A data description model for reusing, sharing and integrating geo-analysis models

Songshan Yue, Yongning Wen, Min Chen, Guonian Lu, Di Hu & Fu Zhang

Environmental Earth Sciences

ISSN 1866-6280
Volume 74
Number 10

Environ Earth Sci (2015) 74:7081-7099
DOI 10.1007/s12665-015-4270-5

Springer

Environ Earth Sci (2015) 74:7081–7099
DOI 10.1007/s12665-015-4270-5

THEMATIC ISSUE

A data description model for reusing, sharing and integrating geo-analysis models

Songshan Yue[1,2,3] · Yongning Wen[1,2,3] · Min Chen[1,2,3] · Guonian Lu[1,2,3] · Di Hu[1,2,3] · Fu Zhang[1,4]

Received: 25 September 2014 / Accepted: 20 February 2015 / Published online: 15 March 2015

Abstract Geo-analysis models are necessary tools for understanding various geo-processes and phenomena in Earth's environment. Studying, reusing, sharing and integrating geo-analysis models can help researchers solve complicated and synthetic geo-problems with interdisciplinary knowledge, especially for researchers who collaborate with each other to build virtual geographic environments (VGEs). While the integration frameworks of geo-analysis models could permit their practical use, it is essential for model users to prepare data according to the specific requirements of the different geo-analysis models. Model users should invest adequate effort and time into preparing such model data, particularly when employing multi-disciplinary geo-analysis models. This paper proposed a data description model, the Universal Data eXchange (UDX) model that can reduce the effort and difficulties of model data preparation and pre-processing for model users. With the UDX model, researchers from interdiscipline can build a collaborative workspace in VGEs more conveniently. A hierarchical structure was employed in the UDX model for the flexible description of heterogeneous model data, and a set of basic data node types was designed to provide a relatively stable organization method for the various data contents. In the UDX model, the structural format data (e.g., the *Shapefile* and *NetCDF* data) and the flexible plain text data content can be described in a uniform way. In addition, model data information can be completely and unambiguously described with the items in the attachment libraries (e.g., unit and dimension library, semantic library, spatial reference library, and data description template library). Furthermore, two different model integration case studies were conducted to prove that various data processing methods and efforts can be accumulated and organized with the designed UDX data processing library.

Keywords Geo-analysis models · VGE · Model integration · Data preparation · Data pre-processing

Introduction

Earth's environmental system is extremely complicated and constantly changing (Serreze 2011; Albanesi and Albanesi 2014; Rich et al. 2014). Geo-analysis models are widely used to describe geographic phenomena, simulate geo-processes, and depict regular geographic patterns in the environment (Davies Evan and Simonovic 2011; Sudo et al. 2013; Ramalho et al. 2013; Tung et al. 2013; Shao et al. 2014; Hudak 2014). Several geo-analysis models have been explored and developed in a range of disciplines to solve specific geo-problems (DeVantier and Feldman 1993; Goodchild et al. 1996; Basnyat et al. 2000; He et al. 2002; Dickinson et al. 2006; Todorova et al. 2010; Dennis et al. 2012; Sadat-Noori et al. 2014; Selvam et al. 2014;

✉ Yongning Wen
wenyn@msn.com

1 Key Laboratory of Virtual Geographic Environment (Nanjing Normal University), Ministry of Education, Nanjing 210023, China

2 State Key Laboratory Cultivation Base of Geographical Environment Evolution (Jiangsu Province), Nanjing 210023, China

3 Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China

4 Institute of Resources and Environment, North China University of Water Resources and Electric Power, Zhengzhou 450000, China

Uyan 2014). However, to solve synthetic environmental problems, any single study field is inadequate; thus, the collaboration of different fields of study is required (Guariso et al. 1996; Argent 2004a, b; Rahimi et al. 2014).

Virtual geographic environment (VGE) has been proposed as a new generation of geographic analysis tool to contribute to human understanding of the geographic world and assist in solving geographic problems at a deeper level (Lu 2011; Lin et al. 2013a; Chen et al. 2013a, b). Geo-data base and geographic process model base are two basic cores of VGE (Lin et al. 2013b). Studying, reusing, sharing and integrating geo-analysis models can provide an interactive interface for researchers performing geographical experiments collaboratively in VGE (Lin et al. 2010). In addition to this, model integration frameworks are commonly considered as an efficient way to conduct interdisciplinary modeling studies (Argent 2004a, b; Argent et al. 2006; Aydi et al. 2013; Yal and Akgün 2013). To integrate different geo-analysis models, the major difficulty lies in the heterogeneity of the geo-analysis models, which can be summarized by the examples in Table 1.

To target these heterogeneity difficulties, a range of frameworks or platforms for reusing, sharing and integrating geo-analysis models have been proposed and developed. For model integration in a single discipline, several discipline-related model integration frameworks have been developed to build specific synthetic models (Aminu et al. 2013; Bagdanaviciute and Valiunas 2013). Examples include the Gestion intégrée des Bassins Versants à l'aide d'un système Informatisé (GIBSI) (Mailhot et al. 1997; Rousseau et al. 2000) for watershed management modeling, the Open Modeling Interface (OpenMI) (Blind and Gregersen 2005; Moore and Tindall 2005) for hydrological modeling, ModCom (Hillyer et al. 2003) for the simulation of agro-ecological systems, and CAPRI (Britz et al. 2010) for the bio-economic modeling of agricultural systems. In contrast to these discipline-specific frameworks, there are various multi-disciplinary modeling frameworks that can integrate models over a wider domain, such as the Spatial Modeling Environment (SME) (Maxwell and Costanza 1997a, b), the Earth System Modeling Framework (ESMF) (Hill et al. 2004) and the European Union's Program for Integrated Earth System Modeling (PRISM) (Valcke et al. 2006). In addition, the Interactive Component Modeling System (ICMS) (Reed et al. 1999), the Dynamic Information Architecture System (DIAS) (Campbell and Hummel 1998; Simunich et al. 2002), the Modular Modeling System (MMS) (Leavesley et al. 1996, 2002), Tarsier (Watson and Rahman 2004), the Community Surface Dynamic Modeling System (CSDMS) (CSDMS Working Group 2004), the Object Modeling System (OMS) (Ahuja et al. 2005), the SEAMLESS-IF (Van Ittersum et al. 2008), and the Open Environment for Model Sharing in Cloud Platform (Wen et al. 2013) permit the integration of different geo-analysis models for complicated problem solving. However, to drive the model execution, the model data should be prepared and provided to these integration frameworks (single discipline frameworks or multi-disciplinary frameworks). The model data appear to have the greatest influence on the practicality of the results.

Table 1 Heterogeneities of geo-analysis models

Heterogeneities	Examples
Software platforms	Windows, Unix, Linux, etc.
Hardware platforms	Super computer, work station, personal computer, etc.
Programming languages	C, C++, Fortran, C#, Pascal, Java, Python, etc.
Distribution mode	COM, EXE, DLL, etc.
Input/output data format	Database, plain text, binary file, etc.
Function call interface	Graphical user interface (GUI), Command line interface, Application Programming interface (API), complex call interface, etc.

According to the above analysis, how to integrate two geo-analysis models is related to the model data closely. Various spatio-temporal scales, interpolation methods, data precision, data units, and other data discrepancies must be eliminated before integrating specific models. The features of model data can be summarized as follows: (1) connectivity with the disciplinary background, for example, hydro-specific codes are commonly used for models in hydrology (e.g., SWAT model and POM model) and land use codes are essential information in land use and land cover change (LUCC) studies; (2) complexity of the data structure, for example, for the same raster data, different data structures (e.g., column-row, row-column, and compression matrix methods) can be employed for the model data; and (3) diversity of the data format, such as fixed data formats (e.g., *Shapefile*, *NetCDF*, and *Geotiff*) and flexible plain text data (e.g., the SWAT model's HRU data). Hence, it is cumbersome work for modelers from various disciplines to use and integrate different models, particularly when using models from different study areas because the model data are closely related to disciplinary background knowledge.

Regarding the heterogeneity of model data, researchers in data management and modeling fields have offered solutions to decrease the difficulty of data interchange and conversion. Geographic information system (GIS) methods, functions and tools have been widely used in modeling fields to manage and operate model data. In GIS fields, establishing an intermediate data format (Bivand et al. 2013), drafting national or union standards (Arctur et al.

1998), providing data read and write interfaces (Pascoe and Penny 1990; Gray et al. 2005), and building semantic conversion channels (Lake 1999; OGC 2001, 2003) have provided solutions to data assimilation. Although the data interchange and data conversion solutions in GIS would somewhat solve the interoperation problems of spatial data, these methods cannot solve the entire problem of heterogeneous data in geo-analysis modeling (Parsons 2011). Most model data are composed of a plain text file, and the content structure is too flexible to be converted to a traditional GIS data format (Deshpande et al. 2004; Gunay et al. 2014). In addition, along with the introduction of data access methods, new structures and formats for model data have been developed and used in geo-analysis modeling (Conde et al. 2014). The GIS data interchange and conversion strategies cannot handle these complex circumstances.

Model data have various formats and can be stored in different modes to describe the research targets. The full descriptions of the information of the model data could also fully describe the model data. From this viewpoint, regarding military simulations, the Synthetic Environment Data Representation and Interchange Specification (SEDRIS, http://www.sedris.org/) project has employed data description strategies to simulate a synthetic battlefield environment (Bhatt et al. 2004). The SEDRIS project has proven that data description strategies can be used to convey heterogeneous data in different simulation workspaces. However, the SEDRIS project is more concerned with battlefield environment simulations; the strategy of information description is self-organized which can hardly be extended; and the designation of its data structure is also relatively too complicated to be conducted for modeling issues (Hwam et al. 2011). The data description strategy proposed in this paper calls attention to the transmission between values and information, which includes not only the conversion of simple numerical values but also the conveyance of scientific concepts, such as spatio-temporal information, measurement units, and semantic information. Based on this strategy for the description of model data, the model data can be expressed and formed in a uniform and unambiguous way; thus, the difficulty of model data preparation and pre-processing is reduced.

The remainder of the article is structured as follows. Section "Basic concept of the data description model" introduces the background regarding reusing, sharing and integrating geo-analysis models and analyzes the common features of model data. In addition, a flexible strategy for describing heterogeneous model data is introduced. In Sect. "Universal Data eXchange model", a UDX schema that fully and unambiguously describes model data is discussed. In addition, by using the proposed data process library, the model data preparation and pre-processing work with UDX are explained. Section "Case studies" introduces and proves the capability of UDX in practical model usage using two typical model integration cases. Finally, the conclusions and a discussion of the study are presented in Sect. "Conclusion and future work".

Basic concept of the data description model

According to the above analysis, it is practically impossible to find any single data format that organizes and stores all types of model data because of technical and interdisciplinary circumstances and the uncontrollable data structure designations of individual modelers (Voinov and Cerco 2010; Warner et al. 2008). Hence, this paper discusses a model data description strategy based on potential universal model data in terms of data structures, contents, and formats.

