



Review

From Manual to Intelligent: A Review of Input Data Preparation Methods for Geographic Modeling

Zhi-Wei Hou ^{1,2}, Cheng-Zhi Qin ^{1,2,3},*, A-Xing Zhu ^{1,2,3,4,5}, Peng Liang ^{1,2}, Yi-Jie Wang ^{1,2} and Yun-Qiang Zhu ^{1,2,3}

- State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China
- College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China
- Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China
- School of Geography, Nanjing Normal University, Nanjing 210023, China
- Department of Geography, University of Wisconsin-Madison, Madison, WI 53706, USA
- * Correspondence: qincz@lreis.ac.cn; Tel.: +86-010-6488-8959

Received: 16 July 2019; Accepted: 23 August 2019; Published: 28 August 2019



Abstract: One of the key concerns in geographic modeling is the preparation of input data that are sufficient and appropriate for models. This requires considerable time, effort, and expertise since geographic models and their application contexts are complex and diverse. Moreover, both data and data pre-processing tools are multi-source, heterogeneous, and sometimes unavailable for a specific application context. The traditional method of manually preparing input data cannot effectively support geographic modeling, especially for complex integrated models and non-expert users. Therefore, effective methods are urgently needed that are not only able to prepare appropriate input data for models but are also easy to use. In this review paper, we first analyze the factors that influence data preparation and discuss the three corresponding key tasks that should be accomplished when developing input data preparation methods for geographic models. Then, existing input data preparation methods for geographic models are discussed through classifying into three categories: manual, (semi-)automatic, and intelligent (i.e., not only (semi-)automatic but also adaptive to application context) methods. Supported by the adoption of knowledge representation and reasoning techniques, the state-of-the-art methods in this field point to intelligent input data preparation for geographic models, which includes knowledge-supported discovery and chaining of data pre-processing functionalities, knowledge-driven (semi-)automatic workflow building (or service composition in the context of geographic web services) of data preprocessing, and artificial intelligent planning-based service composition as well as their parameter-settings. Lastly, we discuss the challenges and future research directions from the following aspects: Sharing and reusing of model data and workflows, integration of data discovery and processing functionalities, task-oriented input data preparation methods, and construction of knowledge bases for geographic modeling, all assisting with the development of an easy-to-use geographic modeling environment with intelligent input data preparation.

Keywords: geographic modeling; input data preparation; intelligent geoprocessing; service composition

1. Introduction

Geographic modeling is a fundamental methodology for understanding, simulating, and predicting geographic phenomena and processes within a certain context [1–4]. A crucial step in geographic modeling is preparing input data for geographic models. Input data, including preliminary data or raw

input data (e.g., digital elevation model (DEM)) and derived information (e.g., topographic properties such as slope and area), are not only a prerequisite for model setup, calibration, and validation, but the quantity and quality also directly affect simulation results [1,2,5–8]. Insufficient and inappropriate input data (e.g., lack of observations and inappropriate DEM resolution) might limit both the accuracy of the model results and the applications of geographic models [9–12].

Input data preparation in geographic modeling is particularly challenging due to the input data needed by geographic models often being obtained from distributed data sources and being syntactically and semantically heterogeneous [1,5,6,13,14]. Modelers have to assess the quality, relevance, and suitability of preliminary data for geographic modeling. Then, they need to select and compose a set of applicable and compatible data pre-processing algorithms and their implementations, such as web services, to prepare needed input data. These data preparation steps often contain many operations that are repeated with a traditional manual method for most cases of geographic modeling. This means that considerable time, expertise, and effort are required to set up a new model application, which restricts the reproducibility of previous studies, particularly for those non-expert stakeholders (e.g., policymakers from local government) [5,6,13–16].

Integrated modeling environment (IME) has been proposed as an efficient and convenient tool for sharing, reusing, integrating, and running heterogeneous geographic models [1,14,17–19]. IMEs are shifting the application model of geographic modeling from centralized desktop software systems to distributed and service-oriented online geoprocessing platforms [20–22]. In addition, IMEs are increasingly using advanced computing technologies, such as parallel computing and cloud computing, to meet the computation requirements of large-scale and complex geographic models in the big data era [23–25].

However, most of the IME studies focused on developing new models, or/and sharing and coupling existing models and modules [1,14,22,26–28]. The input data preparation of geographic modeling in IMEs still heavily depends on modelers' modeling knowledge (including knowledge of the geographical domain, knowledge of geographic models and their input/output data, prior modeling experiences, and technical expertise). This situation not only reduces modeling efficiency and the applicability of IMEs, but might also lead to untrustworthy model results [11,12,29,30]. The situation is becoming unavoidable because geographic models are becoming increasingly complicated due to their trend of integrated multi-factor, multi-process, and multi-scale research [4,31]. Therefore, methods that can prepare appropriate input data for geographic models in a user-friendly and efficient way are urgently needed for IMEs.

To address problems related to input data preparation for geographic models, a variety of methods have been proposed. Based on artificial intelligence (AI) technologies such as ontology, logical reasoning, case-based reasoning (CBR), and AI planning, these methods aim to provide an automatic and intelligent way to discover data and the necessary pre-processing applications (e.g., web services) for geographic models [32–38]. Using these methods, the time, expertise, and prior experience requirements for preparing model input data can be reduced significantly to ensure the efficiency and effectiveness of geographic modeling.

In this paper, we conducted a systematic review of the state-of-the-art methods for preparing input data for geographic models and provide recommendations of areas for future study. The remainder of this paper is structured as follows: Section 2 provides an analysis of the factors that influence data preparation for geographic models, then Section 3 outlines the corresponding key tasks that should be accomplished. In Section 4, existing input data preparation methods are classified into three categories: Manual, (semi-)automatic, and intelligent (i.e., not only (semi-)automatic but also adaptive to application context) methods. Then, each of them is discussed according to their influencing factors and key tasks. Section 5 discusses future research directions of intelligent input data preparation methods for geographic models and their integration with IMEs. The last section provides a summary of this review.

2. Factors Influencing Input Data Preparation

The preparation of input data for geographic models is a procedure that typically requires the use of various data pre-processing tools (or software applications) to transform collected multi-source data into information required by geographic models. This procedure is mainly affected by three types of factors:

- 1. Diverse application contexts of a geographic model. In real applications, a geographic model might be used with diverse application contexts, which include the application purpose (or objective), data availability, spatial and temporal scale, and study area characteristics such as climate, topography, and soil [1,34,39–41]. A specific geographic model in different application contexts might need different data pre-processing workflows to meet its input data requirements [41–43]. For example, the optimal spatial resolution of SWAT (Soil and Water Assessment Tool) [44] modeling in a large mountainous watershed changes with the application purpose (e.g., simulation of flow, sediment, or dissolved oxygen) [45]. Thus, resampling or down-scaling for the input data sources with different resolutions thereby might need to be added into the input data preparation workflow. Moreover, among study areas with different characteristics (e.g., high or low relief), the same pre-processing step in the input data preparation workflow might adopt different algorithms. Another example is that, due to data unavailability (e.g., meteorological observations) in data-scarce regions, hydrological models might use satellite data products and corresponding data pre-processing functionalities [10].
- 2. Diverse characteristics of input data. Generally, geographic models require different types of input data, such as DEM, land cover, soil, and many others. These data are increasingly being obtained from geographically distributed data catalogs or geoportals that are established by cross-domain organizations and are heterogeneous in many aspects such as accessing method, metadata, data format, projection, and resolution [6,46–48]. Data quality and spatial/temporal scale are also key issues in geospatial data which have strong influences on the performance of geographic models [49–51]. Thus, to obtain sufficient, suitable, and ready-to-use input data for a geographic model, a set of appropriate functionalities (e.g., reformatting, reprojection, etc.) are needed to process the data to the forms required by the geographic model [6]. Therefore, modelers have to devote significant time and effort to familiarize themselves with the characteristics of the collected data.
- 3. Diversity of data pre-processing tools. A single tool is normally not suitable for all data processing tasks (e.g., clipping, reformatting, and reprojection) for a complex geographic model. Thus, modelers have to employ diverse tools (such as ArcGIS and Matlab) to manually or automatically pre-process the obtained data into ready-to-use forms for a model [5,52-54]. Normally, these tools developed by different organizations adopt different algorithms, run-time environments, application contexts, and input/output data types. For instance, a topographic attribute can be calculated from several algorithms (such as single- and multiple-flow direction algorithms for flow accumulation calculation). Each algorithm is proposed for specific data types (e.g., grid DEM), terrain conditions (e.g., high or low relief), spatial resolutions (e.g., coarse or fine), or application tasks (e.g., drainage network extraction or topographic wetness index calculation) [50,55–57]. Moreover, different algorithms might be implemented in different software and require different pre-processing steps (e.g., pit removing of DEM-preprocessing for flow direction algorithms). Consequently, finding and using appropriate data pre-processing tools to prepare input data for geographic models require considerable expertise and experience. Training time is long and the learning curve is steep for users who want to acquire such expertise and experience.

