Environmental Modelling and Software 163 (2023) 105687

Contents lists available at ScienceDirect

Environmental Modelling and Software

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

Documentation strategy for facilitating the reproducibility of geo-simulation experiments

Zhiyi Zhu [a,b,c], Min Chen [a,b,c,*], Zhen Qian [a,b,c], Hengyue Li [a,b,c], Kai Wu [a,b,c], Zaiyang Ma [a,b,c], Yongning Wen [a,b,c], Songshan Yue [a,b,c], Guonian Lü [a,b,c]

[a] *Key Laboratory of Virtual Geographic Environment (Ministry of Education of PR China), Nanjing Normal University, Nanjing, Jiangsu, China*

[b] *Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing Normal University, Nanjing, Jiangsu, China*

[c] *State Key Laboratory Cultivation Base of Geographical Environment Evolution (Jiangsu Province), Nanjing Normal University, Nanjing, Jiangsu, China*

A R T I C L E I N F O

Handling Editor: Daniel P Ames

Keywords:
Reproducibility
Geo-simulation experiment
Lifecycle
Experimental documentation

A B S T R A C T

Reproducibility is important to verify the credibility and reliability of scientific research. Geoscientists increasingly focus on the reproducibility of Geo-Simulation Experiments (GSEs). However, geoscientists' propensity to record in different styles and extensive speculation about the essential information required make it difficult to reproduce a GSE from a notebook. To facilitate reproducibility of GSEs, this article proposes documentation strategy based on a conceptual model, designed to identify and abstract the GSE lifecycle to demonstrate different stages of GSEs. There are three main parts in this strategy: recommendations of reference descriptions, construction of a GSE's Document (GSEDocument), and processes for evaluating the reproducibility of GSEs. Thus, a reproducible GSEDocument can be constructed to document and share the GSE lifecycle. Moreover, to verify the effectiveness of documentation strategy, a case study from an online competition is conducted. Through this strategy, transparency, utility, and credibility of published research related to GSEs can be improved.

1. Introduction

Currently, concerns about scientific credit are steadily rising (Easterbrook, 2014; Goodman et al., 2016; Ioannidis, 2005), with more than 70% of scientists failing to reproduce others' experiments (Baker, 2016; Open Science Collaboration, 2015). Movements to address the "credibility crisis" of published computational results are arising in fields as disparate as computational science, neuroscience, and bioinformatics in particular (Begley and Ioannidis, 2015; Committee on Reproducibility and Replicability in Science et al., 2019; Peng et al., 2006; Poldrack, 2019). Reproducibility is the precondition and substantial ability to prove research credibility, as reproducible research enables scientists to reimplement experiments reported by other scientists (Barba, 2018; Konkol et al., 2019; McNutt, 2014, 2014, 2014). Generally, reproducibility can be cited as the ability to obtain consistent results to answer the same question, as well as the usage of the same materials and methods (Barba, 2018; Committee on Reproducibility and Replicability in Science et al., 2019; Wolke et al., 2016).

A Geo-Simulation Experiment (GSE) can be regarded as a series of processes carried out to answer a real-world question utilizing computational geographic analysis models and tools (Balci, 1990; Sacks et al., 1989). It is a practical vehicle for geographic simulation, which is an application step of geographic modeling aiming to reflect and predict specific geographic patterns and processes (Batty, 2011; Chen et al., 2020; Ruscheinski et al., 2020).

There are several reasons why improving the reproducibility of GSEs is important. First, complex and uncontrolled natural phenomena and geographic events (such as earthquakes and debris flows) are impossible to fully reproduce since no two phenomena or events are the same (Committee on Reproducibility and Replicability in Science et al., 2019). However, to prove a model's usability as well as the credibility of findings, reproducing natural environments and phenomena can be attempted to some extent by means of reproducing GSEs. Second, with the emergence of quantitative research in geoscience, computers have become an indispensable component of GSEs (Fotheringham and Sachdeva, 2022; Qian et al., 2022a), manifesting as a series of computational

* Corresponding author. Key Laboratory of Virtual Geographic Environment (Ministry of Education of PR China), Nanjing Normal University, Nanjing, 210023, China.

E-mail address: chenmin0902@163.com (M. Chen).

https://doi.org/10.1016/j.envsoft.2023.105687
Received 18 January 2023; Received in revised form 7 March 2023; Accepted 22 March 2023
Available online 28 March 2023

models and tools. The evolution of the computational mode in GSEs has created a demand for its reproducibility in order to meet the requirement of reproducibility of computational experiments (Gil et al., 2016; Goodchild et al., 2021; Li et al., 2021a; Piccolo and Frampton, 2016; Stodden et al., 2018). Third, given the complexity and uncertainty of geoscience, processes of world exploration tend to be iterative in light of previous GSEs (Chen et al., 2021; Gil et al., 2016). Based on reproducing GSEs, the geo-community could understand the specific approaches to geographic problems, correct previous errors, and even disclose some truths of geo-phenomena (Barba, 2018; Goodchild and Li, 2021).

According to the definition of reproducibility above, a GSE is reproducible when the same resources (i.e., data resources and model resources) and processes (i.e., a series of activities that interact to generate a numerical output) are used to provide consistent results in resolving the same geographic question. Making a GSE reproducible is challenging because intrinsic processes of the given GSE are often not fully available to reviewers (Konkol et al., 2019; Novère et al., 2005; Vasilevsky et al., 2013). It is routine to use experiment notebooks and documentation to write a narrative about what exactly these analytical steps are in an experiment lifecycle, from generating a geo-problem to obtaining a satisfactory conclusion (Dirnagl and Przesdzing, 2016; Kluyver et al., 2016; Piccolo and Frampton, 2016). Grimm et al. have proposed TRAnsparent and Comprehensive Ecological modelling documentation (TRACE) to provide better modeling support (Grimm et al., 2014). The Jupyter notebook has now been widely used by scientists from different fields to save their ideas and computational analyses (Kluyver et al., 2016).

However, there are still some challenges to keeping a notebook. The first challenge involves the preparation process for recording in the notebook, which requires considerable effort (Gil et al., 2016; Sandve et al., 2013). This is mainly because keeping a notebook is quite tedious work requiring geoscientists to choose and rearrange related information in a structured way, while they may have no idea which information is core (Essawy et al., 2020; Konkol et al., 2019; Stodden et al., 2018). If geoscientists have not divulged enough information from a complicated experiment, it is probable that independent variables were not effectively controlled, rendering a GSE irreproducible when reviewers attempt to recreate it. The second challenge is that many geoscientists keep notebooks at their own favor and discretion in different forms (Konkol et al., 2019; Pimentel et al., 2019), which is casual and often incomplete. As a result, a notebook may lack crucial information and include an excessive amount of useless information (Ayllón et al., 2021). This will not only lead to difficulty in migration between different platforms but also to confusion in the understanding of experiments by fellow scientists (Dirnagl and Przesdzing, 2016). After some time, these notebooks may not be understood even by the authors themselves.

In this context, this study aims to provide the documentation strategy to facilitate the reproducibility of a GSE. The GSE lifecycle is identified as a framework of GSEs containing several activities, actions, and tasks that are required to solve a geo-problem with acceptable quality characteristics (Balci, 1994, 2012). Meanwhile, the GSE lifecycle can be used to express different and iterative stages of GSEs, and a CMOR model is designed to generalize and abstract it. Moreover, the recommendation of reference descriptions, the construction of a GSE's Document (GSEDocument), and the processes for evaluating the reproducibility of GSEs based on a GSEDocument are three main parts of the proposed strategy. Using the proposed strategy, a GSEDocument can be easily created by geoscientists, making a GSE understood and reproduced step-by-step in a clearer and more standard form by fellow scientists. Furthermore, this study implements the proposed strategy into practice, provides a GSEDocument example for an online competition involving geographic analysis modeling and simulation, and achieves some substantial results finally.

There is also the issue of scope when applying documentation strategy to reality. Above all, this strategy applies only to the computational experiments that are accessible in most domains of geographic simulation, such as hydrology, soil, atmosphere, biology, and human civilization. Moreover, reference descriptions provided in this study only refer to the essential information that should be included when publishing a GSE, without specifying the technical form in which the information should be represented, such as text, graph, or video form. It is important to note that this study is not concerned with verifying the correctness or rationality of the purpose, resources, steps, or results. However, fellow scientists are encouraged to verify the adequacy of information provided to obtain comprehensive results. Finally, this study is not intended to criticize the quality of any existing work or notebook. In other words, it is possible for many GSEs to be reproducible without complying with the documentation strategy. We anticipate that, through the proposed strategy, other scientists will be able to contribute more effectively to related future research.

