Distributed geospatial model sharing based on open interoperability standards

FENG Min[1,2], LIU Shu-guang[3], Ned H. Euliss, Jr.[4], YIN Fang[1]

1. *Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing* 100101, *China*;
2. *University of Maryland Institute for Advanced Computer Studies, College Park, MD*, 20742, *USA*;
3. *U.S. Geological Survey, Earth Resources Observation and Science Center, Sioux Falls, SD*, 57198, *USA*;
4. *U. S. Geological Survey, Northern Prairie Wildlife Research Center*, 8711 37th *St SE, Jamestown, ND*, 58401, *USA*

Abstract: Numerous geospatial computational models have been developed based on sound principles and published in journals or presented in conferences. However modelers have made few advances in the development of computable modules that facilitate sharing during model development or utilization. Constraints hampering development of model sharing technology includes limitations on computing, storage, and connectivity; traditional stand-alone and closed network systems cannot fully support sharing and integrating geospatial models. To address this need, we have identified methods for sharing geospatial computational models using Service Oriented Architecture (SOA) techniques and open geospatial standards. The service-oriented model sharing service is accessible using any tools or systems compliant with open geospatial standards, making it possible to utilize vast scientific resources available from around the world to solve highly sophisticated application problems. The methods also allow model services to be empowered by diverse computational devices and technologies, such as portable devices and GRID computing infrastructures. Based on the generic and abstract operations and data structures required for Web Processing Service (WPS) standards, we developed an interactive interface for model sharing to help reduce interoperability problems for model use. Geospatial computational models are shared on model services, where the computational processes provided by models can be accessed through tools and systems compliant with WPS. We developed a platform to help modelers publish individual models in a simplified and efficient way. Finally, we illustrate our technique using wetland hydrological models we developed for the prairie pothole region of North America.

Key words: geospatial model, model sharing, distributed computing, geospatial interoperability

CLC number: P208 **Document code:** A

1 INTRODUCTION

Geospatial models refer to models built with Geographic Information System (GIS) support and geospatial data that serve as model inputs and outputs. When geosciences became more multidisciplinary and directed towards broader spatial scale issues, GIS became increasingly valuable for scientific modeling (Goodchild, 2005), especially for hydrological, ecological and environmental research (Wei & Chen, 2005). Countless geospatial models have been developed since GIS and remote sensing techniques have been used in scientific research. The underlying principles of those models have been published in scientific journals and presented at conferences, but the computable modules developed by modelers have made little progress in model sharing technology. The modules developed represent scientific knowledge gained from research but in most cases, will have to be rebuilt into new models because sharing techniques were not integrated into them that would have allowed a cost-effective and efficient means of sharing their modules with other modeling efforts. As model simulation and integration become more important in geographic related research and applications, sharing computable modules becomes highly desirable because highly relevant scientific modules can be shared at a great cost savings and overall efficiency (Liu *et al*., 2002). Sharing model modules makes better overall efficient use of those modules, and facilitates interdisciplinary benefits for research applications (Goodchild, 2005). Because models can generate data using special algorithms, shared models have potential to provide more data to support analyses and researches (Crosier *et al*., 2003).

Models have conceptual, mathematical, numerical, and computational module phases (Goodchild, 2003). Sharing computational modules is much more difficult than sharing their basic principles. Accessing a computational model requires a two-way conversion where the model is unavailable until accessed directly by model clients, and both sides communicate without any technical or semantic problem. Most computational models used for scientific use are command-line applications

Received: 2008-07-21; **Accepted**: 2009-05-19

Foundation: National Infrastructure of Science and Technologies (No. 2005DKA32300), National High Technology Research and Development Program (863 Program) (No. 2006AA01A120).

First author biography: FENG Min (1981—), male, assistant researcher, graduated from Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences. E-mail: fengm@lreis.ac.cn.

usually written in Fortran, C, and a host of scripting languages. These applications are fast and efficient but they are often platform-dependent and difficult to integrate with applications from other disciplines. In addition, such modules are kept at different spatial locations and implemented using various architecture and technologies, and programming languages that makes their access and utilization for remote clients difficult and cumbersome. Additionally, standards that describe their input parameters, output results or track monitoring runs generally do not exist.

Since the end of last century, systems have been designed to help scientists share and reuse their geospatial computational models. The modular modeling system (MMS) was developed by United States Geological Survey (USGS) to support stand-alone environmental modeling (Leavesley *et al*., 2006). Morozov *et al*. (2006) developed a system to share seismic models over the Internet. Granell *et al*. (2007) developed an online water resources managing system to help users predict water volume of European rivers. Those systems are able to share geospatial computational models, but there are some difficulties to be resolved before they could be widely used. First, systems running on a single computer or a closed network are not suitable for sharing resources widely. Second, they are based on different data formats and there are different model interactive methods, and this unconformity creates interoperability problems, especially for large and diverse user groups. Third, sharing copyrighted model modules and source codes can create legal infringement issues.

Geospatial theories and technologies have been improved in the last thirty years, and GIS has gone through Single-Tier and Three-Tiers to Service Oriented Architecture (SOA) (David *et al*., 2005). Comparing to Single-Tier and Three-Tier architecture, SOA is better for sharing and integrating resources over wide geographic regions and spans different systems. Sharing geospatial computational models based on SOA provides at least two advantages. First, since models generally require data, models stored on the Internet provide infrastructure for integrating data and accessing models very quickly and efficiently. Second, SOA based model sharing is highly amenable to variant applications, including Browser/Server and Client/Server and distributed computing architectures, such as GRID computing and cloud computing.

Although little progress has been achieved on open standards related to geospatial model sharing, Open Geospatial Consortium (OGC) recently published its first version of Web Processing Service (WPS) standard (Open GIS Consortium, 2008). WPS defines a standardized interface that facilitates publishing of geospatial processes, and discovering of and binding to those processes by clients. The WPS standard and other standards from International Standards Organization (ISO) and OGC, such as standards for geospatial data format (e.g., geospatial metadata) offer a solution for resolving the interoperability problems in distributed geospatial models sharing and accessing (Granell *et al*., 2007). However, problems still exist with current geospatial computer models. First, considering the huge resources provided on the Internet, shared models should not be isolated from other network services and applications. Therefore, the architecture of distributed geospatial model sharing should be designed to enhance integration between model use and other resources, and to promote effective utilization of models. Second, WPS is compatible with OGC's data, metadata, and service standards, but no explicit rule has been given by WPS on data semantic parameters, such as metadata, geospatial reference system, and measurement units. From a computational perspective, those parameters are necessary for models to perform realistic simulations. Third, OGC standards are independent of implementation, but the interface defined by WPS is too generic for practical computational model sharing (Open GIS Consortium, 2003, 2008).

To address these issues, this paper proposes a service oriented architecture for geospatial model sharing and integrating based on open standards. An interactive interface is developed based on OGC standards to reduce interoperability problems and semantic misunderstanding for geospatial model sharing. We propose a platform built to help modelers publish their models in a simplified and efficient way. Finally, we demonstrate an application of using OGC standards that minimizes interoperability and other problems using a common model sharing platform we developed for wetlands in the prairie pothole region of North America.

2 GEOSPATIAL MODEL SERVICE ARCHITECTURE

The Internet provides an effective means to connect and integrate diverse datasets, models and other resources located in different geographic locations to a diverse group of users through a shared modeling platform. Distributed geospatial model sharing is especially well suited to an Internet environment and it has a number of advantages over traditional stand-alone and closed network systems. Based on OGC geographic information service categories (Open GIS Consortium, 2003), herein we propose an architecture for developing open geospatial model sharing and integrating on the Internet.

The architecture we developed has four key components interconnected through interoperability elements (Fig. 1). Each component is a service or system capable of interacting with other components through the Internet. The interoperability elements are based on open standards and specifications (e.g., open data format, model interface, data metadata, and model metadata) that resolve the incompatibility issues associated with geospatial model sharing and integration.