3. Key Tasks in Developing Input Data Preparation Methods for Geographic Models

Dealing with the aforementioned influencing factors requires considerable modeling knowledge, effort, and time for modelers (even for experienced modelers). When developing input data preparation

methods to free modelers from the burdens of preparing appropriate input data for geographic models, three key tasks need to be accomplished:

- 1. Integration of input data preparation tools for geographic models. Currently, except for a few widely used models like SWAT, most geographic models lack a tool to assist modelers to easily and efficiently prepare input data. Coupling related tools to an integrated data pre-processing environment provide a reasonable strategy to solve this problem. On the one hand, these environments reduce the number of tools used, thus reducing software setup and training time. On the other hand, applying widely used standards and specifications could improve the interoperability of coupled tools, which could facilitate data exchange among tools and avoid breaking the data pre-processing workflow [1,6,38,58].
- 2. Developing automatic input data preparation methods for geographic models. A common task in input data preparation for geographic models is repetitive data pre-processing workflows, such as watershed delineation and topographic wetness index (TWI) calculation, which chain a sequence of data processing tools to produce the desired outputs. Automating these sophisticated data pre-processing workflows in geographic modeling environments could allow modelers to concentrate on solving key problems instead of trivial technical details [6,38], and could also improve the reproducibility of existing studies [13,59].
- 3. Developing intelligent methods to support the automatic preparation of input data for geographic models in an application-context-adaptive way. As noted above, the application context of a geographic model strongly influences the selection of both preliminary input data (including data contents and characteristics, e.g., spatial resolution) and corresponding data pre-processing tools/algorithms (including parameter-settings). This requires extensive geographic modeling knowledge, which poses a challenge for modelers, especially for novices. Knowledge-driven intelligent input data preparation methods could overcome the problem and improve efficiency. These methods explicitly and meaningfully formalize and interlink geographic modeling knowledge, thereby reducing the semantic heterogeneity and improving the interoperability of modeling resources, including models, data, algorithms, and algorithm implementations such as software tools and web services. Both explicit and implicit relationships could be inferred through reasoning and semantic similarity calculation. As a consequence, intelligent input data preparation methods could not only automate the discovery and integration of data, models, and data pre-processing workflows, but also ensure the prepared input data match the application context [33,34,37,60].

4. Classification of Existing Input Data Preparation Methods for Geographic Models

To date, various methods have been proposed to address the above mentioned three key tasks. These input data preparation methods can be classified into three categories: Manual, (semi-)automatic, and intelligent (i.e., not only (semi-)automatic but also adaptive to application context) methods, as shown in Figure 1.

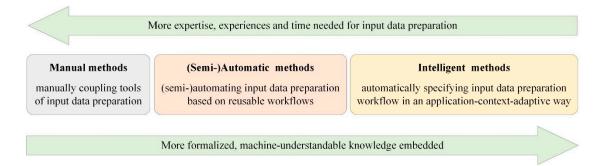


Figure 1. Three categories of input data preparation methods for geographic models.

4.1. Manual Methods

Manual input data preparation methods are methods where modelers manually prepare input data (including data discovery, data quality check, pre-processing functionality selecting, and workflow-building, etc.) for geographic models through human-machine interactive interfaces such as graphic menus, dialogues, or command-line utilities. This is currently the dominant method used in geographic modeling, for example, in distributed hydrological modeling [25].

To reduce data manipulations and simplify data transformations, software used for interactive input data preparation methods, such as ArcGIS and QSWAT [61], is often integrated with geographic models as modules or components. Generally, as depicted in Figure 2, there are four major coupling strategies for integrating data preparation software applications with the model program [62–64].

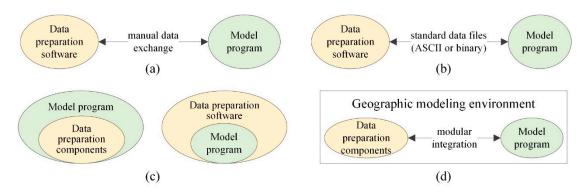


Figure 2. Coupling strategies for integrating data preparation tools with the model program: (a) Stand-alone, (b) loose coupling, (c) tight coupling, and (d) full integration.

The stand-alone strategy (Figure 2a) treats the model program and data preparation tools independently and data are exchanged manually through transformation functionalities. These preparation tools include standard geographic information system (GIS) software, e.g., GRASS GIS [65] and domain-specific analyzing tools, such as TauDEM [66], SimDTA [67], and HydroDesktop [48], in digital terrain analysis (DTA) or hydrological modeling domains.

In the loose coupling strategy (Figure 2b), data preparation tools are developed for a specific geographic model. They exchange data with the model program via both acceptable data formats but run separately without a common user interface. Examples include C-SWAT [68] for the hydrological model SWAT and SPELLmap for the SWATmf, which is a framework that integrates SWAT and the groundwater model MODFLOW [53,69].

The tight (or close) coupling strategy (Figure 2c) has been increasingly adopted in research in recent years [29,58,61,70,71]. This strategy involves embedding the model program into the data preparation system or vice versa via programming. The integrated system has a customized user interface to manage GIS data structures and generate input data files for the geographic model.

In the full integration strategy (Figure 2d), the model program and data preparation tools are coupled as modules or components of an IME, e.g., LIQUID[®] [72], and community surface dynamics modeling system (CSDMS) [5]. Such modules or components use the same data structure and share a common data management component and user interface. This not only facilitates data exchange and management but also reduces the complexity of the model setup process.

No matter which strategy is used, manual input data preparation methods are tedious and error-prone (even for experienced modelers), which prevents the reproducibility [13,38,59]. Modelers should be familiar with the data processing steps and the technical details of the used tools, which require a long time of training and practice. Moreover, during each setup of a model to run, modelers have to manually process a range of input data, which might include many repetitive steps.

4.2. (Semi-)Automatic Methods

Many input data pre-processing steps for geographic models could be (semi-)automated by composing the steps as workflows based on their functionalities and input-output data dependencies [38,73]. Such workflows can support scientists in documenting, sharing, and executing a series of data processing steps [74,75]. Workflows also lower the barriers to promote the efficiency of input data preparation. Currently, according to the method of automating the workflow building process (Figure 3), (semi-)automatic methods can be divided into three categories.