The remainder of this article is structured as follows. In Section 2, the expression of the GSE lifecycle and basic idea for constructing documentation strategy are explained. Section 3 describes the recommendations of reference descriptions, the detailed design to construct a GSEDocument, and the processes for evaluating reproducibility. Section 4 verifies the documentation strategy with a case study, and Section 5 provides the discussion and conclusion of the study.

2. Basic design

The documentation strategy depends on the description of a GSE, which can be articulated by the GSE lifecycle. This section first outlines GSE lifecycle and evaluation for its reproducibility. Then, *integrality*, *effectiveness*, and *objectivity* are identified to guide the construction of this strategy. Finally, a conceptual model is developed to express the abstraction of the GSE lifecycle, as shown in Fig. 1.

2.1. GSE lifecycle

Currently, there are two primary forms of GSEs. Geoscientists discover real-world phenomena and patterns by using geographic analysis models. For instance, some scientists utilize spatial analysis models to simulate urban layouts and environments (Li et al., 2021b; Qian et al., 2020; Zhang et al., 2022). Additionally, geoscientists investigate the robustness and universality of models with a comparison of several models by simulating the target region (Qian et al., 2022b; Tsai et al., 2021).

GSE lifecycle is concluded as a series of ordered tasks completed in GSEs by geoscientists, beginning with (1) scientific geo-problem identification, (2) purpose definition for determination of a GSE, (3) experimental scheme design, (4) data and model resource preparation, (5) model calculation, (6) output validation, (7) output visualization, and (8) result analysis and answers for a geo-problem (Balci, 2012; Ma et al., 2021, 2022; Ruscheinski et al., 2020).

2.2. Evaluation of a GSE's reproducibility

The ultimate purpose of this study is to increase the reproducibility of GSEs. Once a GSE-based paper is published, geoscientists are expected to reproduce the involved experiment.

For this blueprint, evaluation of the reproducibility of GSE is essential, meaning to what extent a GSE can be considered to be reproducible. Three levels of reproducibility were defined to generalize the reproducibility of GSE. First, complete reproducibility means that a GSE can be reimplemented to yield the consistent outputs and answer the same geo-problem. Second, partial reproducibility means that one or some procedures can be reproduced and yield the consistent intermediate output while answering the same geo-problem. Third, irreproducibility refers to any procedures that cannot be re-executed to answer the same geo-problem. In addition, consistent outputs refer to the procedure outputs that are equal to the original outputs, which can be examined by statistical indicators, such as the Mean Square Error (MSE),

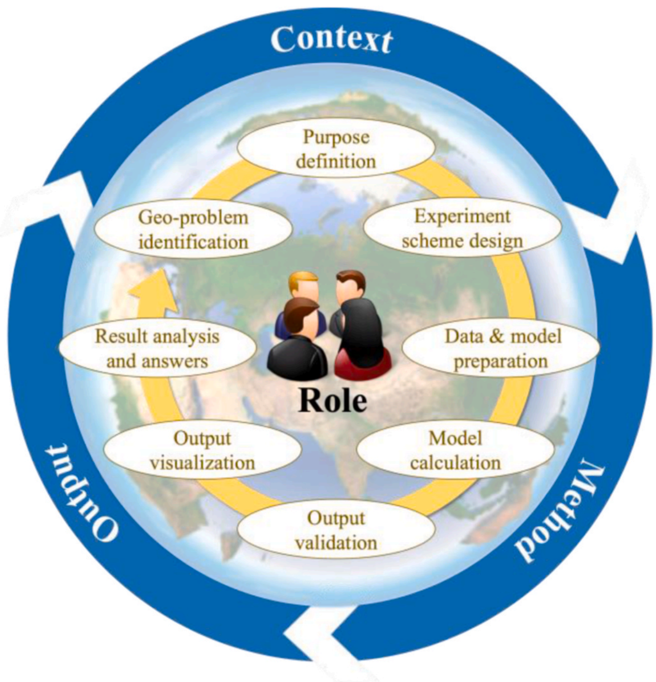

Fig. 1. Abstraction of the Geo-Simulation Experiment (GSE) lifecycle.

Mean Absolute Error (MAE), and R^2.

2.3. Fundamental principles

To guide the construction of documentation strategy, three main goals have been identified:

Integrality: this indicates whether the entire information needed to reproduce a GSE's results is provided publicly (Chirigati et al., 2013). To ensure *integrality*, documentation strategy considers the usefulness and reliability of information, which could influence the difference in results for future reviewers.

Effectiveness: this refers to whether the details are essentially provided. In general, some but not all details are essential for documenting the GSE lifecycle to ensure its reproducibility (Davison, 2012). Thus, careful consideration is given to which information that should be turned into entries in a public document to avoid redundancy.

Objectivity: this means that the result of a GSE should not be influenced by any experimenter (Ferro et al., 2016). To guarantee *objectivity*, this study assumes certain information for steps with identical inputs and outputs. Although some analyses may involve randomness, the use of constant random numbers given the same initial seed can ensure consistency.

2.4. Conceptual model

Following the three principles and in the interest of understandability, the CMOR model is designed to generalize and abstract series tasks of the GSE lifecycle, as shown in equation (1) and Fig. 1.

$$\text{A GSE lifecycle} = <C, M, O\ |R>. \quad (1)$$

R is the ***Role** Item*. Individuals responsible for the GSE lifecycle, as well as individuals and agencies in the cited information, are referred to the *Role Item*. This is used to indicate the provenance of responsible individuals to enhance the credibility and traceability of this GSE.

C refers to the ***Context** Item*. It is a background summary aiming to gain awareness about a geo-problem. The sheer intellectual diversity of geoscience requires precise cognition to allow other geoscientists to comprehend purposes and demands, as well as sufficiently understand the geo-problems (Badham et al., 2019; Ma et al., 2021).

M is the ***Method** Item*. It is the collection of computational processes that includes tools, techniques, and models that can be used to address a problem or problem situation (Mingers, 2000). *Method Item* is the most important part during a GSE lifecycle, designed to understand how to conduct a precise implementation strategy.

O refers to the ***Output** Item*. It is a set of analyses and interpretations of outputs made to generate final results (Sandve et al., 2013) without any numerical change in raw output. The analysis and evaluation of GSE results play a vital role in decision support (Chen et al., 2021; Lü et al., 2019). At the end of the GSE lifecycle, raw outputs from the *Method Item* tend to be analyzed, as well as the textual interpretations aiming to answer the geo-problem mentioned in the *Context Item*.

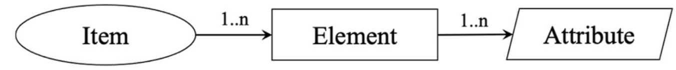

Fig. 2. Relationships of the IEA (<*Item*, *Element*, *Attribute*>) triple.

Table 1
Applicability of *elements* in each *item*.

Item	*Element*	Applicability
Role	*Author*	Submitter or stakeholders of a GSE;
	Cited Authorship	Authorships of any cited information;
Context	*Problem*	Statement of a formulated geo- problem;
	Solution	Response of a geo- problem;
Method	*Resource*	Data resources used and produced in the GSE lifecycle; Model resources used to process initial data and to generate any intermediate or final outputs;
	Procedure	A collection of ordered steps, with the application of different resources;
Output	*Validation*	Analysis steps for validating the accuracy, clarity, and details of the raw numerical output;
	Representation	Textual interpretations analyzing outputs aiming to answer the geo- problem mentioned;

This study also constructed a submodel of the IEA (<*Item*, *Element*, *Attribute*>) triple with equation (2) to represent different tasks in the GSE lifecycle. According to the IEA triple, this paper abstracts the GSE lifecycle to the *attribute* level, which can demonstrate the GSE lifecycle in a structured and comprehensive manner. Here, *item* is used to describe different phases during the GSE lifecycle, divided into *Role Item*, *Context Item*, *Method Item*, and *Output Item*. It also consists of different *elements* and *attributes*. The relationships of the IEA triple are shown in Fig. 2. An *element* can be cited as a subitem in an *item*, which is primarily the union of task-detailed information. In addition, an *attribute* is a specification that defines a property of an *element*. At the *attribute* level, key-value pairs are designed to guide geoscientists to fill out each *attribute* of the GSEs' aspects, where "key" refers to the author-defined attribute and "value" means the detailed information that needs to be described for the "key".

$$\mathit{Item} \sim <\mathit{Element}, \mathit{Attribute}> \quad (2)$$

3. Design of the documentation strategy

3.1. Reference descriptions for the GSE lifecycle

Reference descriptions serve as a standard format on the basis of the CMOR model, which recommends that core information needs to be documented in a notebook. Before examining specific ways to record the GSE lifecycle, it will be useful to develop a vocabulary of reference descriptions. This will assist authors in considering what they can explain in an *item* and what they may do in each *element*. Table 1 presents the applicability of specific *elements* in each *item*. In addition, details and

Table 2
Recommendation for *attributes* in each *element*.