2.1 Geospatial model services

Geospatial models collected and shared on the Internet as a network service are accessed by users through a standard interface. Users follow specific rules to access models through the interface which is application dependent and uses unique platforms or techniques for specific applications. Because geospatial models are usually computationally intense, model services

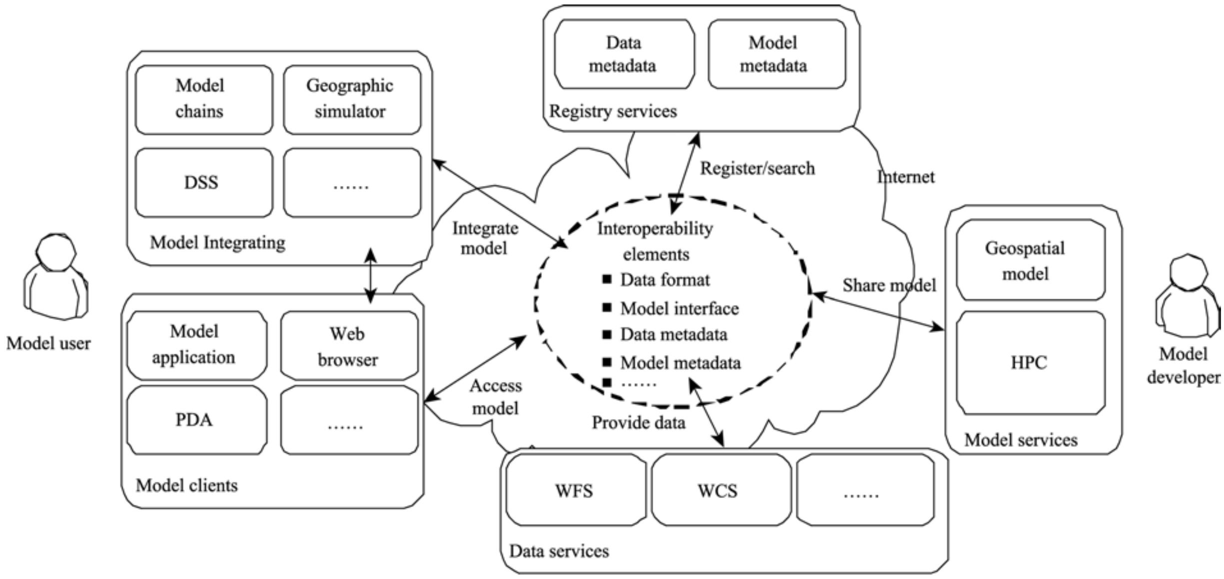

Fig. 1 Architecture of service-oriented geospatial models sharing

should be hosted on servers with high-performance computing (HPC) capability.

2.2 Geospatial data centers

When open geospatial modeling is implemented, each model or submodel functions like a "Black Box" that accepts and processes data, and then outputs results. Data are critical to effective use of shared models. Traditionally, input data are supplied by application users where output goes directly back to the user. However, users who lack access to high performance computers with large storage capacities often cannot process complex applications that require manipulation of large diverse data sets. However, the availability of large data centers is increasing on the Internet (e.g, GLCF [http://www.landcover.org], DayMet [http://www.daymet.org], and Geodata.cn [http://www.geodata.cn]). Professionally hosted data centers eliminate the need to store or transfer data to individual user computers when used through an open geospatial model sharing environment.

2.3 Model clients and integrating

Model sharing is a dynamic process that can generate output data which can be further used for subsequent analyses. Users can effect a cost savings by utilizing a model service (e.g., web browsers) remotely, making it possible to access geospatial models with portable devices. In addition, shared models can be integrated into model chains for use to solve highly sophisticated application problems such as the development of decision support systems (DSS).

2.4 Model registry services

The Internet is so vast that it is hard for users to find the exact model they need. Modelers have a similar problem because potential users of models are not widely known. Therefore, a mechanism is needed for modelers to register their models to maximize their use by modelers. Model registry services can be built into metadata database to provide model registering and discovery capabilities.

The geospatial model sharing architecture we suggest has three advantages. First, sharing geospatial models on the Internet allows users to share models and applications from anywhere in the world. Second, Internet based models have an almost limitless access to resources (e.g., geospatial data, other models, computing environments) that cannot be duplicated in standalone or networked systems. Third, compliance with open standards and specifications for data exchange and model interfaces ensure interoperability of computational models, including those built on different systems using very different tools.

3 MODEL SHARING SERVICE

Model module and model process are two fundamental concepts to access shared models. Model modules are executable programs that implement model algorithms and serve as templates for model processes. In contrast, model process is the actual computing process, an explicit function of the model module. A shared geospatial computational model should go through four phases, i.e., geospatial model module, geospatial model process, geospatial model service, and model client (as shown in Fig. 2). Each geospatial computational model is a model module that has the capability to provide geospatial model processes. To share those models, the model processes will be made accessible to model clients. As a consequence, this will simultaneously resolve communication and interoperability problems.

3.1 Model service interface

To avoid interoperability problems, model service and model client must adopt the same interactive rules. OGC standards were designed to serve a diversity of users representing

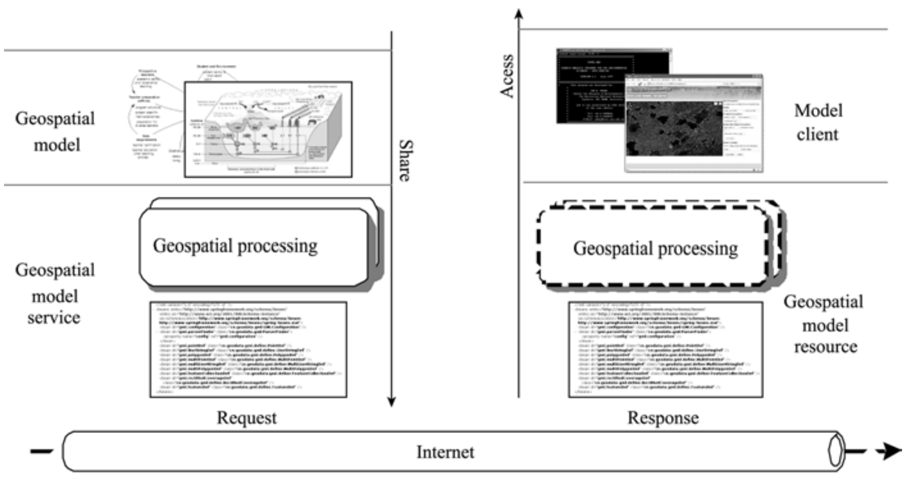

Fig. 2 Access shared geospatial model

different countries and groups, including the general public. In last decade, many OGC standards have been wildly accepted for geospatial applications, especially by open source communities.

WPS define basic operations and data structures for geospatial process based on the Extensible Markup Language (XML). Those operations and data structures do not identify specific processes, and are very generic and abstract. However, WPS allows the development of profiles that comply with basic operations and standardized data structures to be developed for specific uses (Open GIS Consortium, 2008). We developed a specific geospatial model sharing interface based on WPS for geospatial computational model that includes:

3.1.1 Model service metadata

Metadata is essential for sharing and in the discovery of shared models. Metadata of geospatial model service includes two levels information.

(1) Model service metadata provides a general description of the model service (e.g., service identifiers, contactor information, allowed operations) and a list of geospatial processes provided by the service. Hence, the metadata would provide model clients with an overview of the service and convey how user access shared processes.

(2) Geospatial process metadata provides detailed information on specified processes, including identifiers, titles, descriptions, controlling parameters, input and output parameters, and optimized options. Geospatial process metadata provides model clients with information on how specified processes interact, what parameters are required, and what outputs are possible.

3.1.2 Model service interacting

Model client interacts with model service through three predefined operations:

(1) GetCapabilities, retrieves metadata of model services;

(2) DescribeProcess, retrieves metadata of a given process;

(3) Execute, calls the operation, provides model client with input parameters, executes the specified process, checks executing status, and generates results.

When Execute is selected, model client implements an interactive session that invokes data exchange. Model service and model client are automatically programs that utilize Remote Procedure Call (RPC) data exchange, a standardized data format and semantic expressions.

We used Geography Markup Language (GML) 3 data format for exchanging geospatial data in our application. GML 3, the latest geospatial data exchanging standard from OGC, is independent of operating system (OS), and supports many geospatial data models including vector and raster. Model service can also use data from distributed data services supported by GML, such as web feature service (WFS) and web coverage service (WCS) services (Open GIS Consortium, 2004).