- Automatic input data preparation based on hard-coded workflows (Figure 3(a)). This category of
 methods embeds stable data processing workflows into data preparation tools via programming,
 such as the watershed delineation workflows hard-coded in ArcSWAT [76] and HydroTerre [77].
 This way avoids wasteful repetitive efforts and smooths the learning curve for novices, but it is
 costly and difficult to develop these automatic methods. Moreover, automatic methods based on
 hard-coded workflows are "black-boxes", meaning modelers can neither directly understand
 how these methods work nor adjust them for specific application contexts.
- 2. Script-based (semi-)automatic input data preparation (Figure 3(b)). Workflows used to create input data for geographic models can be complicated and require multiple data processing functionalities from different tools (e.g., MATLAB, ArcGIS, and Python packages). This category of methods uses editable scripts (including rule files that control the execution order) to link the required functionality to (semi-)automatic data preparation workflows [29,38,59,78–80]. This way is more flexible and extensible than the hard-code methods. As such, modelers can modify or add new scripts to customize the workflows according to the application contexts. However, this category of (semi-)automatic methods requires extensive user technical expertise and modeling experience.
- 3. Graphic workflow building environment for input data preparation (Figure 3(c)). This method involves the use of graphic modeling panel to assist users to visually and quickly build or reuse, revise, and configure workflows. The generated workflows can then be executed to prepare input data automatically. In recent years, service-oriented graphic workflow building environments, such as Giovanni [81], GeoJModelBuilder [82], and CyberConnector [6,73], have attracted attention. Web services facilitate the sharing, reuse, and coupling of data processing functionalities, workflows, and computing resources. Therefore, this method not only lowers the barriers to build workflows for input data preparation, but also promotes the collaboration of modelers from different disciplines [21,37,83]. However, manually building these input data preparation workflows for novice users is still difficult and laborious. In addition, this category of automatic methods cannot ensure that the generated workflows and configured parameters match the application contexts.

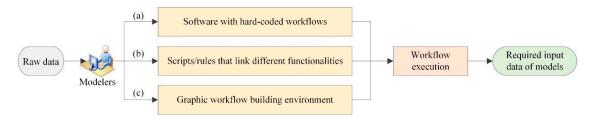


Figure 3. Three ways to automate the workflow building process.

4.3. Intelligent Methods

Knowledge base (KB) and reasoning are two principal aspects of intelligent systems [84]. Intelligent methods for input data preparation for geographic models use advanced AI technologies to build intelligent (i.e., adaptive to application context) input data preparation systems that can address the

problems with manual and the automatic methods presented above. In intelligent methods, semantic web [85] technologies, such as resources description framework (RDF) and ontology, are often used to represent the semantics of modeling resources in unambiguous and machine-understandable forms. Meanwhile, many other AI technologies (e.g., logical reasoning, CBR, and AI planning) have been used to infer implicit relationships and calculate semantic similarities [1,35–37,60,86]. Using these technologies, intelligent methods can improve the interoperability of modeling resources and facilitate on-demand discovery, selection, chaining, and validation.

At present, intelligent input data preparation methods can be divided into two sub-categories: Intelligent building of input data preparation workflow, and the intelligent parameter setting of data processing algorithms. The existing methods of the first sub-category, i.e., the intelligent building of input data preparation workflow, can be classified into the three types as shown in Figure 4.

- 1. Knowledge-supported interactive workflow building (Figure 4a). This type of intelligent building method of the input data preparation workflow requires users to manually build data processing workflows in a graphic workflow building environment with the support of a knowledge base and reasoning. For this type of method, modeling resources are semantically enriched, inter-linked, and published as ontologies and/or linked data [35,86–89], which is different from the automatic methods based on graphic workflow building environments mentioned in Section 4.2. Workflow building knowledge (including tacit experience in existing application cases, relationships between tasks and data, algorithms, and reusable workflows) can also be formalized as CBR cases or ontologies [39,90]. Through querying and reasoning this formalized knowledge, the modeling environment adopting this type of method can assist users in discovering and composing appropriate functionalities to build, validate, and correct or optimize input data preparation workflows [33,39,90,91]. The main problem with the knowledge-supported interactive workflow building is the lack of automation during workflow building.
- Knowledge-driven (semi-)automatic workflow building (Figure 4b). This type of intelligent 2 workflow building methods of input data preparation can (semi-)automatically discover and compose the needed data processing algorithms (or web services) based on semantic matching and reasoning. For example, the heuristic modeling proposed by Jiang, et al. [92] adopts RDF, heuristic modeling, and backward chaining approaches to semi-automatically build abstract workflows. The method starts by selecting an algorithm that can generate outputs matching user-specified target data (i.e., input data for the users' geographic model). Then, for the selected algorithm, users either set its inputs or invoke the system to automatically expand the workflow by adding other algorithms that can generate the required input data. This procedure is repeated until all the input data of the workflow are available. Besides this heuristic modeling proposed by Jiang et al. [92], other researchers have used ontologies, logical reasoning, and forward-chaining or backward-chaining approaches to automatically discover and compose the required services according to users' requests [32,93,94]. The match between users' requests and inputs, outputs, preconditions, and effects/postconditions (IOPE) semantics of web services is based on semantic matching and logical reasoning, such as description logic (DL) reasoning, first-order logic (FOL) reasoning, and rule-based reasoning. The major limitation of knowledge-driven (semi-)automatic workflow building methods is that they cannot ensure the semantic correctness and suitability of the workflow in a specific application context because, except for IO or IOPE, many semantics of a service (e.g., functionality, applicable application contexts, and constraints of data types or formats) are ignored when describing or composing web services.
- 3. Automatic web service composition based on AI planning (Figure 4c). This type of intelligent workflow building method for input data preparation views semantic web services (i.e., services that are semantically annotated using ontologies, e.g., Web Ontology Language for Services (OWL-S) [95] and Web Service Modelling Ontology (WSMO) [96]) as actions, and treats service IOPE semantics as states. Then web service composition becomes a planning problem. To solve this planning problem, AI planning algorithms, for instance, the hierarchical task network (HTN),

can be used to find a sequence of actions (i.e., a plan) to change the initial state satisfying the pre-defined goal state (i.e., desired input data for geographic models) [97–102]. As a result, modelers could use an AI planner together with an ontology inference engine to create plans and translate them to executable service chains for preparing input data for geographic models. The generated web service chains could be optimized based on the quality of service (QoS) using other AI algorithms such as the genetic algorithm and game theory [103,104]. Note that this type of methods faces similar problems as the knowledge-driven (semi-)automatic workflow building methods presented above.

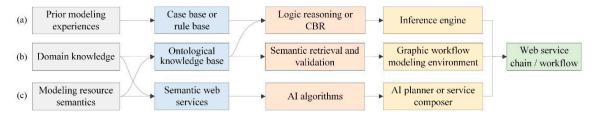


Figure 4. Intelligent building of input data preparation workflow: (a) Knowledge-supported interactive workflow building, (b) knowledge-driven (semi-)automatic workflow building, and (c) automatic web service composition based on AI planning.

The second sub-category of intelligent input data preparation methods is the intelligent parameter setting of data processing algorithms in workflow-building. Parameter setting plays a vital role in the application of algorithms because an inappropriate parameter value will lead to inaccurate results. It requires considerable experience and expertise to set parameter values according to the application context. Some parameters (e.g., the catchment area threshold in drainage network extraction) are empirical and their values should vary with the application context such as landforms and spatial scales. To address this problem, a case-based method has been proposed to automatically set parameter values in digital terrain analysis according to application contexts [34]. As shown in Figure 5, this method first creates a case base by formalizing previous application cases that contain empirical knowledge of parameter settings of algorithms. Then, the case-based method calculates the similarity of application contexts between the new-coming application problem formalized as a case (without solution) and each case (with solution) in the case base to retrieve the most similar case from the case base. Consequently, the application-context-matching parameter value can be automatically recommended. This case-based method can reduce the burden on the users caused by time-intensive learning and try-and-error (especially that of non-expert users). Currently, there are two major issues in this method: Determining how to automatically build such a large-scale case base, and how to expand this method to other application domains.