Item	Element	Recommendation for *Attribute*
Role	*Author*	R1: *Author Element* should be correctly contacted; R2: Authors' type serving the actual position should be provided;
	Cited Authorship	R3: *Cited Authorship Element* should specify literature references of any cited information; R4: Agent information that comes with resources should be cited;
Context	*Problem*	C1: *Problem Element* should specify precise and applied questions; C2: *Problem Element* should have the purpose of a GSE; C3: *Problem Element* should specify geographic characteristics;
	Solution	C4: *Problem Element* should document necessary hypotheses; C5: Sketch of the experimental design should be provided;
Method	*Data Resources*	M1: General information of Data Resources should be specified; M2: Data Resources should be published in a publicly accessible location; M3: Data Resources should have a license; M4: Values of parameters should be specified;
	Model Resources	M5: General information of Model Resources should be specified; M6: Model Resources should be published in a publicly accessible location; M7: Model Resources should have a license; M8: Application/software/platform dependency information; M9: Model Resources should specify execution requirements;
	Procedure	M10. Procedure should have general information of what has been done; M11. Procedure should specify its configuration requirement; M12. Procedure should document the dataflow; M13. Settings of boundary conditions should be documented (If existing);
Output	*Validation*	O1. Validation should have general information of what has been done; O2. Validation should specify its configuration requirement; O3. Validation should document the dataflow; O4. Method applied in a procedure should be specified; O5. Location of raw output from Validation should be specified; O6. Results produced from validation method should be noted.
	Representation	O7. Particular analysis of final results should be specified.

reasons for the recommended information that should be logged in the *attributes* of a GSE are shown in Table 2.

3.1.1. Applicability of elements in each item

***Author* and *Cited Authorship Element*.** Every GSE should have at least one author who serves as a submitter (Chirigati et al., 2013). There are also some stakeholders who make various contributions, from developing technical routes to analyzing raw outputs, such as developers, teammates, and decision makers (Chirigati et al., 2013). Additionally, some information and resources generated are initially produced or declared by other scientists or agencies, which need to be clarified. Thus, one *element* in the *Role Item* is the *Author Element*, which refers to someone who serves as a submitter and stakeholder in a GSE. The *Cited Authorship Element* is the other *element* of *Role Item* related to authorships of any cited information in a GSE. Key contributors of any cited information and resources who are not authors of a GSE should also be assigned (Gil et al., 2016). This *element* is used to indicate the provenance in order to enhance the credibility and traceability of a GSE.

***Problem* and *Solution Element*.** As defined by Booth et al. and Badham et al., the common structure of a *Context Item* consists of two *elements*, including a statement of the problem and response to the problem (Asdal and Moser, 2012; Badham et al., 2019; Booth et al., 2016). Consequently, the *Problem Element* is the first *element* in the *Context Item*, stating the common ground of a shared understanding about a geo-problem (Song et al., 2020). The *Solution Element* is designed to present a formulated plan in order to gather the desired information for solving the problem above at a minimal cost (DellaVigna et al., 2019; Ferro et al., 2016).

Resource and Procedure Element. As defined by Mingers, this paper considers the compatibility of various models. A *Method Item* can be divided into two aspects, the *Resource Element* and *Procedure Element* (Mingers, 2000). Data resources and model resources constituting all resources in a GSE are used and generated in the process of specific implementation procedures (Chen et al., 2020; Zhang et al., 2021). Data resources are data used and produced in a method when performing a GSE, including input data, parameters, intermediate data (i.e., the data are produced in one step and used in the next step), and final outputs (Gil et al., 2016). Model resources are functioning quantitative models and tools involving a set of equations or rules to simulate the earth (Badham et al., 2019; Chen et al., 2020, 2021). The *Procedure Element* is a crucial concern for recording the implementation procedures in a GSE, as seemingly insignificant experimental details can drastically influence the reproduced outputs (Freire and Silva, 2012). To document the specific implementation of how routines of the *Solution Item* mentioned above are executed, the *Procedure Element* is designed as a collection of ordered steps causing a numerical change in the outputs with the application of different resources (Freire et al., 2012).

Validation and Representation Element. As mentioned above, the *Output Item* includes the validation of data to process and the analysis of desired numerical results. First, the *Validation Element* is not necessary but should not be ignored, determining how well a GSE's outputs compare with desired patterns (Balci, 1998; Sargent, 2010). Second, the *Representation Element* is an interpretation and documentation of simulation results (Rabe et al., 2009; Sargent, 2010). Due to the complexity of some simulation results, failing to provide textual interpretations of these results may lead to incorrect geo-problem decisions, although these results are sufficiently credible (Balci, 2012).

3.1.2. Recommendations for attributes in each element

A total of 29 recommendations regarding the essential information of each *element* in a GSE lifecycle are given, as shown in Table 2.

(1) *Attributes* in a *Role Item*

***Attributes* of an *Author Element*.** (R1) Authors and stakeholders should be correctly contacted if there are problems when reviewing or reproducing a GSE. Thus, name, e-mail, professional title, and affiliated institution should be specified, which is the basic contact information (DeRisi et al., 2003). (R2) The author's type should be specified, such as submitters, developers, teammates, and decision makers. As a result, geo-scientists can contact the corresponding author rapidly when encountering troubles in one procedure of a GSE (Orchard et al., 2011).

***Attributes* of a *Cited Authorship Element*.** (R3) A *Cited Authorship Element* should specify literature references including the authors, publication date, and reference citation. For example, GB/T 7774, MLA, and APA are common standards used for reference citation. (R4) Any relevant agent information that comes with resources (e.g., where exactly the data was found in a publication or report) should be cited, such as the publishers' (or repository's) name. If it exists, contact information, affiliated institutions, and other valuable information should also be noted (Novère et al., 2005).

(2) *Attributes* in a *Context Item*

***Attributes* of a *problem Element*.** (C1) When a geo-problem is recognized, a scientist (or team) initiates an experiment by communicating the geo-problem to an analysis. Thus, precise and applied questions should be outlined first (Balci, 1998). (C2) The purpose can be seen as the significance of doing this experiment. By noting the purpose, authors can create a stronger relationship with reviewers because they connect the extension of a GSE to reviewers via their interested fields (Booth et al., 2016). (C3) Geographic characteristics refer to the spatial-temporal scale of this experiment (e.g., spatial-temporal boundaries, dimensions, and scales) (Ma et al., 2021; Phillips, 2022). In doing so, a GSE can be generalized on a similar spatial-temporal scale for further study.

***Attributes* of a *Solution Element*.** (C4) A hypothesis is one or several conditions that can be provisionally accepted as the foundation for further experimentation (Wikipedia, https://en.wikipedia.org/wiki/Hypothesis#Working_hypothesis). (C5) Experimental design can be shown with a flowchart in the form of a workflow or a summarized graph to respond to the problem above.

(3) *Attributes* in a *Method Item*

***Attributes* of a *Data Resource Element*.** Above all, the description of metadata for *Resource Element* is recommended, which documents the integrity of each resource. Resources can be structured in a public, machine-readable format, as either model standards, such as Systems Biology Markup Language (SBML), Overview, Design concepts and Details (ODD) protocol, and Open Geographic Modeling and Simulation (OpenGMS) standards (Chen et al., 2019; Grimm et al., 2020; Hucka et al., 2003), or data standards such as the World Wide Web Consortium provenance (W3C PROV) and ISO 19115 family, and can be supported by specific software applications (Closa et al., 2017; Groth and Frew, 2012; Zhang et al., 2020). If not, other information need to be provided are mentioned below:

(M1) General information is the overview of *Data Resources*, including name and type. In this article, the data resource type ranges from the initial analysis, input data, parameters, and intermediate data (i.e., the data are produced in one step and used in the next step), to the final output, contributing to the construction of the dataflow of procedures.

(M2) A publicly accessible location where *Data Resources* are posted or where exactly authors can find them in a publication or report should be cited. Ideally, data resources can be cited with a permanent unique identifier, enabling other scientists to retrieve the data with directed access, such as Digital Object Unique Identifiers (DOIs), Archival Resource Keys (ARKs), Identifiers.org, and Personalized Uniform Resource Locator (PURL) (DeRisi et al., 2003; Juty et al., 2020; Wilkinson et al., 2016). In addition, if there are some restrictions, access information should be provided.