Based on GML, we adopted several methods that avoid or reduce semantic misunderstanding of geospatial data. Those methods are:

(1) We adopted an open geodetics parameter set compiled and disseminated by the European Petroleum Survey Group (EPSG) to coordinate parameters for geospatial data. EPSG compiled the most commonly used projections to build their database. Users simply refer the coordinates for specific sites of interest to EPSG identifier codes rather than providing all their explicit coordinate parameters. The EPSG identifier is succinct and explicit, and helps reduce geodetic incompatibilities between model service and model client.

(2) Metadata references (usually a Uniform Resource Locator (URL) reference to access the full metadata) are embedded in GML datasets and do not change dataset structure. Further, metadata references provide valuable opportunities for data mining (Feng *et al*., 2007).

(3) Data measurements can be categorized as Nominal, Ordinal, Interval, and Ratio (O'sullivan, 2003). Interval and Ratio data should have explicit measurement units to avoid misunderstanding. However, some measurement units are inheritantly problematic and converting them from one measurement unit to another is inefficient. To avoid this problem, the model service requires that all parameters use International Units, especially for Interval and Ratio type parameters.

3.2 Services communicating

Although model service interface is not constrained to a specific Internet communication technique, we used Web Services and Representational State Transfer (REST) to facilitate the geospatial model sharing service. Web Services and REST are open and widely supported by current systems and tools. These two techniques are complimentary; Web Services are best suited for heavy applications, such as desktop or server based model integrating applications while REST is more suitable for light applications, such as browser based applications.

4 MODEL SHARING PLATFORM

Publishing and serving scientific models on the internet involves many issues including software engineering, service interfaces, network communication, security and others beyond actual model development. Further, it is necessary to develop platforms that help modelers develop and share their models in compliance with model interfaces. Carefully designed platforms will allow modelers to focus on specific model designs and their implementation rather than individually working to ensure sharing and updating of geospatial models.

We developed a geospatial model sharing platform using Java2 Platform Enterprise Edition (J2EE). Java has enhanced network capabilities and numerous libraries available to support the platform we developed. Additionally, open source libraries (e.g., GeoTools, GeoServer, and OpenLayers) can be used to empower the geospatial related features (e.g., geospatial data reading and writing, spatial data checking) of the platform. The platform is OS independent and can be accessed from any OS that supports Java Virtual Machine (e.g., Windows desktops, Linux servers).

The platform we designed follows the Model-View-Controller (MVC) pattern (as shown in Fig. 3), and includes three modules:

(1) Model Integrating Module. This module loads processes from computational models developed by modelers that can be shared for collecting metadata from other models and for interacting with computational models. All computational models are plugged in the platform through a programming level interface, called Model Integrating Interface, rather than through the Model Service Interface. The Model Integrating Interface is easy to use because it only involves programming and involves no network operations.

(2) Model Service Module. This module provides model services using Web Services and REST techniques. This module accepts operations defined by the model service interface, which is independent of implementation and accepts programming operations to translate required conversions.

(3) Controller Module. This module dynamically monitors and manages all model processes to ensure efficient host server resources (e.g., computational resources, memory resources, storage resources). High-performance computing techniques can be introduced into model simulation through this module. The module also provides utility functions, such as translating data back and forth to open standard formats.

The Model Integrating Interface is a set of Java Interfaces and Annotations based on GeoAPI designs, a Java based open library for geospatial related operations. This platform requires all computational models to implement the Model Integrating Interface in one of two ways:

(1) Direct Implementation. Using this approach, the computational model implements the interface directly, and runs in the same application environment as the platform to improve performance. However, models have to be modified when using Direct Implementation, and it can be difficult without model-specific source codes.

(2) Agent Implementation. In this approach, a model agent is developed to implement the interface between the computational model and its platform. Agent Implementation avoids modifications to computational models and can be very useful for sharing models that are hard to modify. However, the model and the platform have to run in separate application environments, potentially causing instability issues.

5 APPLICATION

The prairie pothole region (PPR) is an area where mid-continental climate variations interact with glacial geology to produce one of the most productive ecosystems in North

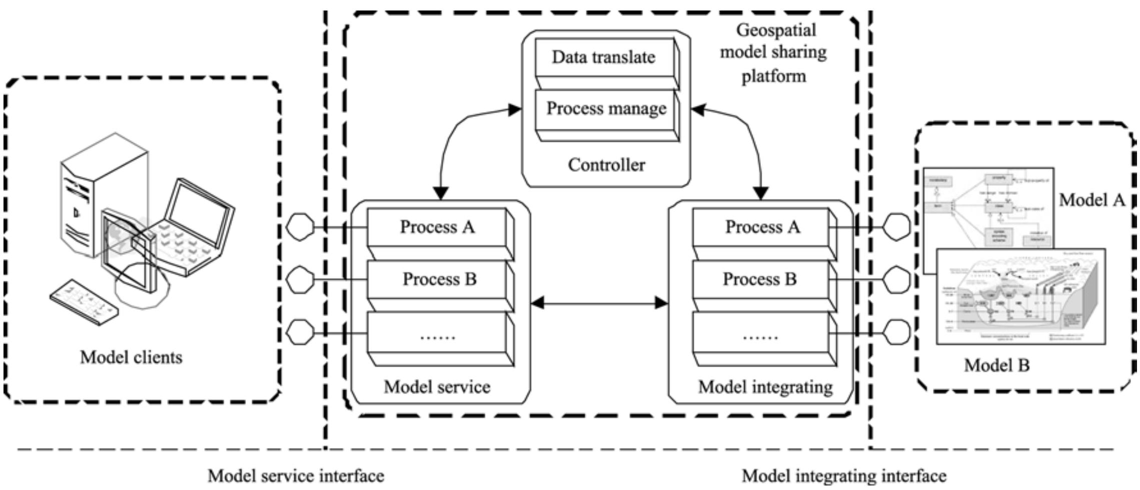

Fig. 3 System structure of geospatial model sharing platform

America, both of agricultural crops and of wildlife. The PPR stretches from Alberta, Saskatchewan, and Manitoba in Canada to Montana, North Dakota, South Dakota, Nebraska, Minnesota, and Iowa in the United States. The PPR is approximately 900000km^2(Mann 1986, Phospahala *et al*., 1974) and may have contained over 20 million ha of wetlands prior to European settlement (Millar, 1973; Tiner, 1984). Soils in the PPR are fertile and the area has been extensively developed for agriculture. Consequently, over 50% of the wetland area in the PPR of the United States (Tiner, 1984) and 71% in Canada (Environment Canada, 1986) have been drained for agricultural development. Prairie wetlands also are of considerable ecological value and support more than 50% of North American migratory waterfowl and they provide numerous other ecosystem services (Gleason *et al*., 2008) such as climate change mitigation and water storage. Because competing land use has highly modified this landscape, we choose the area to demonstrate an application of shared open geospatial models to simulate hydrological and ecological change.

Collaborative research between the Chinese Academy of Sciences and the U.S. Geological Survey's Center for Earth Resources Observation and Science (EROS) and the Northern Prairie Wildlife Research Center has developed several scientific computing models that were published as a model service. The model services we developed are accessible using WPS compliant tools or systems. We used Java to write this model application and implemented it directly using a Model Integrating Interface deployed on a geospatial model sharing platform. Three scientific computing models were shared (i.e. a wetland water table model, a catchment water surface extent model, and an evapotranspiration [ET] model) and the model service gets data from several data services. For example, the model service fetches meteorology data dynamically from DayMet (http://www.daymet.org), which is developed by the U.S. National Center for Atmospheric Research to provide simulated meteorological data for the United States from 1980 to 2003; moreover, the catchment data are fetched from a WFS service we developed for this application.

We developed a user-friendly website, based on distributed model service that integrates model service and WebGIS technologies. Users can simulate water pool depth changes for wetlands of interest and can specify time periods of interest (e.g., day, month, and year) and display the results on an interactive map (as shown in Fig. 4). Users also have the option of downloading data from the application for additional analyses or modeling applications.