Although model data are dispersed and heterogeneous, all executable geo-analysis models should ultimately be implemented via programming. Many programming languages exist in the computer technology field, including those employed by geo-analysis models, such as Fortran, C/C++, Matlab editing language, Python, C#, Java, Visual Basic, and R language. Despite the diversity of languages used for the interpretation of geo-analysis models, there are some common basic features. Data structures and algorithms are the two fundamental elements in model implementation programs, and all model data should be converted to variables in programming codes. Regardless of the complexity of the model data, the program variable types are relatively fixed. Model data can be completely and uniformly described as long as the program variables are fully described. Here, this technical-based data description concept is employed, and relatively stable value types are explored to build the Universal Data eXchange (UDX) model.

Although the program variables, such as *int*, *double*, *struct*, *matrix*, *table*, and *dataset*, are also flexible and dispersive, the variables can be limited to certain types. A simple classification could be implemented by the *Value Type* (e.g., *int*, *double*, and *string*) and *Assemble Type* (e.g., *array*, *list*, and *table*). Other complex data types in the model program code can be constructed using these basic element types. To implement and execute the geo-analysis models, the model data need to be converted to variables of these value types in the program code. From this viewpoint, the heterogeneity of the model data can be reduced by describing them with these two basic and relatively stable value types.

The hierarchy or level method is frequently used in the geo-information modeling field according to the common recognition of geo-information. In addition, hierarchy

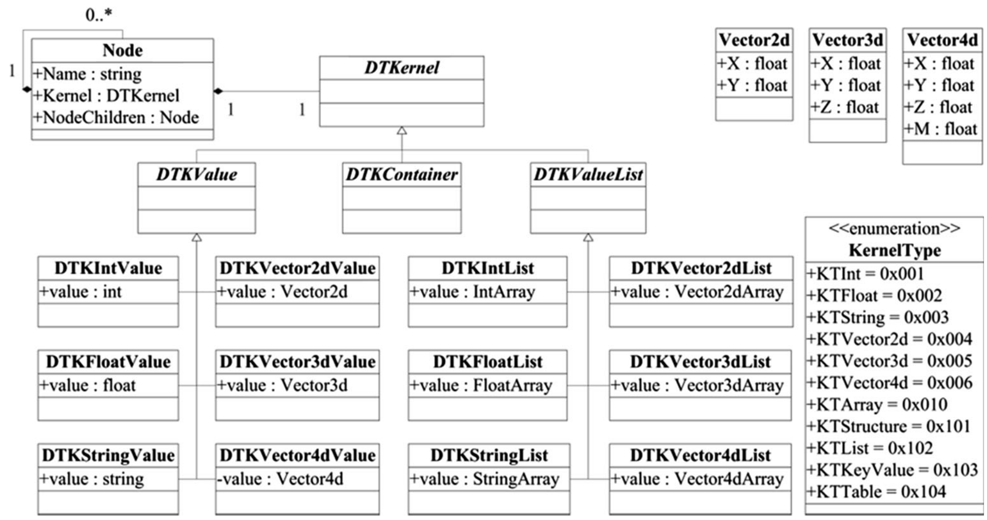


Fig. 1 Basic design of the UDX model

models, such as XML (Extensible Markup Language), HTML (HyperText Markup Language), and JSON (JavaScript Object Notation), are widely used in the data management and information processing fields; they are clearly powerful data-constructed logical approaches. Based on the above analysis, the hierarchy model was employed to organize model data of extreme complexity conveniently and robustly.

The two basic sets of stable value types can be used to describe the internal content of model data, and the hierarchy model can be employed to organize the structure of model data. This approach forms the initial basis of the UDX model. The UML diagram of the UDX model is shown in Fig. 1. The UDX model contains two basic construction elements, *Node* and *Kernel*. *Node* controls the model data's hierarchical structure; *Kernel* controls the specific behavior of *Node*. Every *Node* has a unique *Name* and *Kernel* and possesses a particular number of child *Nodes* (0–*n*).

1. If a *Node*'s kernel type belongs to the *DTKValue*, then the kernel should be limited to *Int type*, *Float type*, *String type*, or *Vector type* (i.e., *Vector2d*, *Vector3d*, and *Vector4d*). In addition, the node cannot have any child node; the value that the node contains must be consistent with the kernel type.
2. The *DTKValueList* kernel indicates that the *Node* has a list of values, and the values must be the same type as one of the kernel types in the *DTKValue* group. Additionally, the *DTKValueList* cannot have any child node.
3. If the node's kernel type belongs to the *DTKContainer*, it can be one of four types: (1) *Structure*, (2) *List*, (3) *Key Value*, or (4) *Table*. The *DTKContainer* node itself does not hold any data value. The specified kernel type also applies to the child nodes, and all the data should be stored in the child nodes.

The detailed kernel types are shown in Table 2. The examples in the table mainly focus on the input and output data of a watershed delineation model. For a group of *DTKValue* types, the *KTInt* type node is explained by "Analysis Window Size" in (a.1), the *KTFloat* type node is explained by "Flow Accumulation Threshold" in (a.2), and so on.

Table 3 shows some typical examples of the UDX data structure and its simple visualization. Simple features (e.g., points, polylines, and polygons) and grid data (e.g., the DEM and raster data) can be easily constructed with UDX, along with other, more flexible data (e.g., plain text and structural table data).

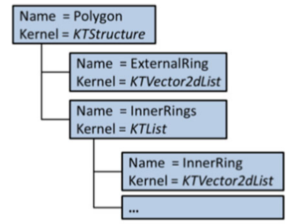
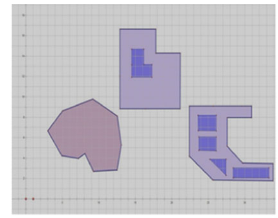
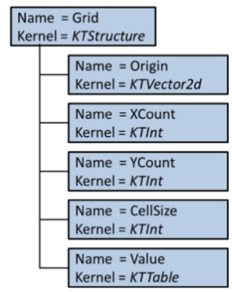
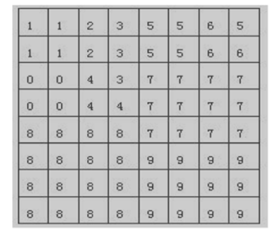
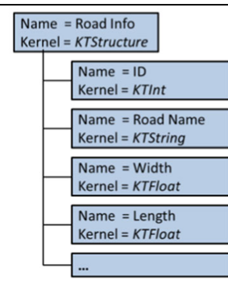
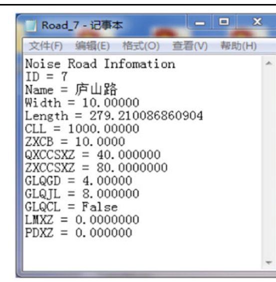
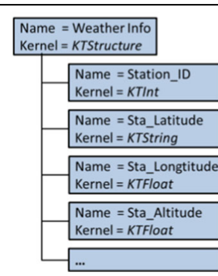
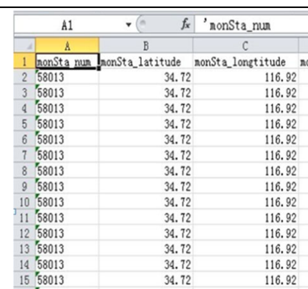
Universal Data eXchange model

Unlike other intermediate data formats in data conversion, the UDX model is not only a data format but also a set of methods to represent model data uniformly and completely. It is platform independent, program-language independent, storage independent, and application transparent. As the meaning of data content largely depends on a modeler's individual understanding, some auxiliary information should also be described to ensure that model providers and model users can communicate unambiguously regarding the model data. Using the basic design of the UDX model in Sect. "Basic concept of the data description model", UDX Data were developed for data value content storage and description (as shown in Table 2). Meanwhile, UDX Schema was developed for data information description and data conversion. UDX Schema is a self-descriptive

Table 2 *Kernel* types of *Node*

<table>
<tr><th>Group</th><th>Kernel Type</th><th>Example</th></tr>
<tr><td>DTKValue</td><td>KTInt ➔ (a.1)
KTFloat ➔ (a.2)
KTString ➔ (a.3)
KTVector2d ➔ (a.4)
KTVector3d ➔ (a.5)
KTVector4d ➔ (a.6)</td><td>(a.1) <XDO name=“Analysis_Window_Size" kernelType="int" value="3" />
(a.2) <XDO name="Flow_Accumulation_Threshod“ kernelType="float" value="7200.0" />
(a.3) <XDO name="Landuse_Type" kernelType="string" value="Grass" />
(a.4) <XDO name="Watershed_Pour_Out" kernelType="vector2d" value="512312.4323, 412332.2989" />
(a.5) <XDO name="Sample_Point_Depth" kernelType="vector3d" value="542166.19, 3546342.10, 0.67" />
(a.6) <XDO name="Sample_Point_Velocity" kernelType="vector4d" value= "542166.19, 3546342.10, 0.845, 0.263" /></td></tr>
<tr><td>DTKValueList</td><td>KTIntList ➔ (b.1)
KTFloatList ➔ (b.2)
KTStringList ➔ (b.3)
KTVector2dList ➔ (b.4)
KTVector3dList ➔ (b.5)
KTVector4dList ➔ (b.6)</td><td>(b.1) <XDO name="Feature_ID" kernelType="int|list" value="0,1,2,3,4" />
(b.2) <XDO name="Watershed_Area" kernelType="float|list" value="103.212, 453.235, 1018.14, 763.213" />
(b.3) <XDO name="Landuse_Code" kernelType="string|list" value="URMD,CRDY,CRIR,GRAS" />
(b.4) <XDO name="Boundary_Line" kernelType="vector2d|list" value="3876.676,2722.342; 4175.325,2816.684;
4473.974,2911.024; 4683.389,3098.370;
4978.841,3195.912; 3876.676,2722.342; …" />
(b.5) <XDO name="Section_Depth" kernelType="vector3d|list" value="166.191,6342.109,0.667; 266.191,6342.109,0.721;
366.191,6342.109,0.721; …" />
(b.6) <XDO name="Section_Velocity" kernelType="vector4d|list" value="166.191,6342.109, 0.845,0.263;
266.191,6342.109, 0.333,0.606;
366.191,6342.109, 0.915,-0.116; …" /></td></tr>
<tr><td>DTKContainer</td><td>KTStructure ➔ (c.1)
KTList ➔ (c.2)
KTKeyValue ➔ (c.3)
KTTable ➔ (c.4)</td><td>(c.1) <XDO name="Watershed_Info" kernelType="structure">
<XDO name="name" kernelType="string" value="" />
<XDO name="area" kernelType="float" value="" />
<XDO name="perimeter" kernelType="float" value="" />
<XDO name="boundary" kernelType="vector2d|list" value=" 3876.676,2722.342; 4175.325,2816.684; …" />
</XDO>
(c.2) <XDO name="Subbasin_Information" kernelType="list">
<XDO name="subbasin_1" kernelType="structure">
<XDO name="id" kernelType="int" value="1" />
<XDO name="area" kernelType="float" value=“419643.843" />
<XDO name="perimeter" kernelType="float" value=“353265" />
</XDO>
<XDO name="subbasin_2" kernelType="structure">
<XDO name="id" kernelType="int" value="2" />
<XDO name="area" kernelType="float" value=“329113.923" />
<XDO name="perimeter" kernelType="float" value=“212314" />
</XDO>
</XDO>
(c.3) <XDO name="Monitor_Station" kernelType="keyvalue">
<XDO name="key" kernelType="int" value="1" />
<XDO name="value" kernelType="structure">
<XDO name="wind_speed" kernelType="float" value=“321.59" />
<XDO name="air_monisture" kernelType="float" value=“1.242" />
<XDO name="air_presure" kernelType="float" value=“98.32" />
</XDO>
</XDO>
(c.4) <XDO name="Basin_Feature" kernelType="table">
<XDO name="id_column" kernelType="int|list" value="0,1,2,3, …" />
<XDO name="area_column" kernelType="float|list" value="77.514, 51.514, 224.96, 169.4, …" />
<XDO name="flow_in_column" kernelType="float|list" value="5.636, 3.915, 27.48, 14.15, … " />
<XDO name="flow_out_column" kernelType="float|list" value="5.632, 3.915, 27.43, 14.13, … " />
</XDO></td></tr>
</table>

