Frehner and Brändlia [10] designed and implemented a Virtual Database consisting of a WebGIS-based framework for retrieval, analysis, and visualization of spatially related environmental data. Horsburgh et al. [11] implemented an integrated system for publishing environmental observation data through web services and demonstrated how to create a national research data network.

The previous studies have demonstrated the applicability of utilizing GIS as an integrated environment for manipulating hydrologic and environmental data in intranet and internet. However, limited studies have discussed its capability in managing, analyzing, and publishing results of numerical models, particularly in cases when the datasets are extremely large. In fact, when integrating the modeling results into a GIS environment for advanced data

* Corresponding author. Tel.: +65 67905288; fax: +65 67921650.
E-mail address: xsqin@ntu.edu.sg (X.S. Qin).

0965-9978/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.advengsoft.2011.10.010

management in both intranet and internet, the following challenges could be encountered:

(1) Many software packages ranging from open sources (e.g. OpenFOAM) to commercial products (e.g. Mike21/3 FM) provide powerful modeling and equation-solving capabilities. The generated results usually represent a number of physical variables (e.g. discharge, velocity, water surface, etc.) with temporal and spatial variations. The file formats are specific to software platforms and organized differently from one vendor to another. It is thus desired that a generic database structure be advanced to manage the large model results uniformly and an open interface be designed to handle the heterogeneous data formats seamlessly.
(2) The size of the model results generated from high resolution numerical models is usually varying from hundreds of megabytes to gigabytes, where the high resolution means the model domain is discretized spatially into lots of small computational meshes. For instance in this study, totally 73,682 meshes with an average of 20 m grid spacing are generated for 68.5 km^2 model area. A considerable challenge is how to develop an approach to store the huge datasets of model results with appropriate spatial and temporal indexes for quick data retrieval (e.g. retrieve time series of the flow flux for a river cross-section).
(3) The traditional GIS software packages (e.g. ArcGIS) is incapable of representing the direct results from numerical models, due to the fact that it is difficult to overlay the GIS layers with the geo-referenced model-results plots to produce an integrated map.
(4) There is a lack of a web-based interactive environment for discovering, exchanging and publishing the geo-referenced model results on the web.

Therefore, the objective of this study is to present the architecture and implementation of a GIS-based management and publication framework (GMPF) for organizing, retrieving and utilizing the dynamic time series of model results in both desktop- and web-based applications. The storage and organization mechanism of respectively structured-grid-based and unstructured-mesh-based model results in the database is proposed and discussed. A database access component is implemented, with capability of performing complex time series queries for user specified locations and criteria. The developed database and access component are then integrated into a desktop-based atlas tool for efficient management of the huge amount of data associated with numerical model results. A web map service (WMS) and a web service interface are implemented to publish the two-dimensional plots and the dynamic time series of model results on the web, which provide a basic infrastructure for discovering, exchanging, and sharing the data available in the individual data site. A real-world case in Hamburg Port, Germany, is selected to demonstrate the applicability of GMPF in managing and publishing hydrodynamic model results under various tide and water surface conditions.

2. Architecture of GMPF

The design and development of the GMPF follow the rules of modular software design, reuse of open source projects, and adoption of the open standards (e.g. XML and web service) to exchange data, such that the features of modern software (e.g. economy, extendibility, open interfaces, easy accessibility and low maintenance) can be obtained. Fig. 1 illustrates the architecture of the GMPF From a logical perspective, it consists of: (i) a *mesh database* to store and manage model results with a large capacity, (ii) a desktop model-results atlas tool to explore digital map and time series interactively, (iii) a WMS and a web service to publish the plots for web mapping and dynamic time series for data analysis, respectively and (iv) a web-based atlas tool for discovering and visualizing the model results on the web.

From Fig. 1, the *mesh database* is the central repository for the model results, geo-spatial data, and related metadata, which is a compulsory component deployed in a database server. The access component (Section 2.2) add-on is used to facilitate communication between the database and applications (e.g. desktop- and web-based atlas tool, WMS and web service). The desktop model-results atlas tool is expected to be used by modeling engineers, administrators or decision makers in the intranet to import/export, manage, and explore the model-results data. A 'network spider' routine (windows service) is scheduled to download the free-use model related data, such as the shuttle radar topography mission data (SRTM) and the tropical rainfall measuring mission (TRMM), and upload them into *mesh database* for usage. The WMS and web service can be optionally deployed in a web server. The WMS adopts the image to serve the two-dimensional plots through HTTP, and the web service adopts the Extendable Markup Language (XML) to transfer time series data using Simple Object Access Protocol (SOAP). They both can be connected to *mesh database* through the access component as we implemented in this study. They are also easy to be customized to support other data sources. The web-based atlas tool uses the WMS and web service as the data source. It can be linked to multiple servers, where WMS and web services deployed, and serve the users without considering the physical location of the data services.

2.1. Mesh database

PostgreSQL [12] with the PostGIS [13] add-on is selected to be the database management system (DBMS) as the basis of the *mesh database*. PostgreSQL is an open source object-relational DBMS and it runs on major operation systems (e.g. Windows, Linux and Unix) with the capability of performing complex queries and supporting various programming languages (e.g. C/C++, .Net, and Java). PostGIS add-on enables PostgreSQL spatially to store and manage the geographic objects in the database. It offers the application programming interfaces (APIs) to perform spatial query and analysis. The employment of the open source DBMS techniques can reduce the cost of the project in the server side without the loss of performance.

The datasets generated by the numerical models (e.g. a large-scale hydrodynamic model) are usually large (e.g. hundreds of gigabytes) with model configurations and simulated values. They can be in different formats due to the different modeling software employed. They are expected to be imported into a database and organized into a series of tables in order to obtain efficient data query and retrieval. Those tables are categorized into three types: (1) metadata tables; (2) geo-spatial tables and (3) model-results tables.

Metadata tables are a collection of tables predefined in the database to store the descriptive information (e.g. modeling area extent, spatial and temporal axis and physical variables simulated) of the model results and geo-spatial data. With the benefit of PostGIS, geo-spatial tables can be created easily with a "geometry" column enabled. Two kinds of the geo-spatial datasets are defined in the system, with one to store the feature layers (e.g. buildings, roads, rivers and land use) of the modeling areas as the background map, and the other to store the geo-referenced polygons representing the numerical calculation meshes (or grids). Model-results tables store the time series of the physical variables (e.g. velocity, flux and water depth) simulated by the models and link to the meshes spatially, such that the extraction of dynamic time series