6 CONCLUSION

To meet the scientific challenges of the coming century,

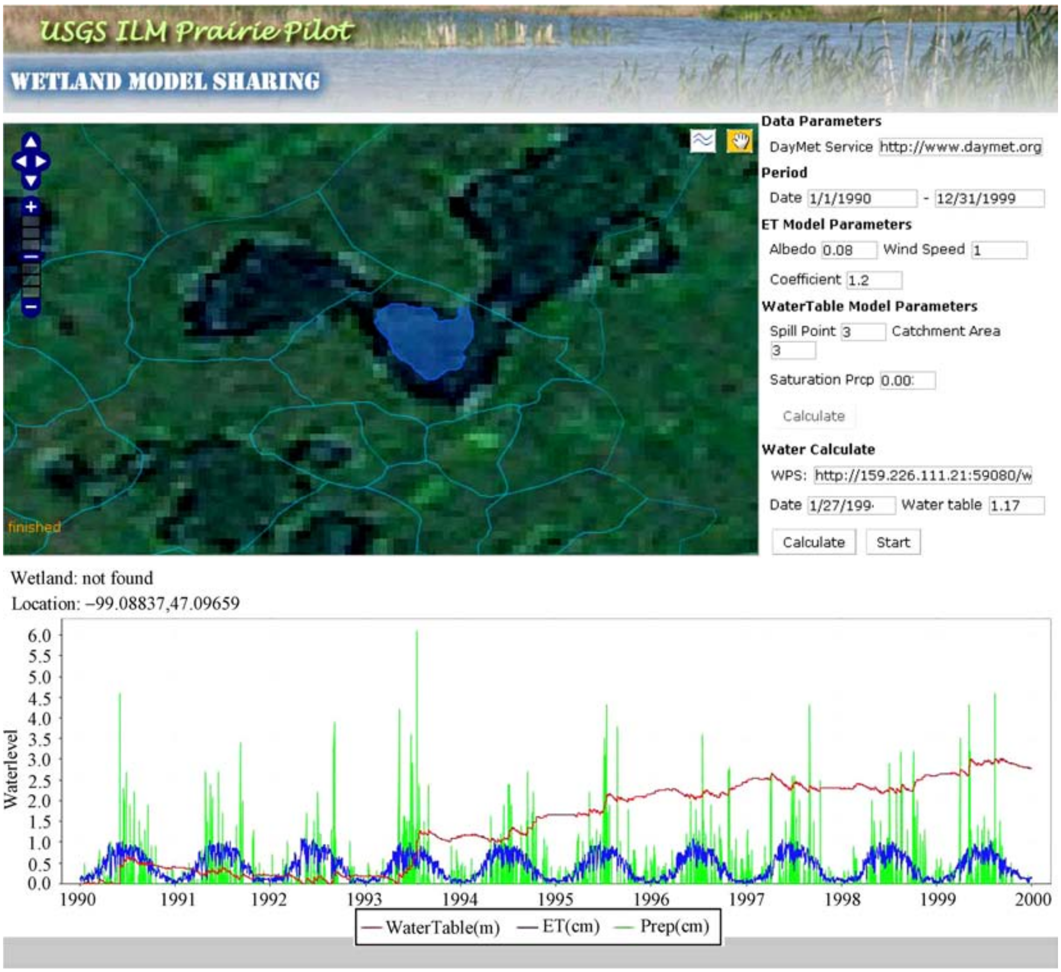

Fig. 4 Water table simulation showed on the model service integrating website.

geosciences will need to address interdisciplinary problems and applications that span national boundaries over wide geographic regions. The capacity of geosciences to handle these future challenges is increasingly obvious as we depend upon more complex workflows for data analysis and simulation tasks (Crosier *et al*., 2003). Sharing and integrating scientific resources over the Internet will be an important approach to meet these future challenges. We have identified methods for sharing geospatial computational models using SOA techniques and open geospatial standards that have far more advantages and utility than traditional stand-alone and closed network model sharing systems. Using service-oriented model sharing not only helps modelers share their models but also makes it possible to capitalize on abundant scientific resources available worldwide to solve more sophisticated application problems. Model sharing also allows model services to be empowered by diverse computational devices and technologies, such as portable devices and GRID computing infrastructures.

Models share computational functionality through model services, so computational models can be shared as model processes using model services. Based on the generic and abstract operations and data structures defined by WPS, we propose an interactive interface for sharing model services. We discussed the role of model metadata, interactive operations and data references issues, and how rules can be added to the interface of geospatial computational models. Although the interface can be designed to reduce interoperability problems between model services and clients, further research is needed to identify model scale, parameter estimation, and model limitations (Leavesley *et al*., 2003).

We developed a geospatial model sharing platform based on J2EE and open-source geospatial libraries. The platform resolves interoperability problems associated with traditional modeling approaches, improving efficiency and allowing modelers more time to focus on model design and implementation. We used a set of hydrological models for prairie pothole wetlands to illustrate and validate the method. The models we developed and shared on this platform, allows users to access model clients though tools compliant with WPS. We also developed a user-friend website for this application as an example for use to solve geospatial simulation problems using an integrating model service.

The theory, specification and technology of distributed data sharing have made much progress in recent years (Zhu *et al*., 2006). However, there has been little progress made in distributed geospatial model sharing despite the enormous potential to advance geosciences. Certainly, achievements on geospatial data sharing and distributed computing technologies (e.g., Web Services, SOA, Grid computing, and Cloud computing) will stimulate further research to improve sharing geospatial models.

REFERENCES

Crosier S J, Goodchild M F, Hill L L and Smith T R. 2003. Developing an infrastructure for sharing environmental models. *Environment and Planning B*: *Planning and Design*, **30**(4): 487—501

David J M. 2005. Towards a GIS platform for spatial analysis and modeling. GIS, Spatial Analysis, and Modeling. Redlands, CA: ESRI Press

Environment Canada. 1986. Wetlands in Canada: a valuable resource. Fact Sheet, Ottawa, Ontario, Lands Directorate

Feng M, Zhu Y Q, Wang J L, Liu R D and Song J. 2007. Research and implement of distributed multi-standard geo-metadata sharing. *Geography and Geo-Information Science*, **23**(6): 8—13

Gleason, R A, M K Laubhan and N H Euliss, Jr.. 2008. Ecosystem services derived from wetland conservation practices in the United States prairie pothole region with an emphasis on the U. S. Department of Agriculture Conservation Reserve and Wetlands Reserve Programs. Professional Paper 1745, U. S. Geological Survey, Reston, Virginia

Goodchild M F. 2003. GIS and Modeling. Workshop on GIS and Modeling, Environmental Systems Research Institute, Redlands, CA

Goodchild M F. 2005. GIS and modeling overview. GIS, Spatial Analysis, And Modeling. Redlands, CA.: ESRI Press

Granell C, Díaz L, Gould M, Pascual V and Guimet J. 2007. Developing geoprocessing services for a hydrological model application. 27th EARSeL Symposium

Leavesley G H, Restrepo P J and Stannard L G. 1996. MMS: A modeling framework for multidisciplinary research and operational applications. GIS and Environmental Modeling: Progress and Research Issues. John Wiley and Sons, Hoboken

Liu J, Peng C, Dang Q, Apps M and Jiang H. 2002. A component object model strategy for reusing ecosystem models. *Computers and Electronics in Agriculture*, **35**(1): 17—33

Mann, L K. 1986. Changes in soil carbon after cultivation. *Soil Science*, **42**: 279—281

Millar, J B. 1973. Vegetation change in shallow marsh wetlands under improving moisture conditions. *Canadian Journal of Botany*, **51**: 1443—1457

Morozov I, Reilkoff B and Chubak G. 2006. A generalized web service model for geophysical data processing and modeling. *Computer* & *Geosciences*, **32**: 1403—1410

Open GIS Consortium. 2003a. OGC Reference Model. Open GIS Consortium

Open GIS Consortium. 2003b. OpenGIS Web Services Architecture. Open GIS Consortium

Open GIS Consortium. 2004. Geographic information Geography Markup Language (GML) 3.1. Open GIS Consortium

Open GIS Consortium. 2008. Web Processing Service (WPS) Specification. Open GIS Consortium

O'sullivan D. 2003. Geographic Information Analysis. John Wiley & Sons, Inc. Hoboken, New Jersey

Pospahala R S, Anderson D R and Henry C J. 1974. Population Ecology of the Mallard. Resources Publication: U. S. Fish and Wildlife Service, Washington, DC

Reed C. 2005. Geospatial Paradigm Shift or Not [EB/OL]. http://www.opengeospatial.org/press/?page=newsletter&year=0&newsletter=89

Tiner R W, Jr. 1984. Wetlands of the United States: current status and recent trends. U. S. Fish and Wildlife Service. U. S. Government Printing Office, Washington, DC.