Table 3 Typical examples of the UDX model and its simple visualizations

Sample data	The UDX structure	Simple Visualization
Simple feature (Polygon)		
Grid Data		
Plain Text Data		
Structural Table Data (MS Excel)		

structure that is strictly in accordance with UDX Data. In addition, the attachment libraries were imported into UDX Schema to support the description of the data information completely.

UDX schema

According to the analysis in Sect. "Basic concept of the data description model", the model data were converted into program-related variables to complete the execution or computation of the models. Different organizational structures can be employed to construct the same data content. For example, consider the *time* value, which is commonly used as the start and end times in the simulation-type models. The *time* value can be constructed as "year/month/day/hour" (2014/5/6 18), "year/month/day" (2014/5/6), "year/day" (2014/126), and so on. Traditionally, model providers offer documents to explain such data, and model users should explore these documents carefully to obtain information on the data. A large amount of effort from model users is required to prepare model data properly, as the illustrative expressions of model providers may cause misunderstanding (e.g., model providers and model users may originate from different countries and use different languages). In addition, some model data are deeply related to the discipline's background knowledge, which is rarely known by model users in other disciplines.

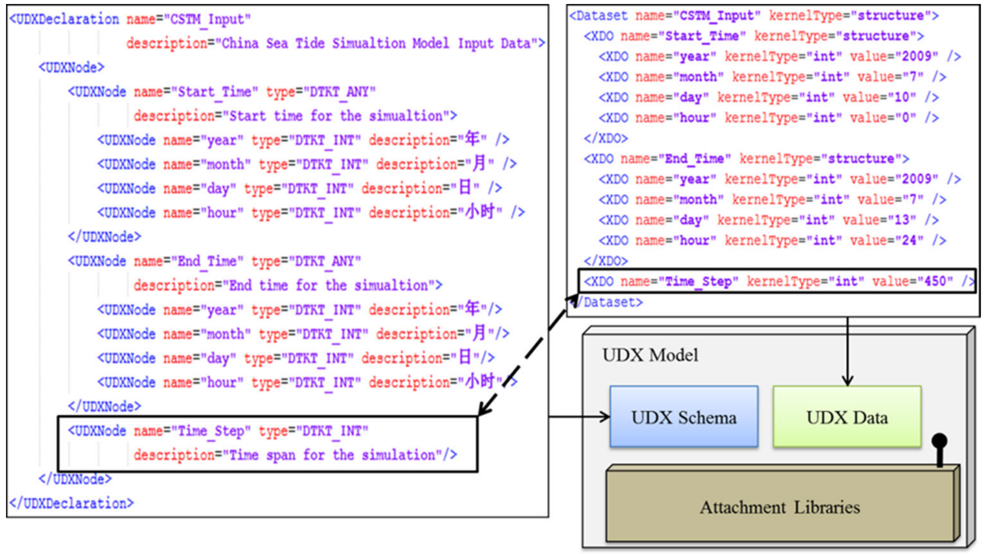


Fig. 2 Sample UDX schema and its corresponding UDX data

Regarding this problem, the proposed UDX model contains two basic structures: (1) UDX Schema and (2) UDX Data. Both UDX Schema and UDX Data are derived from the basic design of the UDX model, namely, the constructed and completed UDX model. UDX Schema mainly describes the model data, and UDX Data mainly organizes and stores model data content.

"China sea tide simulation model input data" is described in Fig. 2: the top-right illustrates UDX Data and the left side is the corresponding UDX Schema. "Start time", "End time" and "Time step" are the 3 main parts of this example data. Therefore, both UDX Data and UDX Schema contain 3 child nodes ("*Start_Time*" node, "*End_Time*" node and "*Time_Step*" node). The "*Start_Time*" node and "*End_Time*" node consist of 4 *int* type child nodes: "*year*", "*month*", "*day*" and "*hour*". The "*Dataset*" node in UDX Data is mapped with the "*UDXDeclaration*" node in UDX Schema, and the "*XDO*" node in UDX Data is mapped with the "*UDXNode*" node in UDX Schema. The "*Time_Step*" node in UDX Data provides "450" as the specific value for the model computation, and the related "*Time_Step*" node in UDX Schema provides "time span for the simulation" as the description for such a data node.

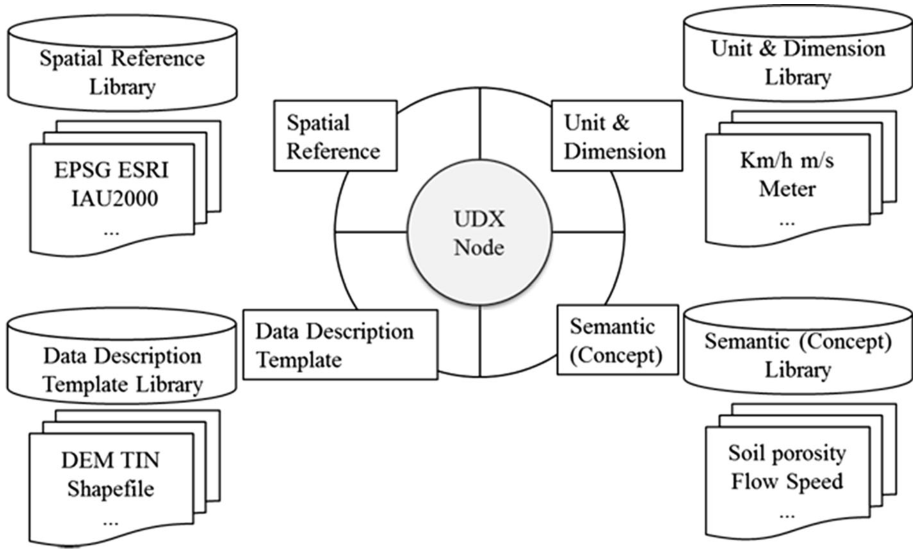
UDX Data and UDX Schema are closely related; they share the same structural info and data type information for each node (as Fig. 2 shows). UDX Data contains the specific data value, while the corresponding UDX Schema contains the description. Although the description attribute within the data node in UDX Schema explains the basic information of the data node, a range of other attachment information should be provided to modelers. In Fig. 2, the *Time_Step* node means "the time span for the simulation"; such a description cannot provide comprehensive information to modelers, such as the unit of the *Time_Step* value (as the top-left area of UDX Data shows, the value 450 could mean 450 s or 450 min). The attachment library strategy was employed to solve such a semantic, explicit problem, as discussed in Sect. "Attachment information of UDX".

Attachment information of UDX

Highly specialized data structures, resolutions, and time/space domains are often used in discipline-specific models (Argent 2004a, b; Jabbar and Zhou 2013). In terms of model data, the value content (commonly represented as numbers) is often directly presented to modelers. However, besides the value content, a bulk of essential information in the data is of interest to modelers and can largely affect the execution results of the models. Through a series of geo-analysis model case studies, we summarize this essential information as (1) the information related to units and dimensions, (2) the semantic (concept) information of the data, (3) the geo-data spatial reference information, and (4) the information related to common formats (data description template information).

This attached information does not always exist in the model data; for instance, the spatial-reference information is only related to geo-spatial data. To satisfy both the flexibility of semantic information and the demands of structural representation, this study designed the attachment strategy for the UDX model. As Fig. 3 shows, four libraries were designed for the attachment information, and a mapping table was used to store the attachment information of a UDX node by linking the node's *id* with the libraries' item *id*. The unit and dimension library provides a full measurement scale. The semantic (or concept) library provides accurate semantic (conceptual or ontological) information for the UDX data by borrowing the semantics

Fig. 3 Model data attachment libraries



systems and webs from other disciplines. The spatial reference library provides entire coordinate or projection information for the UDX data (if the data are spatial data), and the data description template library provides an interface to other domain-related, commonly used data formats (e.g., *Shapefile*, *Geotiff*, and *WKT*).

Unit and dimension library

A unit is a definite magnitude of a physical quantity and defines the measurement scale of a numerical value. Dimension is an expression constructed by basic quantities and is widely used in dimensional analysis (Gruber 1995). According to the SI system (International System of Units) (Burdun 1960), there are seven basic dimensions and units (as Table 4 shows). All the other dimensions are the products of these basic dimensions.

Based on the SI system, this study built a systemic unit and dimension library. Figure 4 shows the design of the Unit and Dimension object model. In Fig. 4, *DimensionClass* (dimension) is one of the attributes in *UnitAbstract* (unit), and both the dimension and unit have the attribute of *Localization*. *Localization* provides the description of the dimension and unit (*name*, *type*, *description*, and *symbol*), and *Localization* is language-related (*LocalizationEnum*): ZH_CN means Chinese and EN_US means English. The units were divided into three types: *UnitBasic* (basic unit), *UnitPrimitive* (primitive unit) and *UnitDerivation* (derivation unit). Consider the length of a road segment as an example. The *kilometer*, *foot*, or *centimeter* units can all be used, but they all share the dimension *Length*. Corresponding to the *Length* dimension, the *meter* (symbol m) unit is the basic unit, and 1000 *meters* is the *kilometer* (symbol km) unit, which belongs to the primitive unit. The velocity unit (e.g., km/h) is composed of two units (*kilometer* and *hour*) and belongs to the derivation unit.