(M3) A license specifies the limitations for *Data Resources*' reuse, including whether resources can be modified before redistribution, whether the creator should be acknowledged, and whether the resources can be used for commercial purposes. For example, the Open Data Commons Attribution License (ODC-By) is offered by Creative Commons, allowing re-users to share, create, and adapt the data resources in any medium or format (https://opendatacommons.org/index.html).

(M4) For parameters, the precise value when implementing the GSE needs to be fixed and specified. The same process will typically yield different outputs every time it is executed when the parameters are not the same.

***Attributes* of a *Model Resource Element*.** Like *Data Resources*, *Model Resources* need to be documented and shared with a similar set of *attributes*. (M5) The general information of *Model Resources* should be noted, including name, type and version. There are three formats of *Model Resources*, including module (e.g., buffer analysis in ArcMap), service (e.g., Soil & Water Assessment Tool (SWAT) service in OpenGMS), and code (e.g., Random Forest source code stored on GitHub.org), which should be specified as the type of *Model Resource*. (M6) All formats of *Model Resources* should be published in an accessible location, ideally with a permanent unique identifier, and (M7) a license should be specified to define their acceptable use. For example, the Open Source Initiative (OSI) provides open source license options for model resources in geoscience, such as the MIT license, the Apache Public License (APL), and the GNU General Public License (GPL) (Fitzgerald, 2006).

(M8) *Model Resources* must be correctly used in the relevant application/software/platform. Thus, the application/software/platform dependency information (e.g., what the dependent application/software/platform is and where the application/software/platform is stored), as well as the dependency library (e.g., R package and Python package), operating system (e.g., Windows, MacOS, and Linux) and dependency techniques (e.g., CPU and GPU requirements), should be explicitly documented.

(M9) *Model Resources* should specify the required necessary input/output information (e.g., input data, parameter, and output data) in several model resources, with reference to the initial values of state variables (if they exist).

***Attributes* of a *Procedure Element*.** (M10) Before obtaining details, each procedure should start with general information. This is a description of what has been done in a procedure, as well as what has been accomplished. It should consist of (i) a title, (ii) an overview of what the procedure is and why the procedure must be done, (iii) the order of different procedures, (iv) the version, and (v) the type of procedure (e.g., data evaluation, data preprocessing, simulation analysis, data postprocessing, uncertainty and sensitivity analysis). (M11) The configuration requirements should be documented. These specify what exactly the *Model Resource* used in a procedure is and which *Data Resource* is applied as the input, parameter, or output of this *Model Resource*. (M12) The *Procedure* should document the dataflow. The dataflow indicates how the intermediate data are generated and traded as input data in the next procedure. (M13) If they exist, the model settings of boundary conditions should be documented, such as stop conditions, number of replicates, and limitation scope at runtime.

(4) *Attributes* in an *Output Item*

***Attributes* of a *Validation Item*.** Similar to a *Procedure Element*, a *Validation Element* needs to be documented as a series of procedures. (O1) General information on what has been done should be provided, (O2) configuration requirements for analyzing the raw output should be documented, (O3) the dataflow of the analysis procedure should be specified, and (O4) the validation method applied in this procedure should not be ignored, such as related indicators (e.g., MSE, RMSE, and MAE). Similar to *Data Resources*, (O5) the location of raw output from validation needs to be specified, showing the credibility of validation, and (O6) results produced through the validation method should be noted.

***Attributes* of a *Representation Item*.** (O7) The particular analysis of final outputs should be specified. This is textual information regarding the analysis and interpretation of raw figures.

3.2. GSEDocument construction

Normally, reference descriptions should be consulted as a document in any form when authors want to publish GSE-related papers or reports. Every form, such as free text or an eXtensible Markup Language (XML) schema, should be accepted both online and offline. Thus, this study proposes pathways for creating a GSEDocument on the basis of reference descriptions.

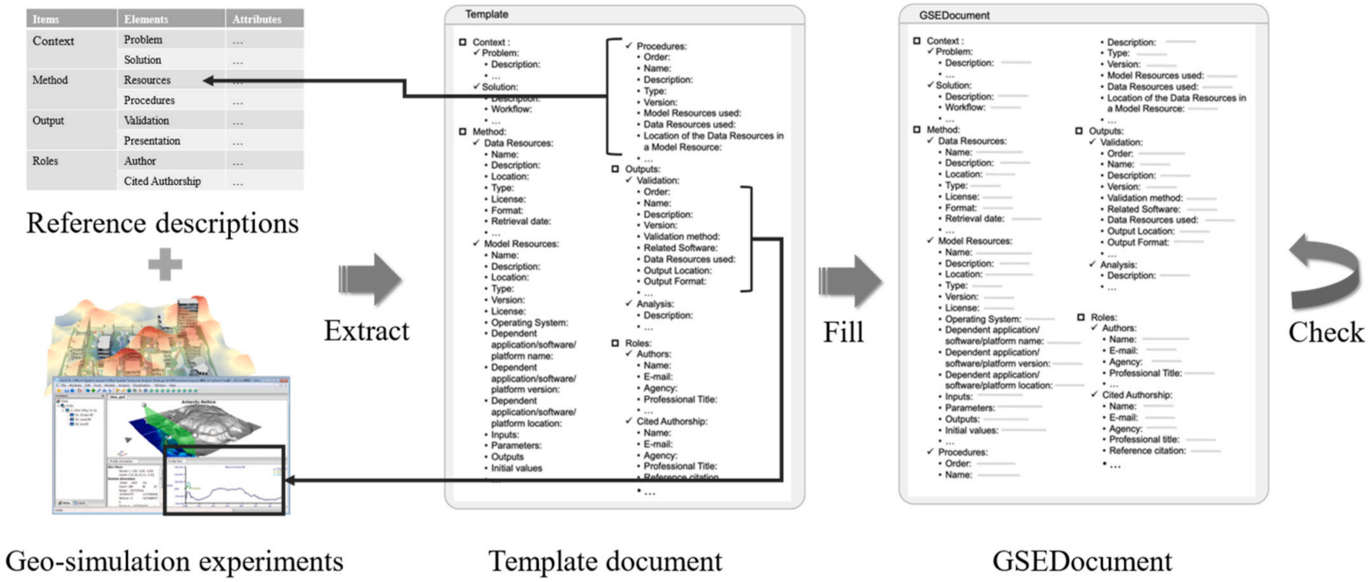

Fig. 3. Flowchart of GSE's Document (GSEDocument) construction.

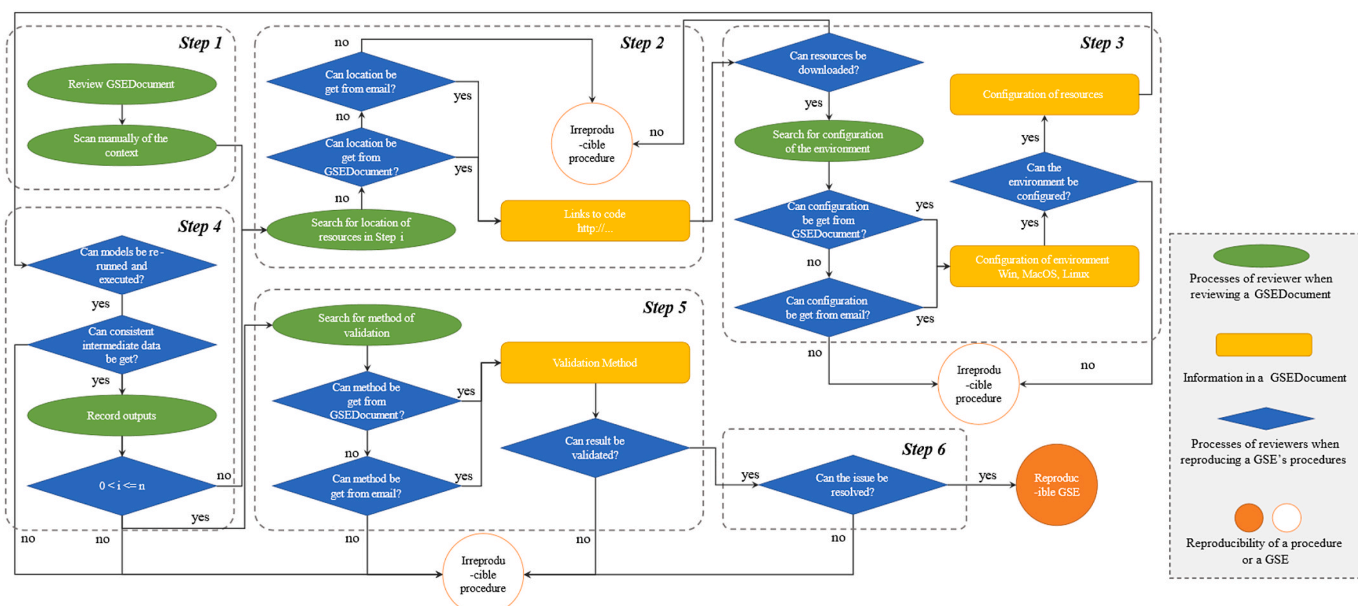

Fig. 4. Processes conducted to evaluate the reproducibility of the GSE lifecycle.