Wei Y C and Cheng S Z. 2005. Principle and Methods of Geographic Modeling. Beijing: Science Press

Zhu Y and Sun J. 2006. The research progress in geo-data sharing based on e-GeoScience. *Advances in Earth Science*, **21**(3): 286—291

基于开放互操作标准的分布式地理空间模型共享研究

冯　敏[1,2], LIU Shu-guang[3], Ned H. Euliss, Jr.[4], 尹　芳[1]

1. 中国科学院 地理科学与资源研究所，北京　100101;
2. University of Maryland Institute for Advanced Computer Studies, College Park, MD, 20742, USA;
3. U.S. Geological Survey, Earth Resources Observation and Science Center, Sioux Falls, SD, 57198, USA;
4. U. S. Geological Survey, Northern Prairie Wildlife Research Center, 8711 37th St SE, Jamestown, ND, 58401, USA

摘　要： 传统的单机环境和封闭式网络环境由于有限的资源利用能力，难以充分支持分散地学数据、模型等资源的共享与应用集成。基于网络环境的信息交换特点，提出了分布式地理空间模型共享的服务体系。该体系以数据、模型、元数据等互操作要素为核心，通过网络将数据、模型等网络节点进行开放式耦合。针对地理空间模型服务的互操作问题，提出了分布式环境下的模型共享服务交互接口，该接口定义了模型服务元数据、模型服务的交互操作、模型服务的通讯方式等交互规则，尽可能地降低模型服务与模型终端之间在数据交换、功能调用等方面的互操作困难。为了降低将模型共享为模型服务的实现难度，设计和开发了地理空间模型共享平台，并介绍了在该平台上发布地理空间模型的 2 种方法。最后介绍了研究成果在 Prairie 生态模型共享方面的应用实践。

关键词： 地理空间模型，模型共享，分布式计算，互操作

中图分类号： P208　　　**文献标识码：** A

1　引　言

地理空间模型是在 GIS(Geographic Information System，地理信息系统)的支持下，采用 GIS 操作来模拟现实世界在一个时刻或一个时间段的地理过程(Goodchild, 2005)。地理空间模型是生态、资源环境等领域科学研究成果的主要形式之一，在未来预测、现象理解、异常诊断、综合分析以及管理决策等方面均是不可或缺的(韦玉春 & 陈锁忠，2005)。随着科学研究的开展，在科学家手中积累了大量地理空间模型，这些模型是经验和知识的具体实现，主要以计算机程序的形式存在。对这些分散在科学家手中的地理空间模型进行共享的意义在于：首先，使这些模型所承载的研究成果得以更充分地发挥，也使其他研究和应用能够利用这些模型的潜在价值；其次，科学计算模型一般有明确的目的和应用边界，但无论在开展科学研究还是解决实际应用问题时，难免要涉及跨学科、跨领域问题，因此模型的共享和集成是深入开展科学研究和应用实践的需要(Liu 等, 2002)。第三，在社会应用尤其是社会经济辅助决策方面需要一些具有科学计算和模拟能力的科学计算模型，共享的地理空间模型为这些应用的开展提供了强有力的支撑；第四，科学数据共享是一种静态的信息共享，而科学计算模型的共享则是动态的信息共享(Crosier & Goodchild, 2003)。与静态的科学数据共享方式相比，动态的科学计算模型能够在使用过程中不断产生和提供信息，其信息量和价值也会随时间和使用过程不断增加。

从 20 世纪末开始，面对已有的地理空间模型，许多组织或研究者尝试采用更有效的模型开发方法，以降低模型重用的难度，提高新模型开发和模型集成的效率，一些与地理空间模型共享和集成相关的研究和项目也相继开展。这方面的研究有：基于单机环境的模型集成系统，如美国地质调查局开发的 MMS(Modular Modeling System)(Leavesley 等, 1996)。基于分布式网络技术实现地理空间模型共享系统，如：Morozov 等(2006)开发了地震数据分析与建模系统。Granell 等(2007)尝试实现融雪水文和环境模型的在线共享，并用于阿尔卑斯山地区的径流预测。

对上述研究总结后发现：在地理空间模型的共享和集成方面，还存在一些突出的问题：第一，很多研究基于单机或小网络环境开展模型的共享和集

收稿日期：2008-07-21；修订日期：2009-05-19

基金项目：国家科技基础条件平台项目(编号：2005DKA32300)和国家 863 项目(编号：2006AA01A120)。

第一作者简介：冯敏(1981—　)，男，助理研究员，博士，主要从事地理信息服务和地学模型共享方面的研究工作。E-mail: fengm@lreis.ac.cn。

成，虽然在短期内和部分应用中取得了较好的成果，但由于相对封闭和有限的资源利用能力，难以适应复杂多样的应用需要；第二，缺少一致的数据交换和模型调用的方式，使互操作问题成为限制广泛应用的重要因素；第三，直接对模型体或源代码共享，易于造成模型著作权和所有权等问题。

随着 GIS 理论、技术的发展和应用实践的深入，GIS 在体系上不断发生着变化，经历了从单层体系(Single-Tier)、3 层体系(3-Tiers)(包括单机 3 层体系和网络 3 层体系)到 SOA(Service Oriented Architecture，面向服务体系)的体系发展阶段。与其他体系相比较，SOA 更能够发挥网络分布式环境的优势，是 GIS 的重要发展方向(David，2005)。基于 SOA 的地理空间模型共享，一方面是推进基于 SOA 的 GIS 发展的重要支撑，另一方面，模型共享服务是建立在该体系其他部分的基础上，需要与数据服务、信息发现服务等紧密协作。

过去 10 年中，在分布式地理空间数据交换的理论方法和标准规范方面有了显著的进步，制定了一系列地理空间数据交换和数据服务的标准规范，并得到了广泛的接受和认可，如：GML(Geography Markup Language，地理标记语言)、WMS(Web Map Service, Web 地图服务)、WFS(Web Feature Services, Web 地理要素服务)、(Web Coverage Service，网络 Coverage 服务)、KML(Keyhole Markup Language, Keyhole 标记语言)等数据互操作规范，以及 ISO 19115 等元数据标准(Granell 等, 2007)。但是，与地理空间数据方面的进展相比较，开放的分布式地理空间模型功能互操作仍处于探索阶段。

2 地理空间模型共享服务体系

在分布式网络环境中开展地理空间模型的共享，需要实现地理空间模型共享基本体系各基本组成的网络服务化。根据地理空间模型的特点，参考 OGC(Open GIS Consortium，开放 GIS 协会)对地理信息服务的分类(Open GIS Consortium，2003a)，提出了地理空间模型分布式共享服务体系(图 1)。

网络环境下的数据、功能、元数据互操作是分布式模型共享的基础，为在网络环境中开展地理空间模型共享、发现和使用，提供必要的信息编码和交互规则约定。通过采用开放的规范和标准，降低或避免模型共享和使用过程中的互操作问题，具体措施包括：采用第 GML 第 3 版本作为地理空间数据交换格式，采用 ISO(International Organization for Standardization，国际标准化组织) 19115/19139 作为地理空间数据元数据标准，基于 OGC 的 WPS(Web Processing Services, Web 处理服务)设计地理空间模型服务的交互接口和模型元数据标准(Open GIS Consortium, 2008)。

以“互操作要素”为中心，在网络环境中将模型、数据等 5 种网络节点联系起来，构成分布式的模型共享体系。这些节点包括：(1)模型服务。以高性能计算的软硬件环境为基础，构建模型共享环境，将模型提供者提供的科学计算模型发布为模型共享服务。(2)数据服务。地理空间数据的网络服务，如：WFS、WCS 以及基于网络传输协议的地理空间数据流等。(3)注册服务。负责维护服务信息库，并接收服务的注册和对服务的查询，包括:“数据服务信息”