Table 4 Basic dimensions and units in the SI system

Quantity	Dimension	Dimension symbol	Unit	Unit symbol
Length	L	L	Meter	m
Mass	M	m	Kilogram	kg
Time	T	t	Second	s
Current	I	I	Ampere	A
Thermodynamic temperature	H	T	Kelvin	K
Amount of substance	N	n	Mole	mol
Luminous intensity	J	Iv	Candle	cd

According to base units of SI system http://en.wikipedia.org/wiki/International_System_of_Units

Semantic (concept) library

Semantic information is usually implied in the data values, as it is information that explains what the numerical value means. Because of recognition differences, a single value without any semantic information may cause ambiguities. For instance, a model user meets the demand of providing input data for a *City Road Noise Simulation* model that requires Wide Road centerline data. Widths greater than either 30 or 35 m could describe the Wide Road, and a four-lane road can also be interpreted as a Wide Road. The simulation result varies according to the Wide Road input

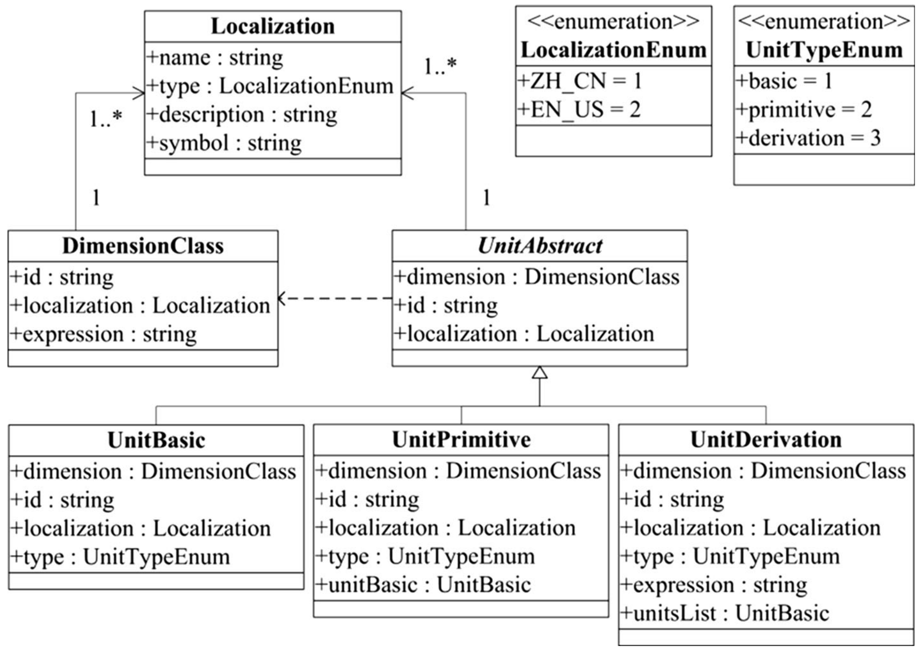


Fig. 4 Design of the unit and dimension library

data provided by the model users. The terms "semantic" or "semantic web" were borrowed from computer sciences to solve such a problem. The term "ontology" was borrowed from philosophy and plays a key role in semantic applications (Maedche and Staab 2001; Jung and Sun 2010).

Different terms may refer to similar concepts and vice versa, which causes semantic heterogeneity in the dataset. Several research projects have provided solutions, such as the NASA's Semantic Web for Earth and Environmental Terminology (SWEET) (Raskin and Pan 2003) and SEDRIS's Environment Data Code Specification (EDCS) (Leite et al. 2006). All these semantics projects have been applied and extended in several fields in the earth sciences, and several tools have been developed for ontological analyses and design, such as Protégé (Knublauch et al. 2004), SWOOP, SMORE (Kalyanpur et al. 2006a, b), and Altova SemanticWorks (http://www.altova.com/products_semanticworks.html). In addition, several reasoning engines, such as the RacerPro Reasoning Engine, have been used in domain-related applications (Chen et al. 2005; Weithöner et al. 2007; López-De-Ipiña et al. 2008; Liu et al. 2014).

This study does not focus on building the semantics or semantic web itself but instead tries to attach model data with the semantic or concept that exists in the external semantic library. The design of such a reference strategy is shown in Fig. 5. The *Concept* class was designed to represent a UDX data node's semantic information, and it possesses a unique identification (*id*), a well-known name (*wkName*), a localization collection (*localization*), a unique resource link website (*urn*) and a set of semantic category

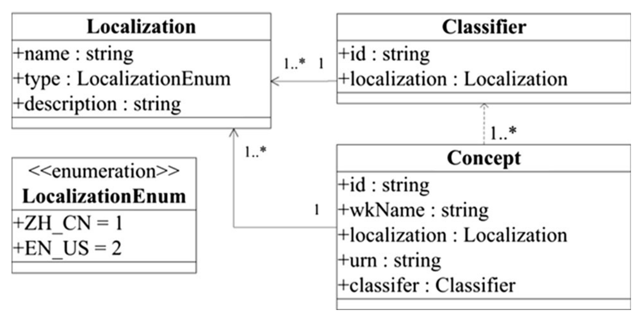


Fig. 5 Design of the semantic (concept) library

systems (*classifier*). The information in the external semantic library was abstracted as a *Classifier* class and has been provided with a unique *id* and *localization* collection. The *Concept* object can belong to several *Classifiers*, which ensures that different external semantic libraries can be integrated and that the information can be conveyed without being lost. For example, consider the "Wet Land" concept item in the EDCS system; its id is "LAND_SURFACE_WET_LAND", its well-known name is "Wet Land", and it belongs to the "EDSC" classifier.

Spatial reference library

As long as the model data have spatial information, the spatial reference system should be defined clearly (Bonham-Carter 1994; Golovko et al. 2014). Location is the most important information for spatial data; without spatial reference information, such a location can be ambiguous. As Fig. 6 shows, the same point can be represented as

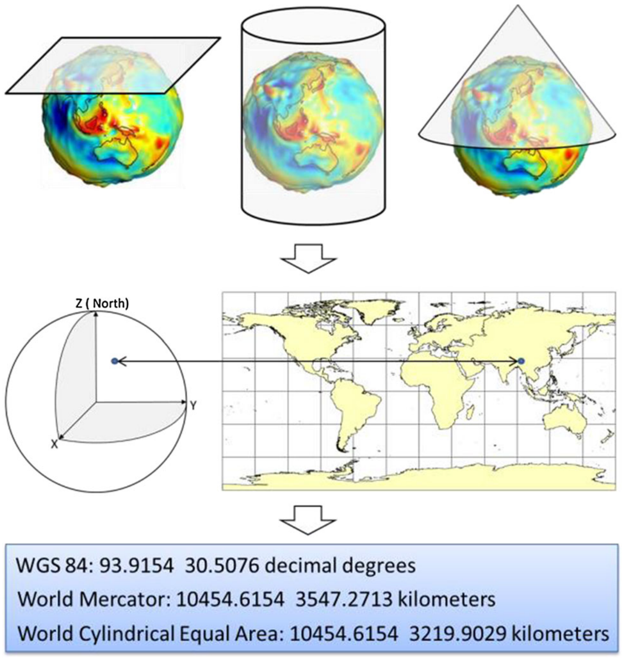


Fig. 6 Different coordinate values of the same point in different spatial reference systems

different coordinates by using different coordinate systems or projection systems (same point, different coordinate values in different coordinate or projection systems). The spatial reference information directly influences the model execution as the coordinate value varies widely according to the projection method.

A spatial reference defines an ellipsoid, a datum using that ellipsoid, and either a geocentric, geographic or projection coordinate system. In addition, the projection is always associated with a geographic coordinate system. The European Petroleum Survey Group (EPSG http://www.epsg.org) has built a huge set of predefined spatial references, in which each has a unique code. These codes are commonly used in several research fields. This study imported the EPSG spatial reference table and employed prj4 (an open source coordinate and projection transformation program library, http://trac.osgeo.org/proj/) to automatically transform coordinates among different spatial references.

Figure 7 shows the design of the spatial reference object model. Similar to the Unit and Dimension library and the Sematic (Concept) library, the *Localization* item provides the basic description information of one *SpatialReference* object. *SpatialRefEnum* demonstrates the type of *SpatialReference* object, and the *type* attribute retains such information. In addition, the *id* attribute indicates the unique identifier of any *SpatialReference* object (mapper with the code in the EPSG spatial reference table), and the *wkName* attribute stores the well-known name information.

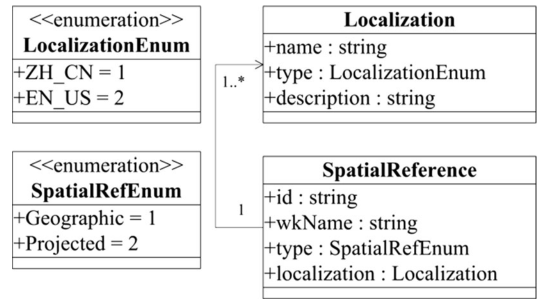


Fig. 7 Design of the spatial reference library

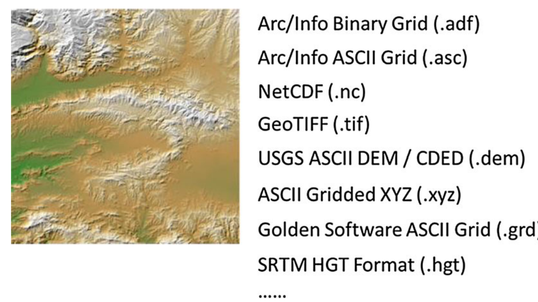


Fig. 8 Commonly used raster data formats

The *SpatialReference* object also contains the information of unit, which can be linked to specific item in unit and dimension library, and the transformation of units also can be conducted with the help the prj4 library.

Data description template library

Through the efforts of researchers in various study areas, most of these systemic disciplines have formed and accumulated a broad array of commonly used and recognized data formats. The data structures and organization methods of these generally accepted model data formats have been studied and proven through a series of modeling research. Most model researchers would use such data formats frequently as input and output data formats. Figure 8 shows several generally used raster data formats.