On the basis of reference descriptions, a GSEDocument is a comprehensive document built from a formal collection of entries about the GSE lifecycle that should be described by authors. The key steps for compiling a GSEDocument are shown in Fig. 3. (1) Extract the <*Element, Attribute* > list in accordance with reference descriptions, and append any additional and necessary information. (2) Construct a template for the GSEDocument upon the extracted lists. Each entry in this template employs key-value pairs, where "key" means relying on the *attributes* contained in the above reference and "value" is the real life experimental information. (3) Retrieve corresponding information from a GSE, including text, graphs, tables, and videos. After this step, the GSEDocument is structured in accordance with the template. (4) Check the completeness of the GSEDocument. With these steps, geoscientists can define any GSEDocument that best meets its requirements for simulation characteristics.

3.3. Processes for evaluating the reproducibility based on GSEDocument

One of the most important parts of the documentation strategy is processes conducted to evaluate the GSE's reproducibility, which refers to how reviewers evaluate the GSE's reproducibility based on a GSEDocument.

According to the three levels of reproducibility definition above, the processes conducted are shown in Fig. 4. It is carried out in six phases and on account of GSEDocument: (1) reviewing a GSEDocument to understand a geo-problem; (2) locating any related resources; (3) attempting to obtain resources and related software and applications, as well as configuring the environment of related resources; (4) rerunning models with related data in each procedure, recording the intermediate data and output, and comparing reproduced intermediate data and output with the initial ones; (5) validating raw output; and (6) checking the analysis and interpretation of the results.

Table 3
The template of the GSEDocument example.

Item	Element	Attributes
Role	*Author*	Name; E-mail; Professional title; Affiliated institution; Type;
	Cited Authorship	Name[a]; E-mail[a]; Professional title[a]; Affiliated institution[a]; Citation of articles in APA format;
Context	*Problem*	Geo-problem; Purpose; Spatial scale; Temporal scale;
	Solution	Experimental design; Hypotheses[a];
Method	*Resource*	Data resource: Name; Description; Type; Location; License; Value of parameter[a]; Access information[a]; Model resource: Name; Description; Type; Location; Version; License; Dependent application/software/platform name; Dependent application/software/platform version; Dependent application/software/platform location; Operating system; Dependency library; Dependency techniques; Necessary inputs, parameters, and outputs;
	Procedure	Order; Name; Description; Type; Version; Model resources used; Data resources used; Location of the data resources in a model resource; Stop conditions[a]; Number of cycles[a];
Output	*Validation*[a]	Order; Name; Description; Version; Validation method used; Output used; Location of the output in a Validation method; Results produced;
	Representation	Analysis of raw result;

[a] Indicates that this *attribute* or *element* is not necessary to describe in this template.

4. Case study

To gradually enable the documentation strategy to be accepted and disseminated by the geo-community, this study provides a GSEDocument example for an online competition involving geographic analysis modeling and simulation. This section introduces the content of the GSEDocument example first and then analyzes the results uploaded by competitors to demonstrate the efficacy of the proposed strategy.

4.1. A GSEDocument example for an online competition

The First National Geographic Analysis Model Development and Application Competition was launched by Nanjing Normal University from 7th October to 7th December 2022. This competition is established to improve the capabilities of development and application of geographic analysis models for undergraduates and graduates. More than 700 participants from universities across China participated in the competition. A GSEDocument template, the content of which is shown in Table 3 and is based on reference descriptions, was created for this competition. At the deadline of this competition, we received 72 GSEDocuments.

According to these three levels of reproducibility above, we found that 30 GSEs could be completely reproduced, yielding identical answers and responses to the same question, while 33 GSEs were partially reproducible, and 9 of them could not be reproduced.

There were four reasons why a complete GSE could not be reproduced or could be only partially reproduced, as shown in Fig. 5. The first challenge is exacerbated by the incompleteness of the GSEDocument, accounting for 57.14% of incompletely reproducible GSEs. Second, nonstandard description detail is another cause, accounting for 47.62%. Third, 7.14% of these GSEs were not completely reproducible because reviewers were not able to achieve consistent results. While they tried to reproduce GSEs in accordance with the information shown in the GSEDocument, they obtained different results, causing irreproducibility of the procedures. Finally, 2.38% of these GSEs' models could not be rerun due to the runtime error. When we tried to execute models mentioned in the GSEDocument with the usage of initial data resources, the models were always run crash, resulting in the runtime error.

4.2. A typical GSE in the competition

Taking one typical case in the competition as an example, this study illustrates how to evaluate a GSE's reproducibility based on the documentation strategy. A brief presentation of this case is shown in Fig. 6.

First, the GSE aims to quantitatively examine the impact of the built environment on crime occurrence from the perspective of multi-scale analysis, utilizing big data of streetscape images and deep learning technology.

Second, this GSE has several models, including FCN-8s (Fully Convolutional Networks-8s), Buffer, GWR (Geographically Weighted Regression), and MGWR (Multi-scale Geography Weighted Regression). These models also correspond to some software, such as ArcMap 10.5, GPU-CUDA-enabled Semantic Segmentation App, MGWR 2.2, Office 2019, and R studio 4.2.1. Additionally, there are numerous data resources, such as Google Street View image, road network, buffer distance, environmental characteristic indicator, and Manhattan crime datasets, used as input, parameters, or intermediate data in a procedure. These data resources are uploaded into the OpenGMS data container (Yue et al., 2015), which can be downloaded by other reviewers. All the initial data resources and model software could be obtained from the provided locations.

Third, the environment was configured in accordance with the description of the model's resources. In particular, the semantic segmentation application requires hardware, that is, GPUs with NVDIA CUDA 9.x.

Fourth, based on the information of each procedure, this study reran the models and recorded all the used and produced data. There were thirteen procedures in this GSE, and the dataflow shown in Fig. 6 was

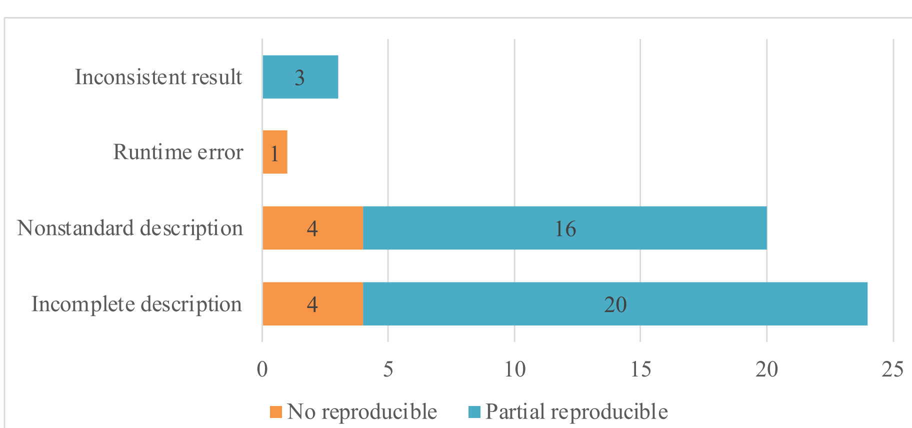

Fig. 5. Analysis of experiments that could not be completely reproduced.