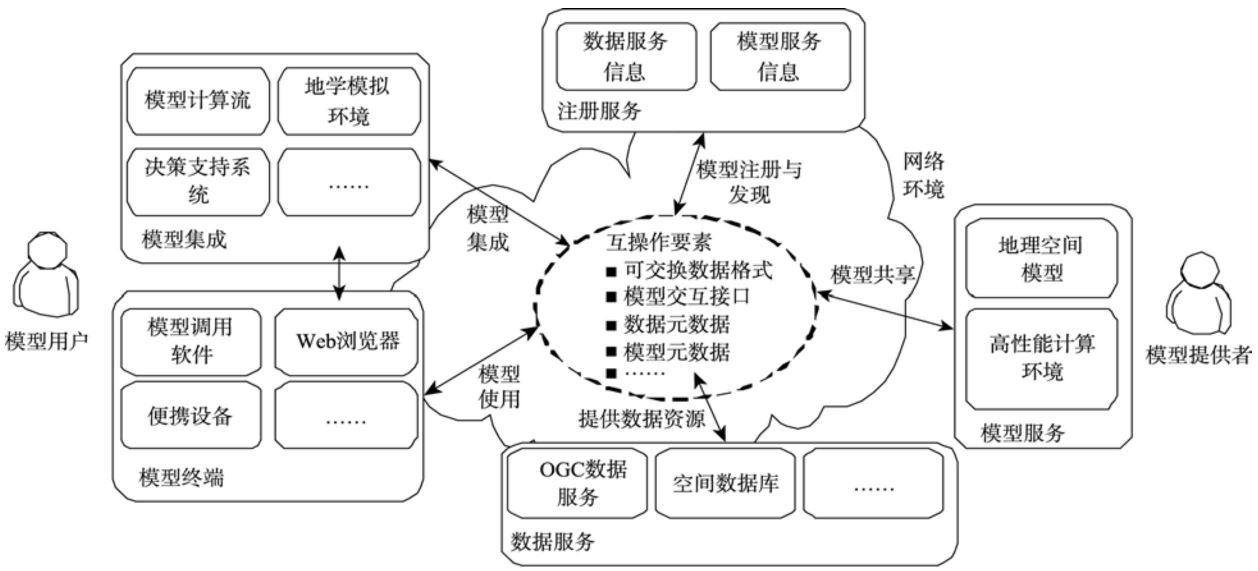

图 1 分布式地理空间模型共享服务体系

和“模型服务信息”的注册与发现。(4)模型终端。用户通过支持模型服务交互接口操作的软件或工具访问模型服务。(5)模型集成。面向科研和社会应用需求，通过网络集成数据、模型等资源，构建模型的应用系统，如：决策支持系统、模型计算流、地学模拟环境等。

该服务体系以分布式网络服务形式支持数据、模型等资源的共享与应用集成，其特点在于：首先，以网络服务形式实现地理空间计算模型的共享，有利于在更广的范围内开展模型应用和集成；其次，有利于在共享模型的访问过程中利用其他网络资源，包括地理空间数据资源、软件资源、高性能计算环境等；第三，通过引入开放的地理空间数据服务和数据编码，在模型访问过程中降低数据交换和模型访问过程中的互操作问题。

3　模型共享服务

地理空间模型体模型使用，需要经历 4 个阶段(图 2)，地理空间模型、地理空间模型服务、地理空间模型资源和模型终端。对于开放分布式地理空间模型共享，最关键的问题在于：如何通过网络，使模型服务转变为模型终端可用的地理空间模型服务资源。在这个环节，不仅需要解决底层网络通讯问题，而且要解决地理空间模型所涉及的数据、元数据等信息的交换以及模型计算功能的交互规则等问题，即模型服务的互操作问题。

3.1　模型服务交互接口

标准化已经成为影响当前 GIS 发展的重要因素，也是 GIS 工业化的基本要求(Reed, 2005)。解决互操作问题需要以交互双方均能够接受的交互规范为基础。交互规范定义了交互双方需要遵守的规则，因此必须能够被通信双方共同接受，需要具有一定的权威性和开放性。本文基于 OWS(OpenGIS Web Services Architecture, OpenGIS Web 服务体系)和 WPS 规范(Open GIS Consortium, 2003b, 2008)，提出了开放的分布式地理空间模型服务互操作接口。交互接口的主要内容如下。

3.1.1　模型服务元数据

地理空间模型元数据包括 2 个层次的元数据信息：(1)模型服务的元数据。对模型共享服务的宏观描述，主要包括服务器的标识信息、联系人信息、交互操作描述和地理空间处理简要信息列表等。通过模型服务元数据，模型终端可以了解模型服务的基本情况以及允许的交互方式；(2)地理空间处理的元数据。对模型服务中共享的地理空间处理的描述，包括地理空间处理的标识、名称、简要描述、控制参数描述、输入参数列表及描述、输出参数列表及描述、执行优化选项等。通过该元数据，模型终端可以了解特定地理空间处理的基本信息以及访问模型处理的方式。

3.1.2　模型服务的交互操作

地理空间模型服务的所有交互通过 3 个操作完成：(1)GetCapabilities。模型终端通过该操作获取模型服务的元数据；(2)DescribeProcess。模型终端通过该操作获取指定地理空间处理的元数据；(3)Execute。通过该操作，模型终端传递输入参数、调用模型、获取运行状态和获取计算结果。

数据交换是模型交互过程中的重要内容。模型终端与地理空间模型服务的交互属于 RPC(Remote procedure call，远程过程调用)过程，要求交互双方在数据表达层和数据语义层均能够一致。在模型服

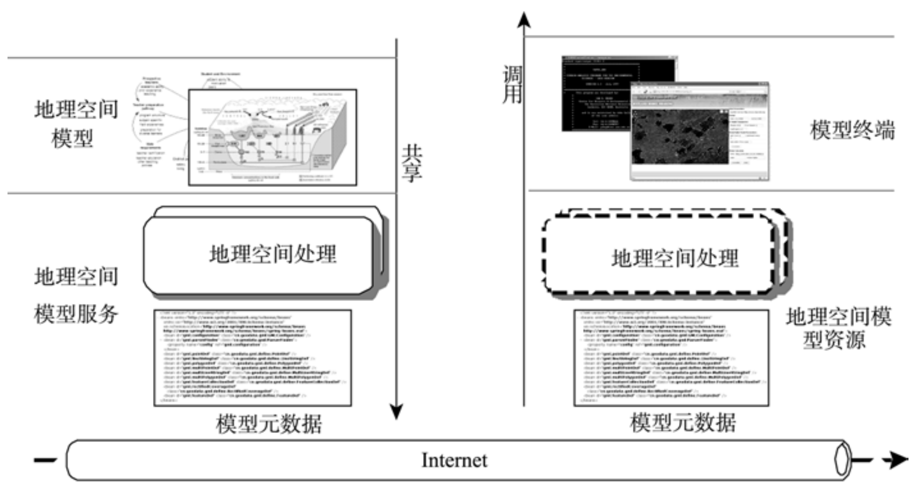

图 2　地理空间模型访问的过程

务接口中采用 GML 作为数据交换格式，以便利用GML 的平台独立性、可扩展性、体系完整性、易于网络交换等特征(Open GIS Consortium, 2004)。在数据语义方面，接口对地理空间坐标参考、地理空间数据元数据、数值单位等关键点上进行了规范，尽可能降低模型交互双方在地理空间数据交换过程中的语义差异。

(1) 在对目前众多地理空间坐标参考表达方式进行分析基础上，提出采用“欧洲石油调查组织”(European Petroleum Survey Group, EPSG)定义的坐标参考编码规范化模型双方的空间坐标参考信息的表达方式。“代码表示”方式避免了对空间参考系统整体传输，减少了网络数据传输，也避免了对同一投影的不同定义问题。

(2) 在数据的交换格式中添加外部数据元数据引用信息，避免在数据体内嵌入元数据体对数据内容和结构造成的影响，而且数据引用方式是一种“弱”联系，模型发现过程可以在这种“弱”联系的基础上设计“数据挖掘”等算法(冯敏等, 2007)。

(3) Stevens 在 1946 年提出了 4 级数据测量级别分类(O’sullivan, 2003)，即：标称(nominal)、序数(ordinal)、区间(interval)和比率(ratio)，地理空间模型共享中涉及的测量单位问题，主要针对“区间”和“比率”类型数据。在接口交互中，通过统一采用“国际单位”，避免由于数值单位偏差造成数据理解的偏差。