The data description template library is an open library because of the imperfect knowledge of any individual researcher or any single research team. As environmental problems worsen, generally accepted data formats change, new formats emerge, and the data description template library continues to grow. To address such a dynamic feature, this study designed a registration strategy by using the *IGeneralDataContainer* interface (as Fig. 9 shows). The *IGeneralData* interface can be implemented and registered to the *GeneralDataContainer*. Thus, the commonly used data formats can be assembled to build the data description template library for modelers. Figure 8 shows some commonly used data formats, such as *ShapefileData* (for vector

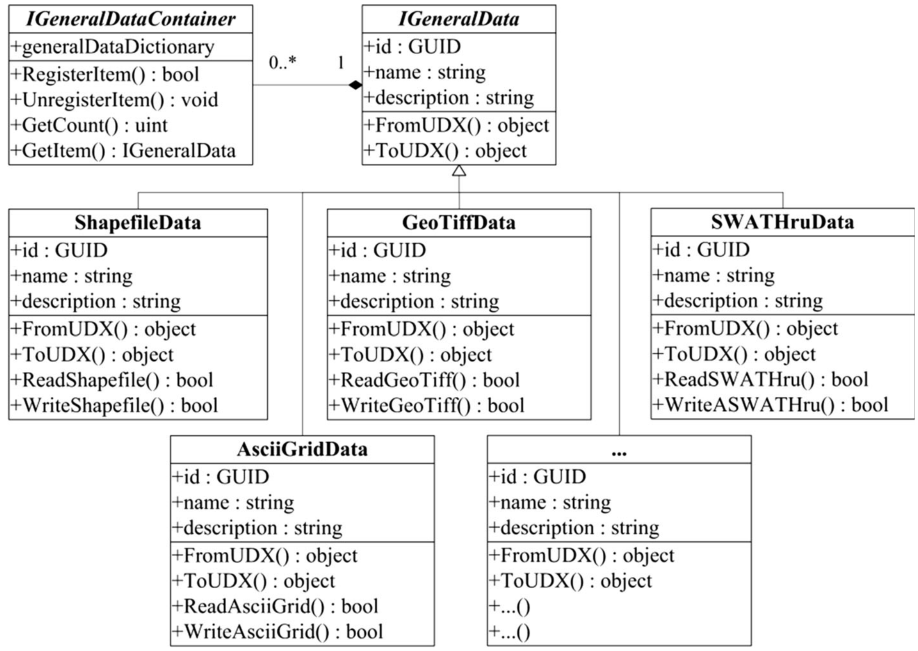


Fig. 9 Basic design of the data description template library

data), *GeotiffData* and *AsciiGridData* (both for raster data), and *SWATHruData* (SWAT Hydrology Response Unit data), which is widely accepted in hydro-modeling areas.

Model data preparation and pre-processing work with the UDX model

According to the above analysis, model data preparation and pre-processing are the most time-consuming tasks towards applying geo-analysis models to practical problem solving. By using UDX to describe model data explicitly and visibly, model users can understand model data more clearly, structurally and unambiguously. Freed from various data formats and unreadable files, model users can focus on the data content. UDX Schema and UDX Data can provide model users with full control of the data structure, content, and attachment information, which include units and dimensions, sematic information, spatial references, and data description templates.

Figure 10 uses the data of a slope analysis model in the digital terrain study area to explain the application of the UDX model. The "Slope_Analysis Input" UDX data shown in Fig. 11a are composed of two main parts: DEM data ("*Dem_Data*" node) and analysis window size ("*winSize*" node). The "*Dem_Data*" node stores the terrain height information of the study area. The basic geographic location information of "*Dem_Data*" is provided by the "*head*" node, and the "*body*" node stores the height information of all the grid cells by using the kernel type of *DTKList*. The "*winSize*" node is defined as the kernel type of *DTKVector2D*, and the value "3, 3" means the analysis window size is 3*3.

In Fig. 10b, the corresponding UDX Schema is explained: the UDX schema of such data consists of *UDXNode* and *SemanticAttachment*. *UDXNode* provides the data structure and content information, while *SemanticAttachment* provides the Unit, Concept, Spatial Reference and Data Description Template information.

To generate UDX Data and UDX Schema, a set of tools was designed and developed so that modelers can construct flexible data structures by using the GUI tool (as Fig. 11 shows). The left side of Fig. 11 is a simple example of model data, which describes time span information. The right side of Fig. 11 describes the DEM data, which is constructed as a *Geotiff* structure. Additionally, there can be more than one data format for the same information. The UDX data can be constructed according to the modeler's demands.

According to the analysis above, several data processing methods should be undertaken to prepare proper model data, such as resampling raster data into the required resolution, clipping the original data into a small-scale study area, and merging multi-data into one hybrid data, to execute the models correctly. Traditionally, such work is conducted separately and individually by either programming or using data processing tools. The data processing work is not easily shared because of the data's heterogeneity and the tools' platform-dependency. Even when simply assembled, none of these processing methods can form a practical data processing library, as different methods require different data formats and the disciplinary understanding of the modeler id based on semantic information. Such a simply constructed method library without any uniform data interface can lead to confusion and thus

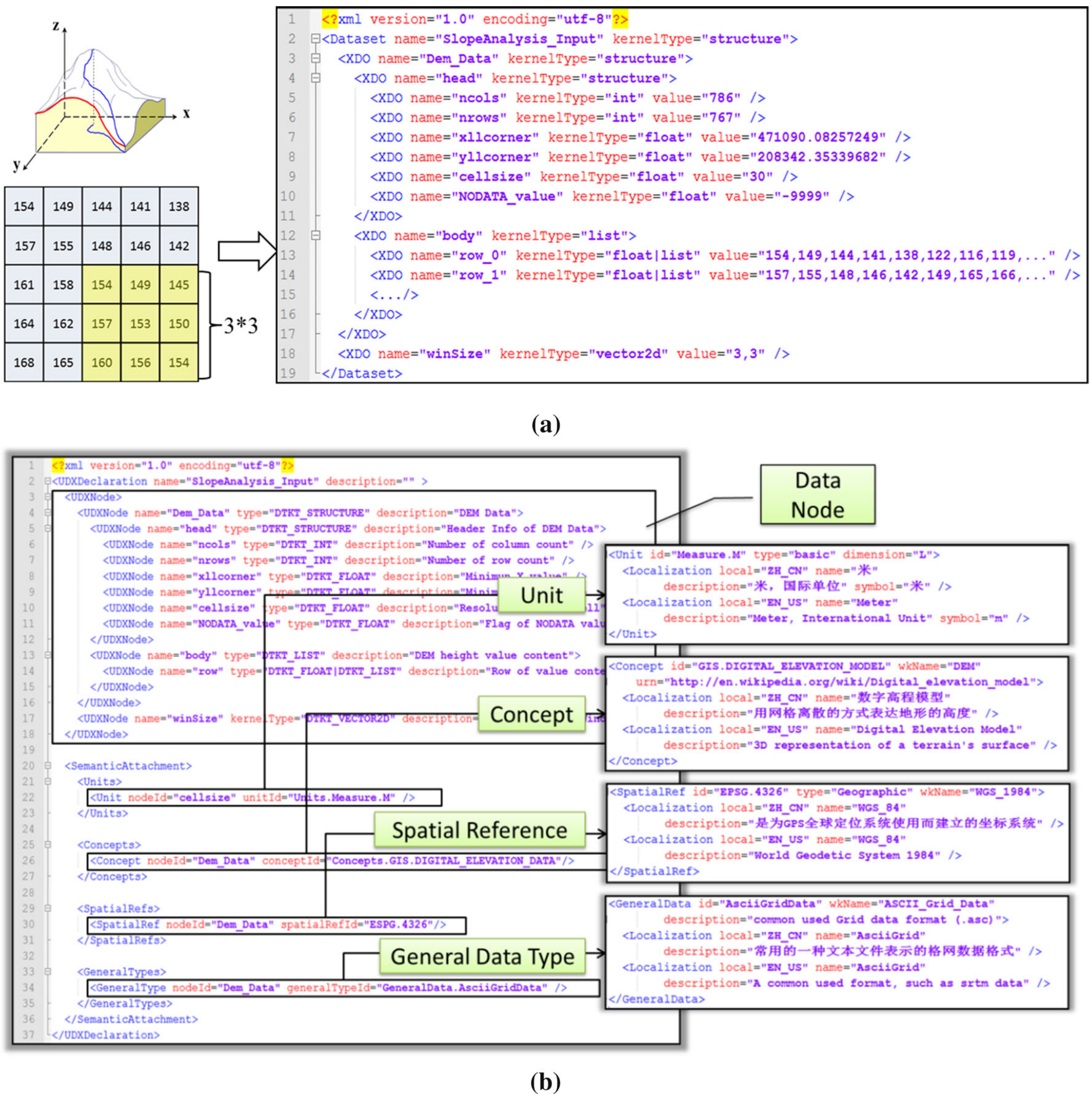
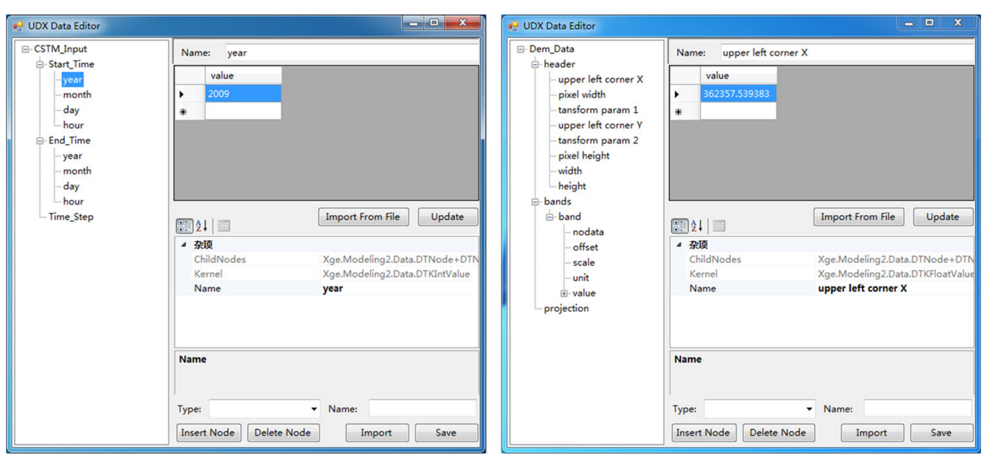


Fig. 10 Example slope analysis input data with UDX data and UDX schema

Fig. 11 Examples using the UDX model to prepare model data



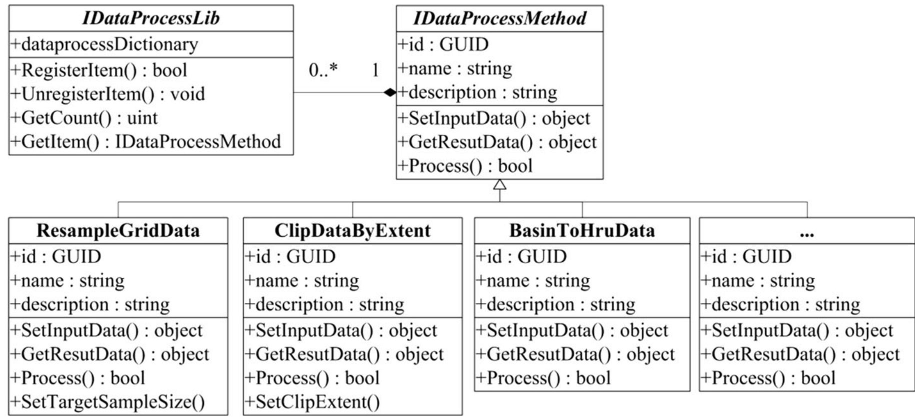


Fig. 12 UDX data processing method library

increase the difficulty of accumulating data process methods.