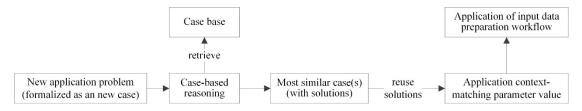


Figure 5. Case-based method for parameter settings of digital terrain analysis algorithm, revised from Qin et al. [34].

5. Future Research Directions

Although many methods have been proposed to improve the efficiency and accuracy of input data preparations and minimize the requirement for extensive modeler expertise, the three key tasks

presented in Section 2 are still far from being accomplished. As geographic models are becoming increasingly complicated through integrating sub-models from diverse domains [4,12,31], input data preparation now requires more time, modeling knowledge, and technical expertise than ever. Increasing numbers of cross-domain stakeholders are engaged in geographic modeling [1,14,25,105]. The need for user-friendly and intelligent input data preparation methods and tools is becoming increasingly urgent.

To fill the gap between the existing methods and the requirement for a highly intelligent and easy-to-use input data preparation environment, knowledge-driven and service-oriented methods for IMEs must be developed. These methods should be able to use domain knowledge and prior experience to solve new modeling problems, automatically discover and pre-process (or reuse) application-context-matching input data for geographic models from distributed data sources and report the uncertainty of the automatically recommended solutions. They will make IMEs easier to use and will be more effective for modelers.

To this end, we recommend the following research priorities in input data preparation for geographic models:

- 1. Publishing, sharing, and reusing model data and data pre-processing workflows. Data involved in geographic modeling can be classified into four types: Preliminary data, intermediate data (processing results used by subsequent steps), prepared input data, and simulation results. Publishing, sharing, and reusing these data and the corresponding workflows could avoid repetitive work in the data pre-processing steps for preparing input data, thus reducing errors, and supporting collaboration and reproducibility. This has been demonstrated by several hydrological model data sharing platforms [106–109] and workflow building environments [73,82]. Whereas a unified, semantically rich, and machine-understandable metadata framework to publish model data and workflows is still lacking. Thus, it is difficult to efficiently discover and reuse multi-source, heterogeneous data and workflows. In addition, due to current sharing platforms being isolated from IMEs, a considerable amount of manual work is required to exchange data between these platforms and IMEs. To solve these problems, web service and semantic web technologies could be used to reduce syntactic and semantic heterogeneities between the data of these platforms and IMEs.
- 2. Integrating both data discovery and processing functionalities into IMEs. As mentioned in Section 2, the integration of data processing functionalities and the geographic model program in IMEs have been extensively researched. However, modelers still have to discover and process input data for geographic models separately. This means that the model input data acquired from data discovery tools, or directly from distributed spatial data infrastructures (SDIs), have to be manually transferred to input data pre-processing tools or IMEs. This procedure is tedious and needs the users to have specialized SDI knowledge (such as metadata standards, protocols, and domain terminologies) and data pre-processing functionalities [5,46]. Recently, integrated geospatial analysis platforms, such as HydroDesktop [48], Google Earth Engine (GEE) [110], and the Joint Research Centre Earth Observation Data and Processing Platform (JEODPP) [111], have attracted increasing interest. They enable users to discover, process, analyze, and visualize the needed data in one platform. Unfortunately, the data discovery and process steps in these platforms have not yet been automated and have not been integrated with IMEs, which means that data have to be exchanged manually. Therefore, integrating both data discovery and processing functionalities into IMEs should be researched in the future.
- 3. Developing task-oriented input data preparation methods. Geographic modeling is inherently task-driven work. These tasks of solving geographic problems are highly dependent on the conceptual knowledge of geographic problem-solving and technical expertise in terms of geographic models, data, data pre-processing tools (including parameter-settings), and workflows. Users can easily understand and express tasks instead of specialized domain knowledge, study area characteristics, and technical details of geographic modeling [112,113]. Recent studies have proposed several task-oriented geospatial data retrieval or processing methods [90,112–115].

- However, these methods are still difficult to use in geographic modeling due to the lack of automation driven by specific task knowledge of geographic modeling, especially of input data preparation.
- 4. Constructing large-scale, high-quality knowledge bases for intelligent geographic modeling. The quantity and quality of formalized geographic modeling knowledge determine the level of automation and intelligence of input data preparation methods and corresponding IMEs [33,35,116]. Currently, a large amount of knowledge on geographic modeling in different domains has not yet been formalized, for example, the knowledge of geoprocessing functionalities, and domain concepts and algorithms of digital terrain analysis [34]. Knowledge fusion and refinement are also urgently needed to alleviate problems of incompleteness, incorrectness, redundancy, and heterogeneity in knowledge bases [116–120]. In addition, few studies have been conducted to address the issue of the representation and reasoning of application-context knowledge [34,39]. Therefore, determining how to construct large-scale and high-quality knowledge bases for intelligent modeling is a key problem in future research. To build these knowledge bases, advanced technologies, such as machine learning, natural language processing, and knowledge graph [121], could be explored to extract, represent, and use the cross-domain modeling knowledge.

6. Summary

Input data preparation for geographic models has been increasingly recognized as a vital step in geographic modeling. An easy-to-use, efficient, and intelligent input data preparation method could not only free modelers from the burden of repetitive work and extensive training but also improve the accuracy of the model results.

We first analyzed factors influencing input data preparation for geographic models, and the corresponding three key tasks that need to be accomplished when developing input data preparation methods. Then, we divided existing input data preparation methods into three categories: Manual methods, (semi-)automatic methods, and intelligent methods. Based on a survey of the state-of-the-art methods, we determined that knowledge-driven intelligent input data preparation for geographic models is the most promising yet challenging research subject. It is still seldom implemented in practical systems. This limits the IMEs' ability to improve the modeling efficiency and to ensure the suitability of model inputs to the application context. Therefore, we discussed four future research directions to improve this situation. With the support of advanced technologies and methods such as web service, semantic web, and AI, input data preparation methods, as well as geographic modeling with IMEs, are entering the era of intelligence. The improvements in these research directions will enable modelers, whether they are domain experts or novices, to easily and effectively prepare sufficient and application-matching input data for geographic models.

Author Contributions: All authors gave substantial contributions to this work. Conceptualization was conducted by all listed authors. Formal analysis and investigation were conducted by Zhi-Wei Hou. Writing—original draft preparation was conducted by Zhi-Wei Hou. Writing—review, and editing were conducted by all authors. Supervision, project administration and funding acquisition were conducted by Cheng-Zhi Qin.