以 WFS、WCS 等形式提供的地理空间数据服务是一种特殊的数据源，一般由拥有数据资源和服务能力的节点单独维护，能够提供较为丰富的数据资源和稳定的性能。在分布式模型调用过程中结合远程数据服务，符合分布式地理空间模型共享服务体系的设计，有利于利用数据服务的丰富数据资源和优化模型的访问策略，如通过将数据服务作为模型输入数据源，可以减轻模型终端的数据传输压力，尤其对于轻量级模型终端具有重要意义，如移动终端等。

3.2　模型服务的通讯方式

模型服务元数据和交互过程的定义均独立于具体的分布式通讯方式，因此对底层的分布式通讯方式没有限制。本文采用 Web Services 和 REST (Representational State Transfer, 代表性状态传输)同时提供实现对地理空间模型服务的支持。Web Services 和 REST 是目前较广泛使用的分布式通讯技术，同时支持两种技术，有助于提高模型服务的适应能力，提高开放性，使模型终端拥有更多的选择。

4　模型发布与共享

将地理空间模型发布为地理空间模型共享服务是一个较为艰巨的过程，不仅需要实现模型服务接口要求的数据、操作等交互规则，而且涉及许多具体的软件工程难题。每个地理空间模型的共享需要重复一些与模型本身不相关的过程。如果能够将这些过程独立出来，使每个模型专注于模型的设计和开发，将大大简化地理空间模型共享难度，也便于实现模型的修改和更新。因此，实现能够支持地理空间模型共享的平台系统，对开展地理空间模型共享工作具有重要意义。

模型共享平台位于模型体与模型终端之间，基于网络技术，实现支持模型服务接口交互规则的网络服务，通过“模型集成接口”实现对模型的搜索、加载和交互过程。“模型集成接口”是在编程级别定义的交互规则，模型通过继承并实现这些接口，实现模型共享平台要求的交互动作，包括获取模型

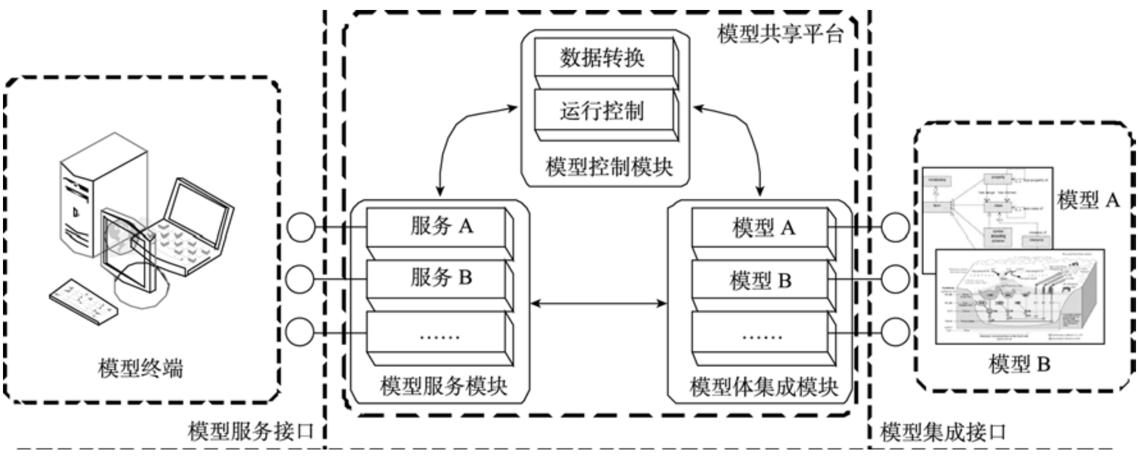

图 3　地理空间模型共享平台的系统结构

运行参数、交换输入/输出数据、启动和管理计算过程等(图 3)。

模型共享平台采用 J2EE 环境实现，一方面利用 Java 的网络系统开发能力和丰富的第三方功能库，另一方面能够实现跨平台部署和运行。

在模型共享平台的支持下，模型不需要实现独立的网络服务功能，仅需要按照模型共享平台的要求，通过模型集成接口发布至模型共享平台。从实现方式上，模型提供者可以通过 2 种方式发布模型为模型服务:

4.1 直接继承接口

模型体直接继承并实现模型集成接口。该方案是最直接有效的集成方式，并且由于运行环境一致，模型的可靠性和稳定性均较好。但对于已有模型，但该方案要求重新编写模型代码，因此必须首先获得模型的详细计算流程或者源代码。

4.2 模型代理继承

该方案是在不修改或少量修改模型体的前提下，通过开发模型的代理模块，实现对已有模型的封装。该方案的优点是避免了对已有模型体的修改，缺点是需要针对不同模型体的特点设计和实现代理模块。由于代理模块往往和模型体在不同的环境下运行，会造成模型运行性能的损失，并且带来潜在的不稳定因素。对该方法的使用，必须依赖于模型体的结构设计和实现方式。对于提供 API 的模型，可以在 API 的基础上设计代理模块，如对于 C/C++、Fortran 模型可以 DLL(Dynamic Link Library, 动态链接库)或静态库的形式提供 API，而 Java、C#等方式实现的模型可以提供 Java 或 C#对象形式的 API。对于不提供 API 的模型体，可以通过重定向 Console 程序的输入/输出流等方式实现模型封装模块与模型体之间交互。

5 应用实践

Prairie Pothole 地区是美国的中北部草原地区，在末次冰期结束时，随着冰川的退缩，在该地区留下了许多低洼地区，继而形成了大量分布的湿地区域。该地区包括加拿大 Alberta、Saskatchewan 和 Manitoba 等省，以及美国 Montana、North Dakota、South Dakota、Nebraska、Minnesota 和 Iowa 等州，面积大约 900000km^2(Mann, 1986; Phospahala 等, 1974)，在欧洲移民到来之前，曾拥有超过 2000 万 hm^2 的湿地(Millar, 1973; Tiner, 1984)。该区域的土壤肥沃，随着农业生产的扩张，大量湿地被排干成为农田，使湿地面积大量减少，其中美国部分 50%(Tiner, 1984)和加拿大部分 71%(Environment Canada, 1986)的湿地消失。该区域的湿地对北美 50%以上候鸟的迁徙具有重要意义，并具有其他生态系统服务功能，如减缓气候变化、储水等(Gleason 等, 2008)。由于该区域的土地利用变化的生态典型性，本文选择该区域作为模型共享的应用实例，示范该方法对水文和生态变化研究和应用的重要性。

本研究由中国科学院地理科学与资源研究所与美国地质调查局地球资源观测系统(Earth Resources Observation Systems, EROS)中心和北部 Prairie 野生生物研究中心(Northern Prairie Wildlife Research Center)科学家合作开展，收集了对该区域研究提出的一系列水文和生态模型，并发布为模型服务，实现模型的分布式调用和集成，为社会公众和政府决策提供平台。

在该应用中，通过采用“直接继承接口”方式，对 EROS 中心生态科学家编写的生态模型进行改造，使其继承模型集成接口，继而在模型共享平台上发布为模型服务。共享的模型包括: 水位模型、水域提取模型、土壤水分蒸发蒸腾损失总量计算模型。模型服务动态从 DayMet 数据源(由美国蒙塔纳大学在 DEM 和气象观测站的日最低和最高气温与降水的基础上，通过模型计算得到的全美国 1980—2003 年气象数据集，支持指定地理位置气象数据的在线查询)获取气象数据，从 WFS 数据服务获取流域几何数据。基于模型服务，构建了基于 Web 的模型操作界面(图 4)，该系统通过远程集成模型服务，向用户提供动态、快捷、可视化的水位模拟计算界面。

6 结　论

全球变化研究和地球系统科学为代表的综合性、跨学科研究的开展，逐渐对地理空间数据、模型等资源的共享和整合提出日益迫切需求。通过网络环境共享和整合地学数据、模型等资源，不仅能够避免单机环境的资源限制因素，支持更广和更深层次的地学资源共享和成果集成与应用，而且有助于促进网络化地球系统科学研究信息化环境(e-GeoScience)的发展(诸云强 & 孙九林, 2006)，是“数据—模型—知识”发展趋势中的重要一环。