Based on the UDX model, model data can be fully described in a uniform way. A model data processing method can easily be implemented with UDX data, as the most commonly used data formats can be bi-directionally transformed. In this context, modelers that possess uniform data with fully described semantic information and other related information can perform all the data processing methods using a transparent interface. Figure 12 shows the basic design of the data processing method library. The interface *IDataProcessMethod* was designed to encapsulate the behavior of the actual data processing method. In addition, the implemented classes of *IDataProcessMethod,* such as *ResampleGridData*, *ClipDataByExtent*, and BasinToHruData, can be registered to the *IDataProcessLib*. Similar to the data description template library, the data processing method library is also an open library, which can be used by all geo-analysis modelers. Through this strategy, various model data processing methods can be accumulated and organized for a wide variety of modelers. Along with the development of the data processing method library, modelers and model users can more conveniently apply these models.

Case studies

According to the above analysis, in this study, four attachment libraries were built and stored in the SQL Server database, which can be queried and extended by modelers. As each of the items in these attachment libraries can be uniquely identified by the attribute *id*, the stored data tables in the database are organized by *id* (as the primary key of the data table), and the contents are storied in the XML field. Based on the SI system, a unit and dimension library was built; based on the EPSG spatial reference table, a spatial reference library was built; imported from the SEDRIS EDCS concept system and the book "geographic concept dictionary" (in Chinese), a semantic library was built; and with the help of GDAL/OGR open source library, a range of common used spatial data formats were imported to the data description template library. Therefore, employing these resource libraries, UDX Schema can be more easily constructed.

Compared with the SEDRIS project, the proposed UDX model in this study put more attention on the understanding of data that serves to provide a better communication approach between model provides and model users. The building of UDX attachment information library has imported many resources in the SEDRIS project that gives more flexibility and extendibility of the UDX model. As the efficiency of data processing is largely depended on the programming implementation, this paper does not focus on the data conversion speed but on the possibility of the model sharing and integration.

To verify the capabilities of the UDX model, two typical model integration case studies were conducted: (1) the TauDEM tool was integrated with the SWAT model, and (2) the CSTM model was integrated with the POM model. The data description capabilities and the comprehensiveness of the UDX model can be verified in both of these case studies. The first case study focuses on data preparation, as the data format varies from the original data to the TauDEM and SWAT data. The second case study focuses on data pre-processing, as the CSTM model and POM model use two different spatial sampling methods.

TauDEM tool integrated with the SWAT model

TauDEM (Terrain analysis using Digital Elevation Models) is a set of tools for the analysis of terrain using digital elevation models (http://hydrology.usu.edu/taudem/taudem5/index.html). It provides specialized functions for hydrology analysis and terrain analysis. However, the SWAT model (Soil and Water Assessment Tool) is a famous model in the watershed research field (SWAT http://swat.tamu.edu/). The input data of the SWAT model

Table 5 Differences between TauDEM and SWAT

Differences	TauDEM	SWAT
Relevant discipline	Terrain analysis	Hydrology
Program language	C++	Fortran
Release mode	Source code/ executable	Source code/executable
Input data type	Binary file: Geotiff and Shapefile	Plain text file: Hydro-responded unit data (.hru)
		Basin data (.sub)
		Ground water data (.gw)
		etc.

include an array of underlying surface information, such as base sub-basin data, hydro-response unit data, soil type data, and ground water data. Generally, SWAT model users should prepare such underlying surface data by incorporating particular GIS software or tools. TauDEM is one of the most frequently used high-quality GIS tools to generate and prepare these data for SWAT model execution.

The base DEM data were downloaded from the SRTM website, which can provide worldwide DEM data with 30 or 90 m precision. The DEM data format is the ASCII grid type. The original grid data type of TauDEM is the *Geotiff* format, and the original vector data type of TauDEM is the *Shapefile* format. Most of the SWAT model data are in flexible plain text format. Table 5 shows the differences between TauDEM and SWAT.

Figure 13 shows the basic integration workflow and the data preparation workflow. The SRTM data can be transformed into the *Geotiff* format with the UDX data description template library, at which point TauDEM can be executed. The TauDEM execution's watershed data results can be transformed into HRU data. When supplemented with other input data, the SWAT model can be executed. Finally, the SWAT output data are transformed into *Shapefile* data. Figure 14 shows a snapshot of the data processing progress. The conversion of ASCII grid data to Geotiff data is shown in Fig. 14a, and the integration of the TauDEM and SWAT models is shown in Fig. 14b.

CSTM model integration with the POM model

In the research field of dynamic water processes, multi-scale models that need to be integrated and nested as a single model rarely meet the complex demands of the simulation of water dynamics. This study chose the integration of the China Sea Tide Model (CSTM) and the Princeton Ocean Model (POM) as a case study to verify the feasibility of the UDX data process library. As Table 6 shows, the CSTM model is a 2-D large-scale model (Lin et al. 1997), while the POM model is a 3-D small-scale model (Blumberg and Mellor 2013).

In the integration process, the CSTM model provides large-scale multi-temporal tidal data within the study area. Based on these data, the model processes the externally input small-scale grid-coordinate data with an interpolation method, and the corresponding small-scale tidal data can be generated. Similarly, through the interpolation of large-scale depth data and small-scale grid-coordinate data, small-scale depth data for the POM model can be generated. Therefore, the POM model can drive the 3-D simulation of tidal data, as the small-scale tidal data and small-scale depth data are the input data for the POM model.

The integration of these two models involves two different spatial discretization schema (CSTM is a large-scale grid and POM is a small-scale grid). As Fig. 15 shows, the user provides the input data for the CSTM model and then uses the CSTM tide result data and the small-grid data to call the resampling method in the data processing method library and obtain the small-scale tide data. Similarly, the

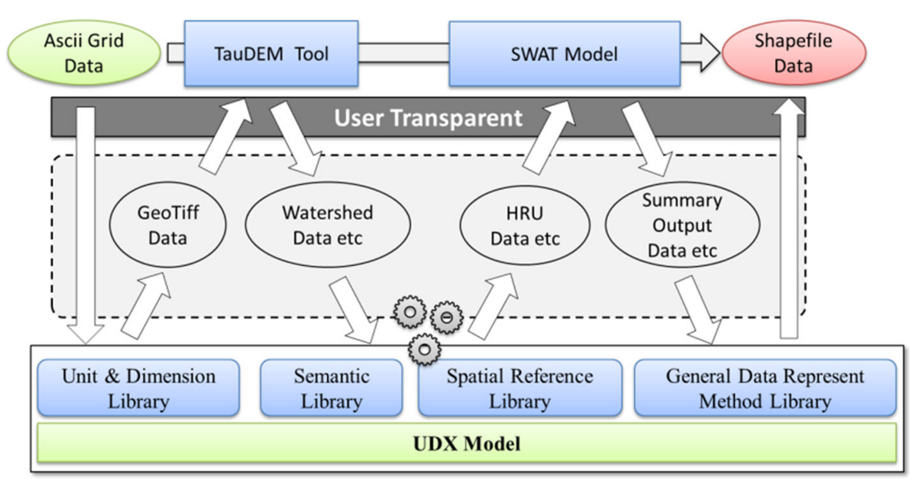


Fig. 13 Flow chart of the integration of TauDEM and SWAT

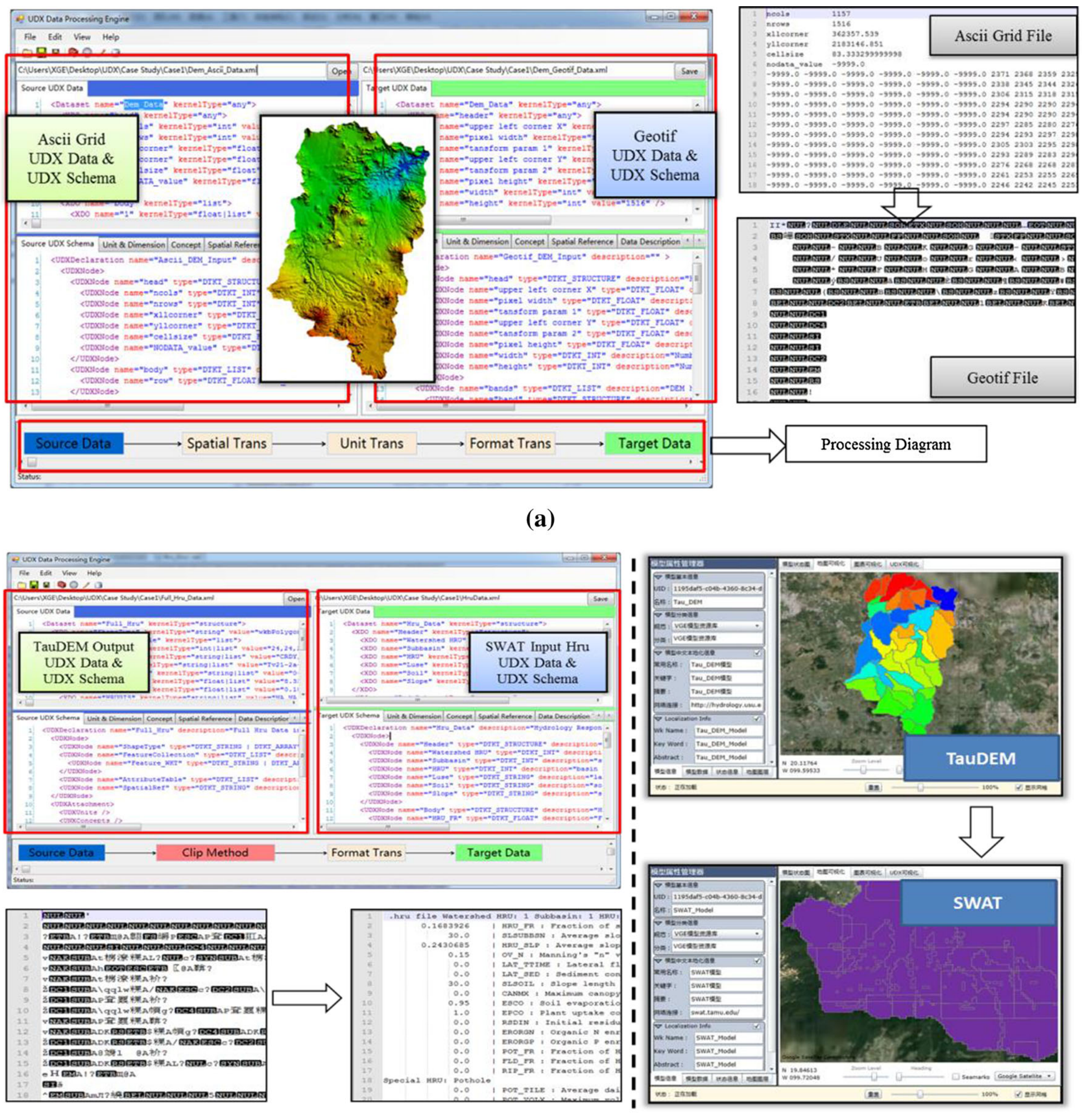


Fig. 14 Data processing with UDX in the integration of TauDEM and SWAT

small-scale depth data can be generated with the resampling method. The small-scale tide data and small-scale depth data are input into the POM execution model to obtain the final result. A snapshot of some of the data and a visualization of the final result are shown in Fig. 16.