Funding: This research was funded by National Natural Science Foundation of China (No. 41431177, 41422109), the Innovation Project of LREIS (No. O88RA20CYA). Supports to A-Xing Zhu through the Vilas Associate Award, the Hammel Faculty Fellow Award, and the Manasse Chair Professorship from the University of Wisconsin-Madison are greatly appreciated.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Laniak, G.F.; Olchin, G.; Goodall, J.; Voinov, A.; Hill, M.; Glynn, P.; Whelan, G.; Geller, G.; Quinn, N.; Blind, M.; et al. Integrated environmental modeling: A vision and roadmap for the future. *Environ. Model. Softw.* **2013**, 39, 3–23. [CrossRef]
- 2. Jakeman, A.; Letcher, R.; Norton, J.; Jakeman, A. Ten iterative steps in development and evaluation of environmental models. *Environ. Model. Softw.* **2006**, *21*, 602–614. [CrossRef]
- 3. Yue, S.; Chen, M.; Wen, Y.; Lu, G. Service-oriented model-encapsulation strategy for sharing and integrating heterogeneous geo-analysis models in an open web environment. *ISPRS J. Photogramm. Remote Sens.* **2016**, 114, 258–273. [CrossRef]
- 4. Peng, S.; Piao, S.; Yu, J.; Liu, Y.; Wang, T.; Zhu, G.; Dong, J.; Miao, C. A review of geographical system models. *Prog. Geo* **2018**, *37*, 109–120. (In Chinese with English abstract).
- 5. Peckham, S.D.; Goodall, J.L. Driving plug-and-play models with data from web services: A demonstration of interoperability between CSDMS and CUAHSI-HIS. *Comput. Geosci.* **2013**, *53*, 154–161. [CrossRef]
- Di, L.; Sun, Z.; Yu, E.; Song, J.; Tong, D.; Huang, H.; Wu, X.; Domenico, B. Coupling of Earth Science Models and Earth Observations through OGC Interoperability Specifications. In Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July 2016; pp. 3602–3605.
- 7. Parsons, M.A. Making data useful for modelers to understand complex Earth systems. *Earth Sci. Inform.* **2011**, *4*, 197–223. [CrossRef]
- 8. Zhu, Y.; Zhu, A.X.; Feng, M.; Song, J.; Zhao, H.; Yang, J.; Zhang, Q.; Sun, K.; Zhang, J.; Yao, L. A similarity-based automatic data recommendation approach for geographic models. *Int. J. Geogr. Inf. Sci.* **2017**, *2*, 1–22. [CrossRef]
- 9. Tan, M.L.; Ficklin, D.L.; Dixon, B.; Ibrahim, A.L.; Yusop, Z.; Chaplot, V. Impacts of DEM resolution, source, and resampling technique on SWAT-simulated streamflow. *Appl. Geogr.* **2015**, *63*, 357–368. [CrossRef]
- 10. Cai, M.; Yang, S.; Zhao, C.; Zhou, Q.; Hou, L. Insight into runoff characteristics using hydrological modeling in the data-scarce southern Tibetan Plateau: Past, present, and future. *PLoS ONE* **2017**, *12*, 0176813. [CrossRef]
- 11. Liu, C.; Bai, P.; Wang, Z.; Liu, S.; Liu, X. Study on prediction of ungaged basins: A case study on the Tibetan Plateau. *J. Hydraul. Eng.* **2016**, *47*, 272–282. (In Chinese with English abstract).
- 12. Fatichi, S.; Vivoni, E.R.; Ogden, F.L.; Ivanov, V.Y.; Mirus, B.; Gochis, D.; Downer, C.W.; Camporese, M.; Davison, J.H.; Ebel, B.; et al. An overview of current applications, challenges, and future trends in distributed process-based models in hydrology. *J. Hydrol.* **2016**, *537*, 45–60. [CrossRef]
- 13. Morsy, M.M.; Goodall, J.L.; Castronova, A.M.; Dash, P.; Merwade, V.; Sadler, J.M.; Rajib, M.A.; Horsburgh, J.S.; Tarboton, D.G. Design of a metadata framework for environmental models with an example hydrologic application in HydroShare. *Environ. Model. Softw.* **2017**, *93*, 13–28. [CrossRef]
- 14. Lü, G.N. Geographic analysis-oriented Virtual Geographic Environment: Framework, structure and functions. *Sci. China Earth Sci.* **2011**, *54*, 733–743. [CrossRef]
- 15. Horsburgh, J.S.; Reeder, S.L. Data visualization and analysis within a Hydrologic Information System: Integrating with the R statistical computing environment. *Environ. Model. Softw.* **2014**, *52*, 51–61. [CrossRef]
- 16. Voinov, A.; Jenni, K.; Gray, S.; Kolagani, N.; Glynn, P.D.; Bommel, P.; Prell, C.; Zellner, M.; Paolisso, M.; Jordan, R.; et al. Tools and methods in participatory modeling: Selecting the right tool for the job. *Environ. Model. Softw.* **2018**, 109, 232–255. [CrossRef]
- 17. Cheng, G.; Li, X.; Zhao, W.; Xu, Z.; Feng, Q.; Xiao, S.; Xiao, H. Integrated study of the water–ecosystem–economy in the Heihe River Basin. *Natl. Sci. Rev.* **2014**, *1*, 413–428. [CrossRef]
- 18. Argent, R.M. An overview of model integration for environmental applications—Components, frameworks and semantics. *Environ. Model. Softw.* **2004**, *19*, 219–234. [CrossRef]
- 19. Granell, C.; Schade, S.; Ostländer, N. Seeing the forest through the trees: A review of integrated environmental modelling tools. *Comput. Environ. Urban Syst.* **2013**, *41*, 136–150. [CrossRef]
- 20. Goodall, J.; Horsburgh, J.; Whiteaker, T.; Maidment, D.; Zaslavsky, I. A first approach to web services for the National Water Information System. *Environ. Model. Softw.* **2008**, 23, 404–411. [CrossRef]
- 21. Hofer, B. Uses of online geoprocessing technology in analyses and case studies: A systematic analysis of literature. *Int. J. Digit. Earth* **2015**, *8*, 901–917. [CrossRef]