基于此，探讨了分布式地理空间模型共享的服务体系，针对模型服务的互操作问题，提出了分布式环境下的模型共享服务交互接口，针对模型共享

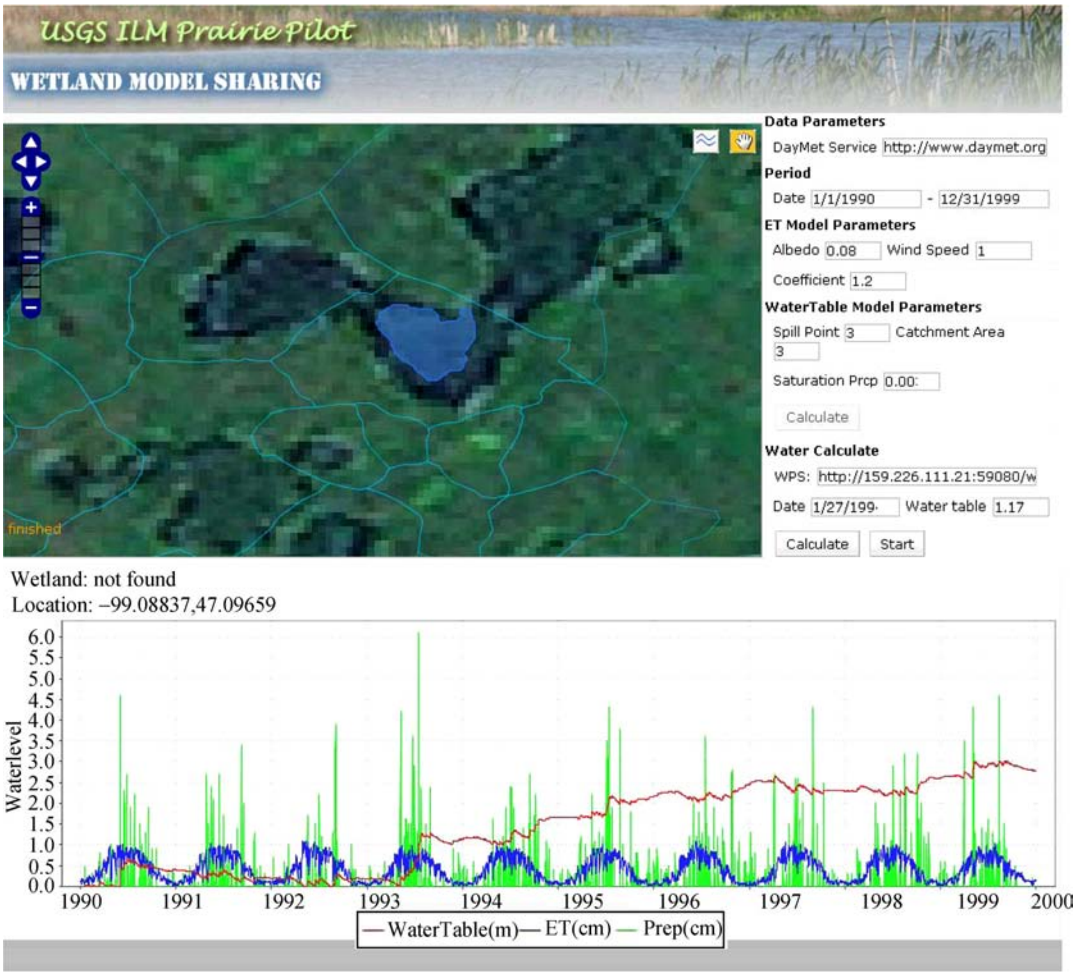

图 4　基于 Web 的 Prairie 模型动态交互页面

的工程困难，提出了模型共享平台和模型发布方法。通过对 Prairie 生态模型的共享和集成，初步实践了所提出的分布式模型共享思路。

目前，分布式数据共享从理论、规范、技术和应用等方面取得了丰富的成果，同时，Web Services、SOA、网格计算、云计算等分布式计算体系和技术也在过去 10 年中相继出现，并获得了长足发展，这一切为分布式地理空间模型共享的深入研究奠定了基础。

REFERENCES

Crosier S J, Goodchild M F, Hill L L and Smith T R. 2003. Developing an infrastructure for sharing environmental models. *Environment and Planning B: Planning and Design*, **30**(4): 487—501

David J M. 2005. Towards a GIS platform for spatial analysis and modeling. GIS, Spatial Analysis, and Modeling. Redlands, CA: ESRI Press

Environment Canada. 1986. Wetlands in Canada: a valuable resource. Fact Sheet, Ottawa, Ontario, Lands Directorate

Feng M, Zhu Y Q, Wang J L, Liu R D and Song J. 2007. Research and implement of distributed multi-standard geo-metadata sharing. *Geography and Geo-Information Science*, **23**(6): 8—13

Gleason, R A, M K Laubhan and N H Euliss, Jr.. 2008. Ecosystem services derived from wetland conservation practices in the United States prairie pothole region with an emphasis on the U. S. Department of Agriculture Conservation Reserve and Wetlands Reserve Programs. Professional Paper 1745, U. S. Geological Survey, Reston, Virginia

Goodchild M F. 2003. GIS and Modeling. Workshop on GIS and Modeling, Environmental Systems Research Institute, Redlands, CA

Goodchild M F. 2005. GIS and modeling overview. GIS, Spatial Analysis, And Modeling. Redlands, CA.: ESRI Press

Granell C, Díaz L, Gould M, Pascual V and Guimet J. 2007. Developing geoprocessing services for a hydrological model application. 27th EARSeL Symposium

Leavesley G H, Restrepo P J and Stannard L G. 1996. MMS: A modeling framework for multidisciplinary research and operational applications. GIS and Environmental Modeling: Progress and Research Issues. John Wiley and Sons, Hoboken

Liu J, Peng C, Dang Q, Apps M and Jiang H. 2002. A component object model strategy for reusing ecosystem models. *Computers and Electronics in Agriculture*, **35**(1): 17—33

Mann, L K. 1986. Changes in soil carbon after cultivation. *Soil Science*, **42**: 279—281

Millar, J B. 1973. Vegetation change in shallow marsh wetlands under improving moisture conditions. *Canadian Journal of Botany*, **51**: 1443—1457

Morozov I, Reilkoff B and Chubak G. 2006. A generalized web service model for geophysical data processing and modeling. *Computer* & *Geosciences*, **32**: 1403—1410

Open GIS Consortium. 2003a. OGC Reference Model. Open GIS Consortium

Open GIS Consortium. 2003b. OpenGIS Web Services Architecture. Open GIS Consortium

Open GIS Consortium. 2004. Geographic information Geography Markup Language (GML) 3.1. Open GIS Consortium

Open GIS Consortium. 2008. Web Processing Service (WPS) Specification. Open GIS Consortium

O'sullivan D. 2003. Geographic Information Analysis. John Wiley & Sons, Inc. Hoboken, New Jersey

Pospahala R S, Anderson D R and Henry C J. 1974. Population Ecology of the Mallard. Resources Publication: U. S. Fish and Wildlife Service, Washington, DC

Reed C. 2005. Geospatial Paradigm Shift or Not [EB/OL]. http://www.opengeospatial.org/press/?page=newsletter&year=0&newsletter=89

Tiner R W, Jr. 1984. Wetlands of the United States: current status and recent trends. U. S. Fish and Wildlife Service. U. S. Government Printing Office, Washington, DC.

Wei Y C and Chen S Z. 2005. Principle and Methods of Geographic Modeling. Beijing: Science Press

Zhu Y D and Sun J L. 2006. The research progress in geo-data sharing based on e-GeoScience. *Advances in Earth Science*, **21**(3): 286—291

附中文参考文献

冯敏，诸云强，王卷乐，刘润达，宋佳. 2007. 分布式多标准地学元数据共享的研究与实践. 地理与地理信息科学，**23**(6): 8—13

韦玉春，陈锁忠. 2005. 地理建模原理与方法. 北京:科学出版社

诸云强，孙九林. 2006. 面向 e-GeoScience 的地学数据共享研究进展. 地球科学进展, **21**(3): 286—291