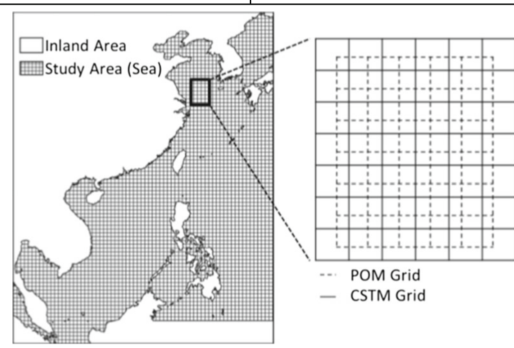
Conclusion and future work

The purpose of this study is to reduce the difficulty of model data preparation and pre-processing work, which is important in the reuse, sharing and integration of geo-analysis models. By taking advantage of the proposed UDX model, model data can be described in a uniform and unambiguous way. With the UDX model, the model data processing method library was designed to accumulate and organize various data processing methods, thus making it more convenient for modelers to reuse, share and integrate heterogeneous geo-analysis models.

However, as geo-modeling and geo-model integration research is synthetic work, future research is needed:

1. Transmission control of UDX Data in a network. As model data are generally "big data", the volume and size of the corresponding UDX Data make it difficult to transmit data in a network environment. A progressive data transmission strategy should be designed by taking advantage of theoretical or technological

Table 6 Differences between the CSTM model and POM model

Differences	CSTM	POM
Relevant discipline	Ocean tide simulation	Ocean circulation
Program language	Fortran	Fortran
Release mode	Executable	Source code / executable
Spatial dimensions	2D	3D
Input data	Plan text file (split by *tab* char)	Plan text file (split by *space* char)
		

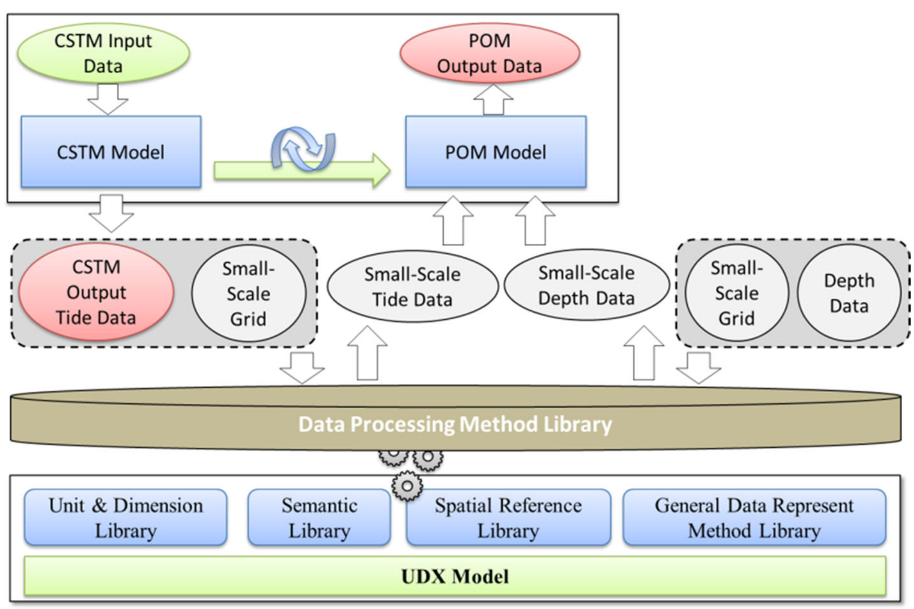


Fig. 15 The flow chart of the CSTM integration with POM

approaches in IT and data management domains, such as data compression technology and the level of detail (LOD) method.

2. Import the semantic reasoning method into the UDX model. This paper imported external semantic libraries into the UDX model to draw different conclusions from

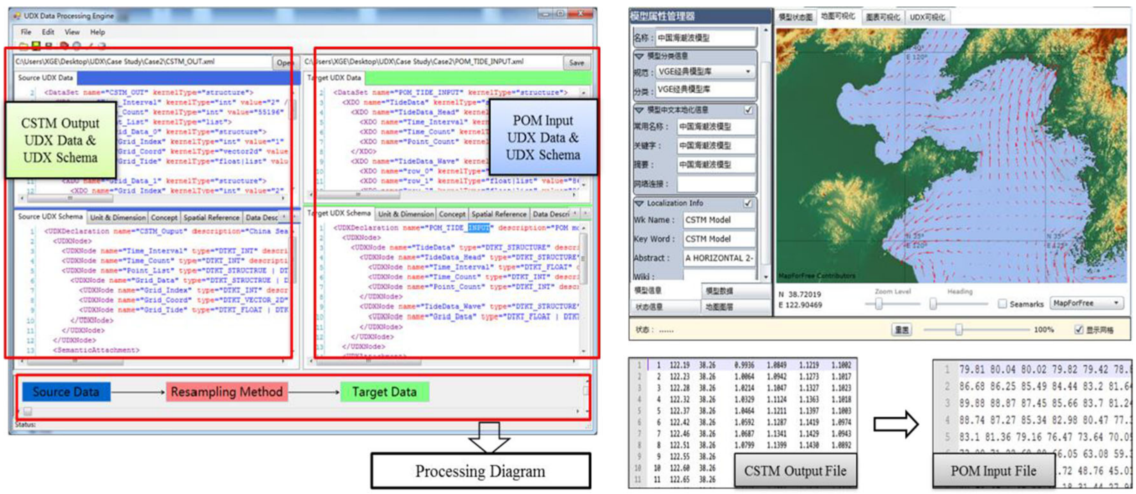


Fig. 16 Model data processing with UDX in the integration of CSTM and POM

model data in a uniform way. Although the UDX model can provide an interface to access such semantic information, the semantic conversion of model data still cannot be conducted. A simple but effective interface should be designed and developed to import an external semantic reasoning engine within the UDX model.

3. The implementation and improvement of model data processing method library and the development of the UDX data operation script language. As the model data processing method library is extendable, much more common used data processing methods should be implemented to improve the usability. And although the proposed model data processing library can reduce the effort and difficulty for model users, much programming work has to be performed to build such a library. Scripting languages have been proven as an effective method for numerical calculations in the IT domain. Several mature technologies can be employed to establish UDX data operation scripting languages, such as the Domain Specific Language (DSL) and YACC/LEX.

Acknowledgments We appreciate the detailed suggestions and comments from the editor and the anonymous reviewers. As the work described in this article involves many geo-analysis models, several research programs supported the work in this article: the National Key Basic Research Program of China (Grant No. 2015CB954102), the National Natural Science Foundation of China (Grant No. 41471317 and Grant No. 41101498), Project of NSF of Jiangsu Province, China (Grant No. BK20130904), the Priority Academic Program Development of Jiangsu Higher Education Institutions.

References

Ahuja LR, Ii JA, David O (2005) Developing natural resource models using the object modeling system: feasibility and challenges. Adv Geosci 4:29–36

Albanesi MG, Albanesi R (2014) A Decision-making support system for land use estimation based on a new anthropentropy predictive model for environmental preservation-theory, model, and web-based implementation. Int J Adv Intell Syst 7(1 and 2):85–102

Aminu M, Ludin ANBM, Matori AN et al (2013) A spatial decision support system (SDSS) for sustainable tourism planning in Johor Ramsar sites, Malaysia. Environ Earth Sci 70(3):1113–1124

Arctur D, Hair D, Timson G et al (1998) Issues and prospects for the next generation of the spatial data transfer standard (SDTS). Int J Geogr Inf Sci 12(4):403–425

Argent RM (2004a) An overview of model integration for environmental applications—components, frameworks and semantics. Environ Model Softw 19(3):219–234

Argent RM (2004b) Concepts, methods and applications in environmental model integration. Environ Model Softw 19(3):217

Argent RM, Voinov A, Maxwell T et al (2006) Comparing modelling frameworks—a workshop approach. Environ Model Softw 21(7):895–910

Aydi A, Zairi M, Dhia HB (2013) Minimization of environmental risk of landfill site using fuzzy logic, analytical hierarchy process, and weighted linear combination methodology in a geographic information system environment. Environ Earth Sci 68(5):1375–1389

Bagdanavičiūtė I, Valiūnas J (2013) GIS-based land suitability analysis integrating multi-criteria evaluation for the allocation of potential pollution sources. Environ Earth Sci 68(6):1797–1812

Basnyat P, Teeter LD, Lockaby BG et al (2000) The use of remote sensing and GIS in watershed level analyses of non-point source pollution problems. For Ecol Manage 128(1):65–73

Bhatt M, Rahayu W, Sterling G (2004) sedOnto: a web enabled ontology for synthetic environment representation based on the SEDRIS specification. In: Fall Simulation Interoperability Workshop

Bivand RS, Pebesma E, Gómez-Rubio V (2013) Spatial data import and export in applied spatial data analysis with R. Springer, New York, pp 83–125

Blind M, Gregersen JB (2005) Towards an open modelling interface (OpenMI) the HarmonIT project. Adv Geosci 4:69–74

Blumberg AF, Mellor GL (2013) A description of a three-dimensional coastal ocean circulation model. Three Dimens Coast Ocean Models. doi:10.1029/CO004p0001

Bonham-Carter G (1994) Geographic information systems for geoscientists: modelling with GIS, vol 13. Elsevier, Amsterdam

Britz W, Domínguez IP, Heckelei T (2010) A comparison of CAPRI and SEAMLESS-IF as integrated modelling systems. Environmental and agricultural modelling. Springer, The Netherlands, pp 257–274

Burdun GD (1960) International system of units. Meas Tech 3(11):913–919

Campbell AP, Hummel JR (1998) The dynamic information architecture system: an advanced simulation framework for military and civilian applications (No. ANL/DIS/CP-94722). Argonne National Lab., IL (US)

Chen H, Finin T, Joshi A (2005) Semantic web in the context broker architecture. Maryland Univ Baltimore Dept. of Computer Science and Electrical Engineering

Chen M, Lin H, Hu M et al (2013a) Real geographic scenario based virtual social environment: integrate geography with social research. Environ Plan B-Plan Design 40(6):1103–1121

Chen M, Lin H, Wen Y et al (2013b) Construction of a virtual lunar environment platform. Int J Digital Earth 6(5):469–482

Conde FC, Martínez SG, Ramos JL et al (2014) Building a 3D geomodel for water resources management: case study in the Regional Park of the lower courses of Manzanares and Jarama Rivers (Madrid, Spain). Environ Earth Sci 71(1):61–66

CSDMS Working Group (2004) Community surface dynamics modeling system (CSDMS) Science Plan [online]. Available from: http://csdms.colorado.edu/mediawiki/images/CSDMS_Science_Plan_Aug04.pdf. Accessed 3 May 2012