- 22. Nativi, S.; Mazzetti, P.; Geller, G.N. Environmental model access and interoperability: The GEO Model Web initiative. *Environ. Model. Softw.* **2013**, *39*, 214–228. [CrossRef]
- 23. Yang, C.; Huang, Q.; Li, Z.; Liu, K.; Hu, F. Big Data and cloud computing: Innovation opportunities and challenges. *Int. J. Digit. Earth* **2017**, *10*, 13–53. [CrossRef]
- 24. Wang, S.; Anselin, L.; Bhaduri, B.; Crosby, C.; Goodchild, M.F.; Liu, Y.Y.; Nyerges, T.L. CyberGIS software: A synthetic review and integration roadmap. *Int. J. Geogr. Inf. Sci.* **2013**, 27, 2122–2145. [CrossRef]
- 25. Jiang, J.; Zhu, A.X.; Qin, C.; Liu, J.; Chen, L.; Wu, H. Review on distributed hydrological modelling software systems. *Prog. Geo* **2014**, *33*, 1090–1100. (In Chinese with English abstract).
- 26. Moore, R.V.; Tindall, C.I. An overview of the open modelling interface and environment (the OpenMI). *Environ. Sci. Policy* **2005**, *8*, 279–286. [CrossRef]
- 27. Belete, G.F.; Voinov, A.; Laniak, G.F. An overview of the model integration process: From pre-integration assessment to testing. *Environ. Model. Softw.* **2017**, *87*, 49–63. [CrossRef]
- 28. Peckham, S.D.; Hutton, E.W.; Norris, B. A component-based approach to integrated modeling in the geosciences: The design of CSDMS. *Comput. Geosci.* 2013, 53, 3–12. [CrossRef]
- 29. Wi, S.; Ray, P.; DeMaria, E.M.; Steinschneider, S.; Brown, C. A user-friendly software package for VIC hydrologic model development. *Environ. Model. Softw.* **2017**, *98*, 35–53. [CrossRef]
- 30. Guillera-Arroita, G.; Lahoz-Monfort, J.J.; Elith, J.; Gordon, A.; Kujala, H.; Lentini, P.E.; McCarthy, M.A.; Tingley, R.; Wintle, B.A.; Guillera-Arroita, G.; et al. Is my species distribution model fit for purpose? Matching data and models to applications. *Glob. Ecol. Biogeogr.* **2015**, *24*, 276–292. [CrossRef]
- 31. Fu, B. Thoughts on the recent development of physical geography. *Prog. Geo* **2018**, *37*, 1–7. (In Chinese with English abstract).
- 32. Lutz, M. Ontology-based descriptions for semantic discovery and composition of geoprocessing services. *GeoInformatica* **2007**, *11*, 1–36. [CrossRef]
- 33. Hofer, B.; Mäs, S.; Brauner, J.; Bernard, L. Towards a knowledge base to support geoprocessing workflow development. *Int. J. Geogr. Inf. Sci.* **2016**, *31*, 1–23. [CrossRef]
- 34. Qin, C.Z.; Wu, X.W.; Jiang, J.C.; Zhu, A.X. Case-based formalization and reasoning method for knowledge in digital terrain analysis Illustrated by determining the catchment area threshold for extracting drainage networks. *Hydrol. Earth Syst. Sci. Discuss.* **2016**, 20, 1–40. [CrossRef]
- 35. Di, L.; Zhao, P.; Yang, W.; Yue, P. Ontology-Driven Automatic Geospatial-Processing Modeling Based on Web-Service Chaining. In Proceedings of the Sixth Annual NASA Earth Science Technology Conference, College Park, MD, USA, 27–29 June 2006; pp. 27–29.
- 36. Qiu, L.; Du, Z.; Zhu, Q.; Fan, Y. An integrated flood management system based on linking environmental models and disaster-related data. *Environ. Model. Softw.* **2017**, *91*, 111–126. [CrossRef]
- 37. Yue, P.; Baumann, P.; Bugbee, K.; Jiang, L. Towards intelligent GIServices. *Earth Sci. Inform.* **2015**, *8*, 463–481. [CrossRef]
- 38. Billah, M.M.; Goodall, J.L.; Narayan, U.; Essawy, B.T.; Lakshmi, V.; Rajasekar, A.; Moore, R.W. Using a data grid to automate data preparation pipelines required for regional-scale hydrologic modeling. *Environ. Model. Softw.* **2016**, *78*, 31–39. [CrossRef]
- 39. Lu, Y.; Qin, C.Z.; Zhu, A.X.; Qiu, W. Application-Matching Knowledge Based Engine for a Modelling Environment for Digital Terrain Analysis. In Proceedings of the GeoInformatics, Hong Kong, China, 15–17 June 2012.
- 40. Schmolke, A.; Thorbek, P.; DeAngelis, D.L.; Grimm, V. Ecological models supporting environmental decision making: A strategy for the future. *Trends Ecol. Evol.* **2010**, 25, 479–486. [CrossRef]
- 41. Fenicia, F.; Kavetski, D.; Savenije, H.H.G. Elements of a flexible approach for conceptual hydrological modeling: Motivation and theoretical development. *Water Resour. Res.* **2011**, *47*, W11510. [CrossRef]
- 42. Voinov, A.; Fitz, C.; Boumans, R.; Costanza, R. Modular ecosystem modeling. *Environ. Modell. Softw.* **2004**, 19, 285–304. [CrossRef]
- 43. Xu, Z. Hydrological models: Past, present and future. *J. Beijing Norm. Univ. Nat. Sci.* **2010**, *46*, 278–289. (In Chinese with English abstract).
- 44. Arnold, J.G.; Moriasi, D.N.; Gassman, P.W.; Abbaspour, K.C.; White, M.J.; Srinivasan, R.; Santhi, C.; Harmel, R.D.; Van Griensven, A.; Van Liew, M.W.; et al. SWAT: Model Use, Calibration, and Validation. *Trans. ASABE* **2012**, *55*, 1491–1508. [CrossRef]

- 45. Zhang, P.; Liu, R.; Bao, Y.; Wang, J.; Yu, W.; Shen, Z. Uncertainty of SWAT model at different DEM resolutions in a large mountainous watershed. *Water Res.* **2014**, *53*, 132–144. [CrossRef] [PubMed]
- 46. Gui, Z.; Yang, C.; Xia, J.; Liu, K.; Xu, C.; Li, J.; Lostritto, P. A performance, semantic and service quality-enhanced distributed search engine for improving geospatial resource discovery. *Int. J. Geogr. Inf. Sci.* 2013, 27, 1109–1132. [CrossRef]
- 47. Peckham, S.D. The CSDMS Standard Names: Cross-Domain Naming Conventions for Describing Process Models, Data Sets and Their Associated Variables. In Proceedings of the International Environmental Modelling and Software Society, 7th International Congress on Environmental Modeling and Software, San Diego, CA, USA, 15–19 June 2014.
- 48. Ames, D.P.; Horsburgh, J.S.; Cao, Y.; Kadlec, J.; Whiteaker, T.; Valentine, D. HydroDesktop: Web services-based software for hydrologic data discovery, download, visualization, and analysis. *Environ. Model. Softw.* **2012**, 37, 146–156. [CrossRef]
- 49. Goodchild, M.F. Scale in GIS: An overview. Geomorphology 2011, 130, 5–9. [CrossRef]
- 50. Tang, G. Progress of DEM and digital terrain analysis in China. *Acta Geogr. Sin.* **2014**, *69*, 1305–1325. (In Chinese with English abstract).
- 51. Servigne, S.; Ubeda, T.; Puricelli, A.; Laurini, R. A Methodology for Spatial Consistency Improvement of Geographic Databases. *GeoInformatica* **2000**, *4*, 7–34. [CrossRef]
- 52. Khan, K.A.; Akhter, G.; Ahmad, Z. OIL—Output input language for data connectivity between geoscientific software applications. *Comput. Geosci.* **2010**, *36*, 687–697. [CrossRef]
- 53. Guzman, A.J.; Moriasi, D.N.; Chu, M.L.; Starks, P.; Steiner, J.; Gowda, P. A tool for mapping and spatio-temporal analysis of hydrological data. *Environ. Model. Softw.* **2013**, *48*, 163–170. [CrossRef]
- 54. Granell, C.; Diaz, L.; Gould, M. Service-oriented applications for environmental models: Reusable geospatial services. *Environ. Model. Softw.* **2010**, *25*, 182–198. [CrossRef]
- 55. Qin, C.Z.; Zhu, A.; Pei, T.; Li, B.; Zhou, C.; Yang, L. An adaptive approach to selecting a flow-partition exponent for a multiple-flow-direction algorithm. *Int. J. Geogr. Inf. Sci.* **2007**, *21*, 443–458. [CrossRef]
- 56. Wilson, J.P. Digital terrain modeling. Geomorphology 2012, 137, 107–121. [CrossRef]
- 57. Hengl, T.; Reuter, H.I. (Eds.) *Geomorphometry: Concepts, Software, Applications*; Elsevier: Amsterdam, The Netherlands, 2008; p. 772.
- 58. Rossetto, R.; De Filippis, G.; Borsi, I.; Foglia, L.; Cannata, M.; Criollo, R.; Vázquez-Suñé, E. Integrating free and open source tools and distributed modelling codes in GIS environment for data-based groundwater management. *Environ. Model. Softw.* **2018**, *107*, 210–230. [CrossRef]
- 59. Gardner, M.A.; Morton, C.G.; Huntington, J.L.; Niswonger, R.G.; Henson, W.R. Input data processing tools for the integrated hydrologic model GSFLOW. *Environ. Model. Softw.* **2018**, *109*, 41–53. [CrossRef]
- 60. Villa, F.; Athanasiadis, I.N.; Rizzoli, A.E. Modelling with knowledge: A review of emerging semantic approaches to environmental modelling. *Environ. Model. Softw.* **2009**, 24, 577–587. [CrossRef]
- 61. Dile, Y.T.; Daggupati, P.; George, C.; Srinivasan, R.; Arnold, J. Introducing a new open source GIS user interface for the SWAT model. *Environ. Model. Softw.* **2016**, *85*, 129–138. [CrossRef]
- 62. Goodchild, M.; Haining, R.; Wise, S. Integrating GIS and spatial data analysis: Problems and possibilities. *Int. J. Geogr. Inf. Syst.* **1992**, *6*, 407–423. [CrossRef]
- 63. Sui, D.; Maggio, R. Integrating GIS with hydrological modeling: Practices, problems, and prospects. *Comput. Environ. Urban Syst.* **1999**, 23, 33–51. [CrossRef]
- 64. Nyerges, T. Coupling GIS and Spatial Analytic Models. In Proceedings of the 5th International Symposium on Spatial Data Handling, San Fransisco, CA, USA, 12–15 May 1992; pp. 534–543.
- 65. Neteler, M.; Bowman, M.H.; Landa, M.; Metz, M.; Bowman, M. GRASS GIS: A multi-purpose open source GIS. *Environ. Model. Softw.* **2012**, *31*, 124–130. [CrossRef]
- 66. Tarboton, D.G. *Terrain Analysis Using Digital Elevation Models (TauDEM)*; Utah State Universityp: Logan, UT, USA, 2005.
- 67. Qin, C.; Lu, Y.; Bao, L.; Zhu, A.; Qiu, W.; Cheng, W. Simple Digital Terrain Analysis Software (SimDTA 1.0) and Its Application in Fuzzy Classification of Slope Positions. *J. Geo Inf. Sci.* **2009**, *11*, 737–743. (In Chinese with English abstract). [CrossRef]
- 68. Yen, H.; Ahmadi, M.; White, M.J.; Wang, X.; Arnold, J.G. C-SWAT: The Soil and Water Assessment Tool with consolidated input files in alleviating computational burden of recursive simulations. *Comput. Geosci.* **2014**, 72, 221–232. [CrossRef]