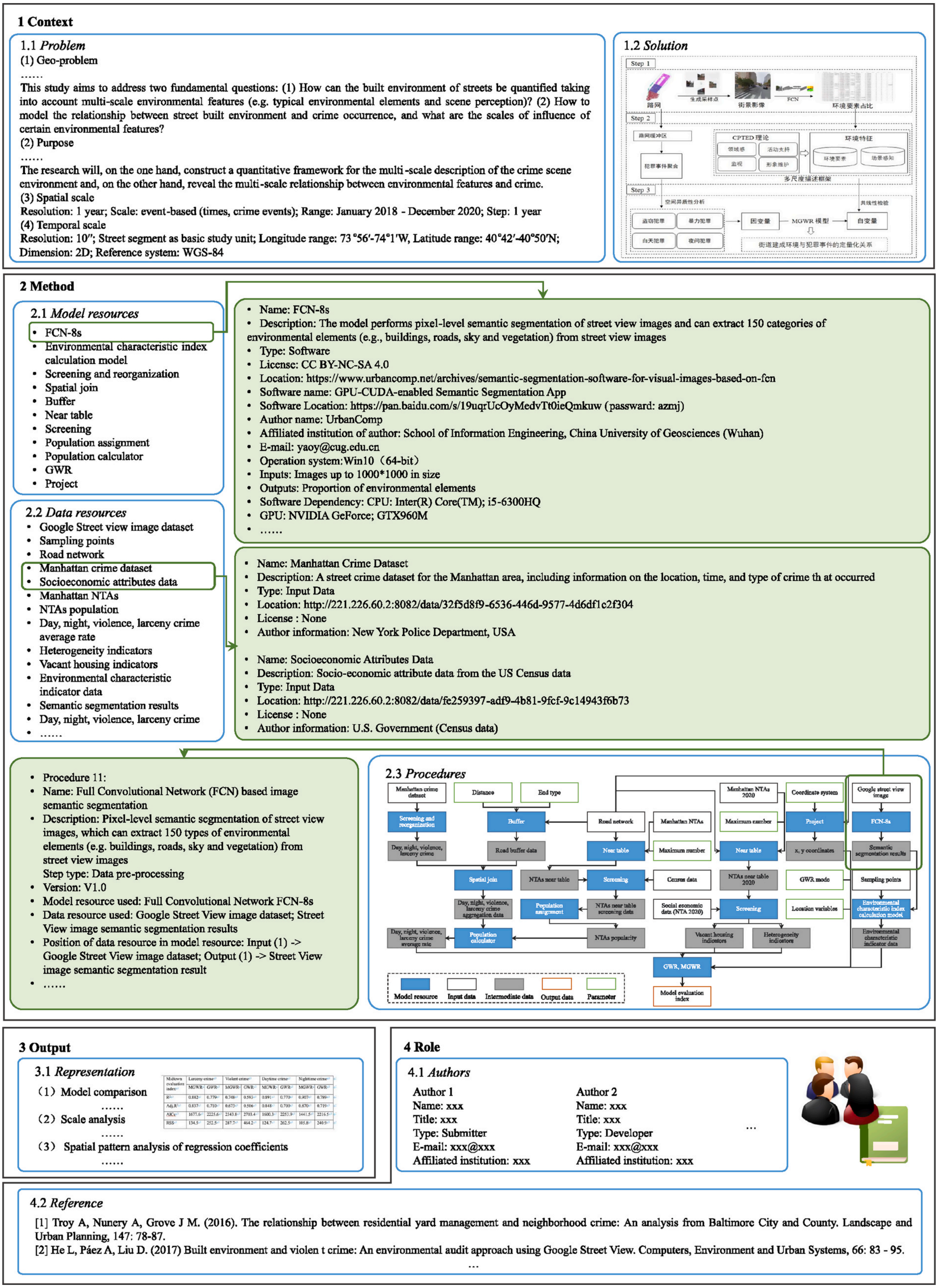

Fig. 6. A brief presentation of one case study.

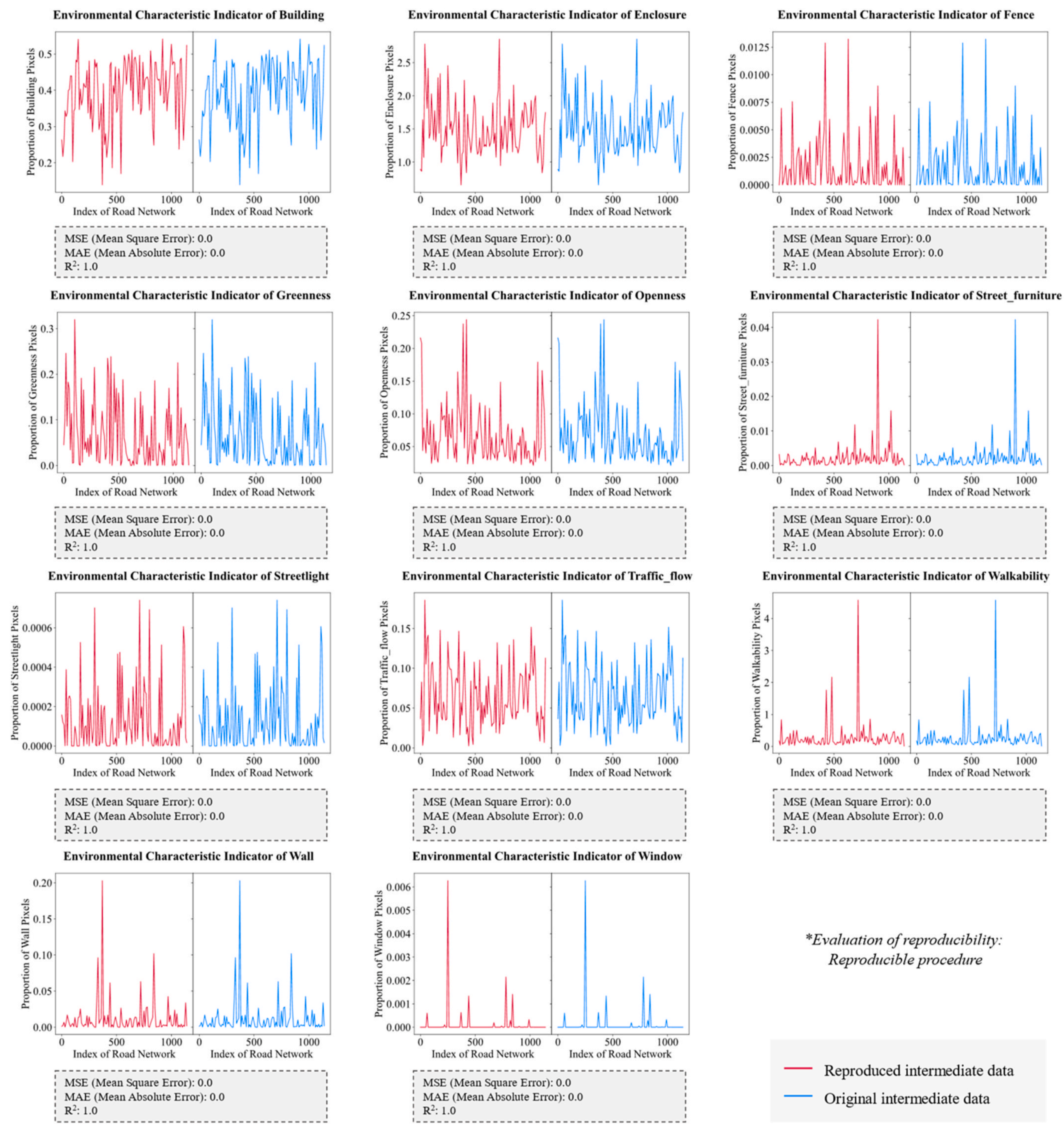

Fig. 7. Comparison of partial reproduced outputs and original outputs in *Procedure 11*.

established to represent all the procedures in a brief way.

Fifth, all the intermediate data, output, and analysis and interpretation of results are checked. Therefore, this study uses MSE, MAE, and R^2 to evaluate the consistency between the reproduced data and original data (provided by the authors). MSE and MAE are two typical indicators to evaluate the difference between reproduced and initial outputs (Hodson, 2022). R^2 is used to access the goodness of fit between reproduced outputs and the original outputs (Chicco et al., 2021). We found that these procedural intermediate data and outputs are exactly constant with the original ones. Some substantial outputs are shown in Fig. 7, and Table 4.

Finally, this GSE analyzes the raw outputs in Table 4, and the author reported conclusions in three aspects, that is, *comparison of GRW and MGRW*, *scale analysis*, and *spatial pattern analysis of regression coefficients*. In summary, through the development of the experimental case, on the one hand, the multiscale descriptive quantitative framework of crime scene environments is constructed; moreover, the results revealed the multiscale relationship between environmental characteristics and crime, answering the geo-problem mentioned above.

Table 4
Comparison of reproduced outputs and original outputs in this case study.