Davies Evan GR, Simonovic SP (2011) Global water resources modeling with an integrated model of the social–economic–environmental system. Adv Water Resour 34(6):684–700

Dennis JM, Vertenstein M, Worley PH et al (2012) Computational performance of ultra-high-resolution capability in the Community Earth System Model. Int J High Perform Comput Appl 26(1):5–16

Deshpande A, Guestrin C, Madden SR et al (2004) Model-driven data acquisition in sensor networks. In: Proceedings of the Thirtieth international conference on very large data bases-Volume 30. VLDB Endowment, pp 588–599

DeVantier BA, Feldman AD (1993) Review of GIS applications in hydrologic modeling. J Water Resour Plan Manag 119(2):246–261

Dickinson RE, Oleson KW, Bonan G et al (2006) The Community Land Model and its climate statistics as a component of the Community Climate System Model. J Clim 19(11):2302–2324

Golovko D, Roessner S, Behling R et al (2014) GIS-based integration of heterogeneous data for a multi-temporal landslide inventory. In: Landslide science for a safer geoenvironment. Springer International Publishing, Berlin, pp 799–804

Goodchild MF, Steyaert LT, Parks BO (eds) (1996) GIS and environmental modeling: progress and research issues. Wiley, USA

Gray J, Liu DT, Nieto-Santisteban M et al (2005) Scientific data management in the coming decade. ACM SIGMOD Record 34(4):34–41

Gruber TR (1995) Toward principles for the design of ontologies used for knowledge sharing? Int J Hum Comput Stud 43(5):907–928

Guariso G, Hitz M, Werthner H (1996) An integrated simulation and optimization modelling environment for decision support. Decis Support Syst 16(2):103–117

Gunay A, Akcay O, Altan MO (2014) Building a semantic based public transportation geoportal compliant with the INSPIRE transport network data theme. Earth Sci Inf 7(1):25–37

He HS, Larsen DR, Mladenoff DJ (2002) Exploring component-based approaches in forest landscape modeling. Environ Model Softw 17(6):519–529

Hill C, DeLuca C, Suarez M et al (2004) The architecture of the earth system modeling framework. Comput Sci Eng 6(1):18–28

Hillyer C, Bolte J, van Evert F et al (2003) The ModCom modular simulation system. Eur J Agron 18(3):333–343

Hudak PF (2014) Inundation patterns and plant growth in constructed wetland characterized by dynamic water budget model. Environ Earth Sci 72(6):1821–1826

Hwam WK, Chung Y, Kwon Y et al (2011) Conversion the time dependent grid data of NetCDF to SEDRIS Transmittal Format. In: 2011 International Conference on Information Technology and Multimedia (ICIM), IEEE, pp 1–5

Jabbar MT, Zhou JX (2013) Environmental degradation assessment in arid areas: a case study from Basra Province, southern Iraq. Environ Earth Sci 70(5):2203–2214

Jung CT, Sun CH (2010) Ontology-driven problem solving framework for spatial decision support systems. In: Proceedings, global spatial data handling. 12:19–22

Kalyanpur A, Hendler J, Parsia B et al (2006a) SMORE-semantic markup, ontology, and RDF editor. Maryland Univ College Park Dept. of Computer Science

Kalyanpur A, Parsia B, Sirin E et al (2006b) Swoop: a web ontology editing browser. Web Semant Sci Serv Agents World Wide Web 4(2):144–153

Knublauch H, Fergerson RW, Noy NF et al (2004) The Protégé OWL plugin: an open development environment for semantic web applications. The Semantic Web–ISWC 2004. Springer, Berlin Heidelberg, pp 229–243

Lake R (1999) Introduction to GML geography markup language. Retrieved 20 June 2003

Leavesley GH, Markstrom SL, Brewer MS et al (1996) The Modular Modeling System (MMS)—the physical process modeling component of a database-centered decision support system for water and power management. In: Clean water: factors that influence its availability, quality and its use. Springer, The Netherlands, pp 303–311

Leavesley GH, Markstrom SL, Restrepo PJ et al (2002) A modular approach to addressing model design, scale, and parameter estimation issues in distributed hydrological modelling. Hydrol Process 16(2):173–187

Leite MJ, Dobey V, Gravitz P (2006) SEDRIS: a tool for the management and exchange of multi-domain environmental data. In: Earth and Space 2006: Engineering, Construction, and Operations in Challenging Environment. ASCE, pp 1–8

Lin H, Lu GN, Song ZY et al (1997) Study on modeling the tide wave system of East China Sea with GIS. ACTA Geographic Sinica. vol 52. pp 161–169. CNKI: http://www.cnki.com.cn/Article/CJFDTOTAL-DLXB1997S1018.htm. Accessed 7 Jan 2015

Lin H, Zhu J, Gong J et al (2010) A grid-based collaborative virtual geographic environment for the planning of silt dam systems. Int J Geogr Inf Sci 24(4):607–621

Lin H, Chen M, Lu G et al (2013a) Virtual geographic environments (VGEs): a new generation of geographic analysis tool. Earth Sci Rev 126:74–84

Lin H, Chen M, Lu G (2013b) Virtual geographic environment: a workspace for computer-aided geographic experiments. Ann Assoc Am Geogr 103(3):465–482

Liu K, Yang C, Li W et al (2014) Using semantic search and knowledge reasoning to improve the discovery of earth science records: an example with the ESIP semantic testbed. Int J Appl Geospat Res (IJAGR) 5(2):44–58

López-De-Ipiña D, Almeida A, Aguilera U et al (2008) Dynamic discovery and semantic reasoning for next generation intelligent environments. http://digital-library.theiet.org/content/conferences/10.1049/cp_20081119. Accessed 26 Feb 2015

Lu GN (2011) Geographic analysis-oriented virtual geographic environment: framework, structure and functions. Sci China Earth Sci 54(5):733–743

Maedche A, Staab S (2001) Ontology learning for the semantic web. IEEE Intell Syst 16(2):72–79

Mailhot A, Rousseau AN, Massicotte S et al (1997) A watershed-based system for the integrated management of surface water quality: the GIBSI system. Water Sci Technol 36(5):381–387

Maxwell T, Costanza R (1997a) An open geographic modeling environment. Simulation 68(3):175–185

Maxwell T, Costanza R (1997b) A language for modular spatio-temporal simulation. Ecol Model 103(2):105–113

Moore RV, Tindall CI (2005) An overview of the open modelling interface and environment (the OpenMI). Environ Sci Policy 8(3):279–286

OGC (2003). http://www.opengeospatial.org/standards/is. Accessed 26 Feb 2015

Open GIS Consortium (2001) Inc. Open GIS specification-geography markup language (GML). Technical Report version 2.0 (01-029), OGC

Parsons MA (2011) Making data useful for modelers to understand complex Earth systems. Earth Sci Inf 4(4):197–223

Pascoe RT, Penny JP (1990) Construction of interfaces for the exchange of geographic data. Int J Geogr Inf Syst 4(2):147–156

Rahimi S, Roodposhti MS, Abbaspour RA (2014) Using combined AHP–genetic algorithm in artificial groundwater recharge site selection of Gareh Bygone Plain, Iran. Environ Earth Sci 72(6):1979–1992

Ramalho EC, Dill AC, Rocha R (2013) Assessment of the leachate movement in a sealed landfill using geophysical methods. Environ Earth Sci 68(2):343–354

Raskin R, Pan M (2003) Semantic web for earth and environmental terminology (sweet). In: Proc. of the Workshop on Semantic Web Technologies for Searching and Retrieving Scientific Data

Reed M, Cuddy SM, Rizzoli AE (1999) A framework for modelling multiple resource management issues—an open modelling approach. Environ Model Softw 14(6):503–509

Rich FJ, Newsom L, Meyer B et al (2014) Radiocarbon dates and the genesis of phytogenic near-shore sediments on St. Catherines Island, Georgia, USA. Environ Earth Sci 72(8):2985–2997

Rousseau AN, Mailhot A, Turcotte R et al (2000) GIBSI—an integrated modelling system prototype for river basin management. Springer, The Netherlands, pp 465–475

Sadat-Noori SM, Ebrahimi K, Liaghat AM (2014) Groundwater quality assessment using the Water Quality Index and GIS in Saveh-Nobaran aquifer, Iran. Environ Earth Sci 71(9):3827–3843

Selvam S, Manimaran G, Sivasubramanian P et al (2014) GIS-based Evaluation of Water Quality Index of groundwater resources around Tuticorin coastal city, south India. Environ Earth Sci 71(6):2847–2867

Serreze MC (2011) Climate change: rethinking the sea-ice tipping point. Nature 471(7336):47–48

Shao H, Liu M, Shao Q, Sun X et al (2014) Research on eco-environmental vulnerability evaluation of the Anning River Basin in the upper reaches of the Yangtze River. Environ Earth Sci 72(5):1555–1568

Simunich KL, Sydelko P, Dolph J et al (2002) Dynamic information architecture system (DIAS): multiple model simulation management. http://www.ipd.anl.gov/anlpubs/2002/05/43066.pdf. Accessed 26 Feb 2015

Sudo K, Takemura T, Klimont Z et al (2013) Global modeling and projection of short-lived climate pollutants in an Earth System Model. In: AGU Fall Meeting Abstracts. vol. 1, p 08

Todorova A, Syrakov D, Gadjhev G et al (2010) Grid computing for atmospheric composition studies in Bulgaria. Earth Sci Inf 3(4):259–282

Tung S, Leung JKC, Jiao JJ et al (2013) Assessment of soil radon potential in Hong Kong, China, using a 10-point evaluation system. Environ Earth Sci 68(3):679–689

Uyan M (2014) MSW landfill site selection by combining AHP with GIS for Konya, Turkey. Environ Earth Sci 71(4):1629–1639

Valcke S, Guilyardi E, Larsson C (2006) PRISM and ENES: A European approach to Earth system modelling. Concur Comput Pract Exp 18(2):247–262

Van Ittersum MK, Ewert F, Heckelei T et al (2008) Integrated assessment of agricultural systems—a component-based framework for the European Union (SEAMLESS). Agric Syst 96(1):150–165

Voinov A, Cerco C (2010) Model integration and the role of data. Environ Model Softw 25(8):965–969

Warner JC, Perlin N, Skyllingstad ED (2008) Using the Model Coupling Toolkit to couple earth system models. Environ Model Softw 23(10):1240–1249

Watson Fred GR, Rahman JM (2004) Tarsier: a practical software framework for model development, testing and deployment. Environ Model Softw 19(3):245–260

Weithöner T, Liebig T, Luther M et al (2007) Real-world reasoning with OWL. The semantic web: research and applications. Springer, Berlin Heidelberg, pp 296–310

Wen Y, Chen M, Lu G et al (2013) Prototyping an open environment for sharing geographical analysis models on cloud computing platform. Int J Digital Earth 6(4):356–382

Yal GP, Akgün H (2013) Landfill site selection and landfill liner design for Ankara, Turkey. Environ Earth Sci 70(6):2729–2752