- 69. Guzman, J.; Moriasi, D.; Gowda, P.; Steiner, J.; Starks, P.; Arnold, J.; Srinivasan, R. A model integration framework for linking SWAT and MODFLOW. *Environ. Model. Softw.* **2015**, *73*, 103–116. [CrossRef]
- 70. Bhatt, G.; Kumar, M.; Duffy, C.J. A tightly coupled GIS and distributed hydrologic modeling framework. *Environ. Model. Softw.* **2014**, *62*, 70–84. [CrossRef]
- 71. Lewis, E.; Birkinshaw, S.; Kilsby, C.; Fowler, H.J. Development of a system for automated setup of a physically-based, spatially-distributed hydrological model for catchments in Great Britain. *Environ. Model. Softw.* **2018**, *108*, 102–110. [CrossRef]
- 72. Branger, F.; Braud, I.; Debionne, S.; Viallet, P.; Dehotin, J.; Henine, H.; Nedelec, Y.; Anquetin, S. Towards multi-scale integrated hydrological models using the LIQUID® framework. Overview of the concepts and first application examples. *Environ. Model. Softw.* **2010**, 25, 1672–1681. [CrossRef]
- 73. Sun, Z.; Di, L.; Hao, H.; Wu, X.; Tong, D.Q.; Zhang, C.; Virgei, C.; Fang, H.; Yu, E.; Tan, X. CyberConnector: A service-oriented system for automatically tailoring multisource Earth observation data to feed Earth science models. *Earth Sci. Inform.* **2018**, *11*, 1–17. [CrossRef]
- 74. Ludäscher, B.; Goble, C. Guest editors' introduction to the special section on scientific workflows. *ACM SIGMOD Rec.* **2005**, *34*, 3. [CrossRef]
- 75. Barker, A.; Van Hemert, J. Scientific Workflow: A Survey and Research Directions. In *Proceedings of the 7th International Conference on Parallel Processing and Applied Mathematics, Gdansk, Poland, 9–12 September 2007*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 746–753.
- 76. Olivera, F.; Valenzuela, M.; Srinivasan, R.; Choi, J.; Cho, H.; Koka, S.; Agrawal, A. Arcgis-swat: A geodata model and gis interface for swat. *JAWRA J. Am. Water Resour. Assoc.* **2006**, 42, 295–309. [CrossRef]
- 77. Leonard, L.; Duffy, C.J. Automating data-model workflows at a level 12 HUC scale: Watershed modeling in a distributed computing environment. *Environ. Model. Softw.* **2014**, *61*, 174–190. [CrossRef]
- 78. Omran, A.; Dietrich, S.; Abouelmagd, A.; Michael, M.; Märker, M. New ArcGIS tools developed for stream network extraction and basin delineations using Python and java script. *Comput. Geosci.* **2016**, *94*, 140–149. [CrossRef]
- 79. Essawy, B.T.; Goodall, J.L.; Xu, H.; Rajasekar, A.; Myers, J.D.; Kugler, T.A.; Billah, M.M.; Whitton, M.C.; Moore, R.W. Server-side workflow execution using data grid technology for reproducible analyses of data-intensive hydrologic systems. *Earth Space Sci.* **2016**, *3*, 163–175. [CrossRef]
- 80. De Mulder, C.; Flameling, T.; Weijers, S.; Amerlinck, Y.; Nopens, I. An open software package for data reconciliation and gap filling in preparation of Water and Resource Recovery Facility Modeling. *Environ. Model. Softw.* **2018**, 107, 186–198. [CrossRef]
- 81. Berrick, S.; Leptoukh, G.; Farley, J.; Rui, H. Giovanni: A Web Service Workflow-Based Data Visualization and Analysis System. *IEEE Trans. Geosci. Remote Sens.* **2009**, 47, 106–113. [CrossRef]
- 82. Yue, P.; Zhang, M.; Tan, Z. A geoprocessing workflow system for environmental monitoring and integrated modelling. *Environ. Model. Softw.* **2015**, *69*, 128–140. [CrossRef]
- 83. Zhao, P.; Foerster, T.; Yue, P. The Geoprocessing Web. Comput. Geosci. 2012, 47, 3–12. [CrossRef]
- 84. Feigenbaum, E.A. Expert systems: Principles and Practice. In *The Encyclopedia of Computer Science and Engineering*; Wiley: Hoboken, NJ, USA, 1992.
- 85. Berners-Lee, T.; Hendler, J.; Lassila, O. The Semantic Web. Sci. Am. 2001, 284, 34-43. [CrossRef]
- 86. Zhao, P.; Di, L.; Yu, G.; Yue, P.; Wei, Y.; Yang, W. Semantic Web-based geospatial knowledge transformation. *Comput. Geosci.* **2009**, *35*, 798–808. [CrossRef]
- 87. Yue, P.; Guo, X.; Zhang, M.; Jiang, L.; Zhai, X. Linked Data and SDI: The case on Web geoprocessing workflows. ISPRS-J. Photogramm. *Remote Sens.* **2016**, *114*, 245–257.
- 88. Scheider, S.; Ballatore, A. Semantic typing of linked geoprocessing workflows. *Int. J. Digit. Earth* **2017**, 11, 113–138. [CrossRef]
- 89. Yue, P.; Gong, J.; Di, L.; He, L.; Wei, Y. Integrating semantic web technologies and geospatial catalog services for geospatial information discovery and processing in cyberinfrastructure. *GeoInformatica* **2011**, *15*, 273–303. [CrossRef]
- 90. Sun, Z.; Yue, P.; Lu, X.; Zhai, X.; Hu, L. A Task Ontology Driven Approach for Live Geoprocessing in a Service-Oriented Environment. *Trans. GIS* **2012**, *16*, 867–884. [CrossRef]
- 91. Hofer, B.; Papadakis, E.; Mäs, S. Coupling Knowledge with GIS Operations: The Benefits of Extended Operation Descriptions. *ISPRS Int. J. Geo Inf.* **2017**, *6*, 40. [CrossRef]