Midtown evaluation index	Larceny crime		Violent crime		Daytime crime		Nighttime crime	
	MGWR	GWR	MGWR	GWR	MGWR	GWR	MGWR	GWR
R^2	0.882	0.779	0.748	0.593	0.891	0.770	0.907	0.789
R^2[a]	0.882	0.779	0.748	0.593	0.891	0.770	0.907	0.789
Adj.R^2	0.837	0.710	0.673	0.506	0.848	0.700	0.870	0.719
Adj.R^2[a]	0.837	0.710	0.673	0.506	0.848	0.700	0.870	0.719
AICc	1677.6	2225.6	2343.8	2703.4	1600.3	2253.9	1441.5	2216.5
AICc[a]	1677.6	2225.6	2343.8	2703.4	1600.3	2253.9	1441.5	2216.5
RSS	134.5	252.5	287.7	464.2	124.7	262.5	105.8	240.9
RSS[a]	134.5	252.5	287.7	464.2	124.7	262.5	105.8	240.9

[a] Indicates that this indicator was reproduced by reviewers.

5. Conclusion and outlook

Keeping a notebook or document of the GSE lifecycle is a routine way to make a GSE reproducible. As mentioned earlier, major challenges to keeping a notebook include extensive speculation about essential information for keeping a notebook and the preference to record experiments in the geoscientist's own style. In this paper, we propose documentation strategy to help address these barriers through a concise and practical articulation of GSEDocument construction.

Despite some benefits, there is still considerable room for improvement of GSEs' reproducibility. This study next presents three recommendations for the future development of GSEs' reproducibility to make the proposed strategy more acceptable, useable, and comprehensive.

First, the reproducibility of a GSE is closely tied to the evaluation criteria. There are still many questions about how to evaluate an experiment is reproducible (Barba, 2018; Essawy et al., 2020; Goodman et al., 2016). When conforming to randomness of method, accuracy of data, manual error, and some other factors may influence the consistency of results, definition and evaluation criteria may have a deeper and more complex meaning.

Second, from our perspective, there are still some shortcomings of the proposed strategy. As with the ODD protocol, TRACE notebook, and Minimum Information Required in the Annotation of Biochemical Models (MIRIAM) (Ayllón et al., 2021; Evans, 1969; Grimm et al., 2014, 2020; Novère et al., 2005; Orchard et al., 2011), our study aims to learn from the experience of existing guidelines, as well as to reference and welcome geo-communities to provide feedback by contacting the lead author. We also hope that documentation strategy can be widely adopted by other scientists and the geo-community and ultimately galvanize discussion because each constituent entry in reference descriptions may change over time.

Finally, one important concern in implementing a documentation strategy is a lack of tools to facilitate the recording processes. Numerous tools already exist for problem formulation, resource distribution, metadata documentation, computational process provenance capture, and many other features described in this study (Cerutti et al., 2021; Freire and Silva, 2012; Pimentel et al., 2019; Steeves et al., 2020). These tools can be implemented as part of the proposed strategy to save time and effort. Another important concern is the demand for an online GSE-supportive database that allows geoscientists to query, evaluate, re-execute, and make new modifications to the original experiment. Although these technologies are not yet integrated effortlessly into GSEDocument practices, their integration should be considered in future research.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

We appreciate the detailed suggestions and comments from the editor and anonymous reviewers and express heartfelt thanks to the other members of the OpenGMS team. This work was supported by the National Natural Science Foundation of China [Grant No. 42071363] and Postgraduate Research & Practice Innovation Program of Jiangsu Province [Grant No. KYCX22_1566].

References

Asdal, K., Moser, I., 2012. Experiments in context and contexting. Sci. Technol. Hum. Val. 37, 291–306. https://doi.org/10.1177/0162243912449749.
Ayllón, D., Railsback, S.F., Gallagher, C., Augusiak, J., Baveco, H., Berger, U., Charles, S., Martin, R., Focks, A., Galic, N., Liu, C., van Loon, E.E., Nabe-Nielsen, J., Piou, C., Polhill, J.G., Preuss, T.G., Radchuk, V., Schmolke, A., Stadnicka-Michalak, J., Thorbek, P., Grimm, V., 2021. Keeping modelling notebooks with TRACE: good for you and good for environmental research and management support. Environ. Model. Software 136, 104932. https://doi.org/10.1016/j.envsoft.2020.104932.
Badham, J., Elsawah, S., Guillaume, J.H.A., Hamilton, S.H., Hunt, R.J., Jakeman, A.J., Pierce, S.A., Snow, V.O., Babbar-Sebens, M., Fu, B., Gober, P., Hill, M.C., Iwanaga, T., Loucks, D.P., Merritt, W.S., Peckham, S.D., Richmond, A.K., Zare, F., Ames, D., Bammer, G., 2019. Effective modeling for Integrated Water Resource Management: a guide to contextual practices by phases and steps and future opportunities. Environ. Model. Software 116, 40–56. https://doi.org/10.1016/j.envsoft.2019.02.013.
Baker, M., 2016. 1,500 scientists lift the lid on reproducibility. Nature 533, 452–454. https://doi.org/10.1038/533452a.
Balci, Osman, 1990. Guidelines for Successful Simulation Studies. Institute of Electrical and Electronics Engineers (IEEE).
Balci, O., 1994. Validation, verification, and testing techniques throughout the life cycle of a simulation study. Ann. Oper. Res. 53, 121–173.
Balci, O., 1998. In: Banks, J. (Ed.), Verification, Validation and Testing,"Handbook of Simulation (Ney York).
Balci, O., 2012. A life cycle for modeling and simulation. Simulation 88, 870–883. https://doi.org/10.1177/0037549712438469.
Barba, L.A., 2018. Terminologies for Reproducible Research arXiv:1802.03311 [cs].
Batty, M., 2011. Modeling and simulation in geographic information science: integrated models and grand challenges. Procedia - Social and Behavioral Sciences 21, 10–17. https://doi.org/10.1016/j.sbspro.2011.07.003.
Begley, C.G., Ioannidis, J.P.A., 2015. Reproducibility in science: improving the standard for basic and preclinical research. Circ. Res. 116, 116–126. https://doi.org/10.1161/CIRCRESAHA.114.303819.
Booth, W.C., Colomb, G.G., Williams, J.M., Bizup, J., FitzGerald, W.T., 2016. The craft of research. In: Chicago Guides to Writing, Editing, fourth ed. The University of Chicago Press, Chicago. and publishing.
Cerutti, V., Bellman, C., Both, A., Duckham, M., Jenny, B., Lemmens, R.L.G., Ostermann, F.O., 2021. Improving the reproducibility of geospatial scientific workflows: the use of geosocial media in facilitating disaster response. Spatial Sci. 66, 383–400. https://doi.org/10.1080/14498596.2019.1654944.
Chen, M., Yue, S., Lü, G., Lin, H., Yang, C., Wen, Y., Hou, T., Xiao, D., Jiang, H., 2019. Teamwork-oriented Integrated Modeling Method for Geo- Problem Solving. Environmental Modelling & Software, vol. 119, pp. 111–123. https://doi.org/10.1016/j.envsoft.2019.05.015.
Chen, M., Voinov, A., Ames, D.P., Kettner, A.J., Goodall, J.L., Jakeman, A.J., Barton, M. C., Harpham, Q., Cuddy, S.M., DeLuca, C., Yue, S., Wang, J., Zhang, F., Wen, Y., Lü, G., 2020. Position paper: open web-distributed integrated geographic modelling and simulation to enable broader participation and applications. Earth Sci. Rev. 207, 103223 https://doi.org/10.1016/j.earscirev.2020.103223.

Chen, M., Lv, G., Zhou, C., Lin, H., Ma, Z., Yue, S., Wen, Y., Zhang, F., Wang, J., Zhu, Z., Xu, K., He, Y., 2021a. Geographic modeling and simulation systems for geographic research in the new era: some thoughts on their development and construction. Sci. China Earth Sci. 64, 1207–1223. https://doi.org/10.1007/s11430-020-9759-0.

Chen, Y., Lin, H., Xiao, L., Jing, Q., You, L., Ding, Y., Hu, M., Devlin, A.T., 2021b. Versioned geoscientific workflow for the collaborative geo-simulation of human-nature interactions – a case study of global change and human activities. International Journal of Digital Earth 14, 510–539. https://doi.org/10.1080/17538947.2020.1849439.

Chicco, D., Warrens, M.J., Jurman, G., 2021. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Computer Science 7, e623. https://doi.org/10.7717/peerj-cs.623.

Chirigati, F., Troyer, M., Shasha, D., Freire, J., 2013. A computational reproducibility benchmark. Bulletin of the IEEE Computer Society Technical Committee on Data Engineering 36, 54–59.

Closa, G., Masó, J., Proß, B., Pons, X., 2017. W3C PROV to describe provenance at the dataset, feature and attribute levels in a distributed environment. Comput. Environ. Urban Syst. 64, 103–117. https://doi.org/10.1016/j.compenvurbsys.2017.01.008.