- 92. Jiang, J.; Zhu, A.X.; Qin, C.Z.; Zhu, T.; Liu, J.; Du, F.; Liu, J.; Zhang, G.; An, Y. CyberSoLIM: A cyber platform for digital soil mapping. *Geoderma* **2016**, *263*, 234–243. [CrossRef]
- 93. Yue, P.; Di, L.; Yang, W.; Yu, G.; Zhao, P. Semantics-based automatic composition of geospatial Web service chains. *Comput. Geosci.* **2007**, *33*, 649–665. [CrossRef]
- 94. Lutz, M.; Lucchi, R.; Friis-Christensen, A.; Ostländer, N. A Rule-Based Description Framework for the Composition of Geographic Information Services. In Proceedings of the International Conference on GeoSpatial Sematics, Mexico City, Mexico, 29–30 November 2007; pp. 114–127.
- 95. Martin, D.; Burstein, M.; Hobbs, J.; Lassila, O.; McDermott, D.; McIlraith, S.; Narayanan, S.; Paolucci, M.; Parsia, B.; Payne, T. OWL-S: Semantic markup for web services. *W3C Memb. Submiss.* **2004**, 22, 2007–2004.
- 96. Roman, D.; Keller, U.; Lausen, H.; Bruijn, J.D.; Stollberg, M.; Polleres, A.; Feier, C.; Bussler, C.; Fensel, D. Web Service Modeling Ontology. *Appl. Ontol.* **2005**, *1*, 77–106.
- 97. Yue, P.; Di, L.; Yang, W.; Yu, G.; Zhao, P.; Gong, J. Semantic Web Services-based process planning for earth science applications. *Int. J. Geogr. Inf. Sci.* **2009**, 23, 1139–1163. [CrossRef]
- 98. Farnaghi, M.; Mansourian, A. Automatic composition of WSMO based geospatial semantic web services using artificial intelligence planning. *J. Spat. Sci.* **2013**, *58*, 235–250. [CrossRef]
- 99. Peer, J. Web service composition as AI planning: A survey; University of St. Gallen Switzerland: Gallen, Switzerland, 2005.
- 100. Cruz, S.A.B.; Monteiro, A.M.V.; Santos, R. Automated geospatial Web Services composition based on geodata quality requirements. *Comput. Geosci.* **2012**, *47*, 60–74. [CrossRef]
- 101. Farnaghi, M.; Mansourian, A. Disaster planning using automated composition of semantic OGC web services: A case study in sheltering. *Comput. Environ. Urban Syst.* **2013**, *41*, 204–218. [CrossRef]
- 102. Farnaghi, M.; Mansourian, A. Multi-Agent Planning for Automatic Geospatial Web Service Composition in Geoportals. *ISPRS Int. J. Geo Inf.* **2018**, 7, 404. [CrossRef]
- 103. Li, H.; Zhu, Q.; Yang, X.; Xu, L. Geo-information processing service composition for concurrent tasks: A QoS-aware game theory approach. *Comput. Geosci.* **2012**, *47*, 46–59. [CrossRef]
- 104. Yue, P.; Tan, Z.; Zhang, M. GeoQoS: Delivering Quality of Services on the Geoprocessing Web. In Proceedings of the OSGeo's European Conference on Free and Open Source Software for Geospatial (FOSS4G-Europe 2014), Bremen, Germany, 15–17 July 2014.
- 105. Voinov, A.; Kolagani, N.; McCall, M.K.; Glynn, P.D.; Kragt, M.E.; Ostermann, F.O.; Pierce, S.A.; Ramu, P. Modelling with stakeholders—Next generation. *Environ. Model. Softw.* **2016**, 77, 196–220. [CrossRef]
- 106. Giuliani, G.; Rahman, K.; Ray, N.; Lehmann, A. OWS4SWAT: Publishing and Sharing SWAT Outputs with OGC standards. *Int. J. Adv. Comput. Sci. Appl.* **2013**, *3*, 90–98. [CrossRef]
- 107. Rajib, M.A.; Merwade, V.; Kim, I.L.; Zhao, L.; Song, C.; Zhe, S. SWATShare—A web platform for collaborative research and education through online sharing, simulation and visualization of SWAT models. *Environ. Model. Softw.* **2016**, *75*, 498–512. [CrossRef]
- 108. Horsburgh, J.S.; Morsy, M.M.; Castronova, A.M.; Goodall, J.L.; Gan, T.; Yi, H.; Stealey, M.J.; Tarboton, D.G. HydroShare: Sharing Diverse Environmental Data Types and Models as Social Objects with Application to the Hydrology Domain. *J. Am. Water Resour. Assoc.* **2016**, *52*, 873–889. [CrossRef]
- 109. Guigoz, Y.; Lacroix, P.; Rouholahnejad, E.; Ray, N.; Giuliani, G. SCOPED-W: SCalable Online Platform forextracting Environmental Data and Water-related model outputs. *Trans. GIS* **2017**, 21, 748–763. [CrossRef]
- 110. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **2017**, 202, 18–27. [CrossRef]
- 111. Soille, P.; Burger, A.; De Marchi, D.; Kempeneers, P.; Rodriguez, D.; Syrris, V.; Vasilev, V. A versatile data-intensive computing platform for information retrieval from big geospatial data. *Future Gener. Comput. Syst.* **2018**, *81*, 30–40. [CrossRef]
- 112. Wiegand, N.; García, C. A Task-Based Ontology Approach to Automate Geospatial Data Retrieval. *Trans. GIS* **2007**, *11*, 355–376. [CrossRef]
- 113. Li, M.; Guo, W.; Duan, L.; Zhu, X. A case-based reasoning approach for task-driven spatial–temporally aware geospatial data discovery through geoportals. *Int. J. Digit. Earth* **2017**, *10*, 1146–1165. [CrossRef]
- 114. Qiu, L.Y.; Zhu, Q.; Gu, J.Y.; Du, Z.Q. A task-driven disaster data link approach. *ISPRS Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2015**, 179–186. [CrossRef]
- 115. Hu, L.; Yue, P.; Zhang, M.; Gong, J.; Jiang, L.; Zhang, X. Task-oriented Sensor Web data processing for environmental monitoring. *Earth Sci. Inform.* **2015**, *8*, 511–525. [CrossRef]

- 116. You, L.; Lin, H. Towards a research agenda for knowledge engineering of virtual geographical environments. *Ann. GIS* **2016**, 22, 1–9. [CrossRef]
- 117. Delgado, F.; Martínez-González, M.M.; Finat, J. An evaluation of ontology matching techniques on geospatial ontologies. *Int. J. Geogr. Inf. Sci.* **2013**, 27, 2279–2301. [CrossRef]
- 118. Yu, L.; Qiu, P.; Liu, X.; Lu, F.; Wan, B. A holistic approach to aligning geospatial data with multidimensional similarity measuring. *Int. J. Digit. Earth* **2017**, *11*, 845–862. [CrossRef]
- 119. Sun, K.; Zhu, Y.; Song, J. Progress and Challenges on Entity Alignment of Geographic Knowledge Bases. *ISPRS Int. J. Geo Inf.* **2019**, *8*, 77. [CrossRef]
- 120. Paulheim, H. Knowledge Graph Refinement: A Survey of Approaches and Evaluation Methods. *Semant. Web* **2017**, *8*, 489–508. [CrossRef]
- 121. Ehrlinger, L.; Wöß, W. Towards a Definition of Knowledge Graphs. In Proceedings of the SEMANTiCS2016, Leipzig, Germany, 13–14 September 2016.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).