Committee on Reproducibility and Replicability in Science, 2019. Board on behavioral, cognitive, and sensory sciences, committee on national statistics, division of behavioral and social sciences and education, nuclear and radiation studies board, division on earth and life studies, board on mathematical sciences and analytics, committee on applied and theoretical statistics, division on engineering and physical sciences, board on research data and information, committee on science, engineering, medicine, and public policy, policy and global affairs, national academies of sciences, engineering, and medicine. Reproducibility and Replicability in Science. https://doi.org/10.17226/25303. National Academies Press, Washington, D.C.

Davison, A., 2012. Automated capture of experiment context for easier reproducibility in computational research. Comput. Sci. Eng. 14, 48–56. https://doi.org/10.1109/MCSE.2012.41.

DellaVigna, S., Pope, D., Vivalt, E., 2019. Predict science to improve science. Science 366, 428–429. https://doi.org/10.1126/science.aaz1704.

DeRisi, S., Kennison, R., Twyman, N., 2003. The what and whys of DOIs. PLoS Biol. 1, e57. https://doi.org/10.1371/journal.pbio.0000057.

Dirnagl, U., Przesdzing, I., 2016. A pocket guide to electronic laboratory notebooks in the academic life sciences. F1000Res 5, 2. https://doi.org/10.12688/f1000research.7628.1.

Easterbrook, S.M., 2014. Open code for open science? Nat. Geosci. 7, 779–781. https://doi.org/10.1038/ngeo2283.

Essawy, B.T., Goodall, J.L., Voce, D., Morsy, M.M., Sadler, J.M., Choi, Y.D., Tarboton, D. G., Malik, T., 2020. A taxonomy for reproducible and replicable research in environmental modelling. Environ. Model. Software 134, 104753. https://doi.org/10.1016/j.envsoft.2020.104753.

Evans, R.A., 1969. The principle of Minimum information. IEEE Transactions on Reliability R- 18, 87–90. https://doi.org/10.1109/TR.1969.5216992.

Ferro, N., Fuhr, N., Jarvelin, K., Kando, N., Lippold, M., Zobel, J., 2016. Increasing reproducibility in IR: findings from the dagstuhl seminar on "reproducibility of data-oriented experiments in e-science. ACM SIGIR Forum 50, 15.

Fitzgerald, 2006. The transformation of open source software. MIS Q. 30, 587. https://doi.org/10.2307/25148740.

Fotheringham, A.S., Sachdeva, M., 2022. Modelling spatial processes in quantitative human geography. Spatial Sci. 28, 5–14. https://doi.org/10.1080/19475683.2021.1903996.

Freire, J., Silva, C.T., 2012. Making computations and publications reproducible with VisTrails. Comput. Sci. Eng. 14, 18–25. https://doi.org/10.1109/MCSE.2012.76.

Freire, J., Bonnet, P., Shasha, D., 2012. Computational reproducibility: state-of-the-art, challenges, and database research opportunities. In: Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data, pp. 593–596.

Gil, Y., David, C.H., Demir, I., Essawy, B.T., Fulweiler, R.W., Goodall, J.L., Karlstrom, L., Lee, H., Mills, H.J., Oh, J., Pierce, S.A., Pope, A., Tzeng, M.W., Villamizar, S.R., Yu, X., 2016. Toward the Geoscience Paper of the Future: best practices for documenting and sharing research from data to software to provenance. Earth Space Sci. 3, 388–415. https://doi.org/10.1002/2015EA000136.

Goodchild, M.F., Li, W., 2021. Replication across space and time must be weak in the social and environmental sciences. Proc. Natl. Acad. Sci. USA 118, e2015759118. https://doi.org/10.1073/pnas.2015759118.

Goodchild, M.F., Fotheringham, A.S., Kedron, P., Li, W., 2021. Introduction: forum on reproducibility and replicability in geography. Ann. Assoc. Am. Geogr. 111, 1271–1274. https://doi.org/10.1080/24694452.2020.1806030.

Goodman, S.N., Fanelli, D., Ioannidis, J.P.A., 2016. What does research reproducibility mean? Sci. Transl. Med. 8 https://doi.org/10.1126/scitranslmed.aaf5027, 341ps12-341ps12.

Grimm, V., Augusiak, J., Focks, A., Frank, B.M., Gabsi, F., Johnston, A.S.A., Liu, C., Martin, B.T., Meli, M., Radchuk, V., Thorbek, P., Railsback, S.F., 2014. Towards better modelling and decision support: documenting model development, testing, and analysis using TRACE. Ecol. Model. 280, 129–139. https://doi.org/10.1016/j.ecolmodel.2014.01.018.

Grimm, V., Railsback, S.F., Vincenot, C.E., Berger, U., Gallagher, C., DeAngelis, D.L., Edmonds, B., Ge, J., Giske, J., Groeneveld, J., Johnston, A.S.A., Milles, A., Nabe-Nielsen, J., Polhill, J.G., Radchuk, V., Rohwäder, M.-S., Stillman, R.A., Thiele, J.C., Ayllón, D., 2020. The ODD protocol for describing agent-based and other simulation models: a second update to improve clarity, replication, and structural realism. J. Artif. Soc. Soc. Simulat. 23, 7. https://doi.org/10.18564/jasss.4259.

Groth, P., Frew, J. (Eds.), 2012. Provenance and Annotation of Data and Processes. Lecture Notes in Computer Science. Springer Berlin Heidelberg, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-34222-6.

Hodson, T.O., 2022. Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not. Geosci. Model Dev. (GMD) 15, 5481–5487. https://doi.org/10.5194/gmd-15-5481-2022.

Hucka, M., Finney, A., Sauro, H.M., Bolouri, H., Doyle, J.C., Kitano, H., , the rest of the SBML Forum, Arkin, A.P., Bornstein, B.J., Bray, D., Cornish-Bowden, A., Cuellar, A. A., Dronov, S., Gilles, E.D., Ginkel, M., Gor, V., Goryanin, I.I., Hedley, W.J., Hodgman, T.C., Hofmeyr, J.-H., Hunter, P.J., Juty, N.S., Kasberger, J.L., Kremling, A., Kummer, U., Le Novere, N., Loew, L.M., Lucio, D., Mendes, P., Minch, E., Mjolsness, E.D., Nakayama, Y., Nelson, M.R., Nielsen, P.F., Sakurada, T., Schaff, J.C., Shapiro, B.E., Shimizu, T.S., Spence, H.D., Stelling, J., Takahashi, K., Tomita, M., Wagner, J., Wang, J., 2003. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. Bioinformatics 19, 524–531. https://doi.org/10.1093/bioinformatics/btg015.

Ioannidis, J.P.A., 2005. Why most published research findings are false. PLoS Med. 2, e124. https://doi.org/10.1371/journal.pmed.0020124.

Juty, N., Wimalaratne, S.M., Soiland-Reyes, S., Kunze, J., Goble, C.A., Clark, T., 2020. Unique, persistent, resolvable: identifiers as the foundation of FAIR. Data Intellegence 2, 30–39. https://doi.org/10.1162/dint_a_00025.

Kluyver, T., Ragan-Kelley, B., Pérez, F., Bussonnier, M., Frederic, J., Hamrick, J., Grout, J., Corlay, S., Ivanov, P., Abdalla, S., Willing, C., 2016. Jupyter Notebooks—A Publishing Format for Reproducible Computational Workflows 87–90. https://doi.org/10.3233/978-1-61499-649-1-87.

Konkol, M., Kray, C., Pfeiffer, M., 2019. Computational reproducibility in geoscientific papers: insights from a series of studies with geoscientists and a reproduction study. Int. J. Geogr. Inf. Sci. 33, 408–429. https://doi.org/10.1080/13658816.2018.1508687.

Li, W., Zhu, J., Fu, L., Zhu, Q., Guo, Y., Gong, Y., 2021a. A rapid 3D reproduction system of dam-break floods constrained by post-disaster information. Environ. Model. Software 139, 104994. https://doi.org/10.1016/j.envsoft.2021.104994.

Li, W., Zhu, J., Fu, L., Zhu, Q., Xie, Y., Hu, Y., 2021b. An augmented representation method of debris flow scenes to improve public perception. Int. J. Geogr. Inf. Sci. 35, 1521–1544. https://doi.org/10.1080/13658816.2020.1833016.

Lü, G., Batty, M., Strobl, J., Lin, H., Zhu, A.-X., Chen, M., 2019. Reflections and speculations on the progress in Geographic Information Systems (GIS): a geographic perspective. Int. J. Geogr. Inf. Sci. 33, 346–367. https://doi.org/10.1080/13658816.2018.1533136.

Ma, Z., Chen, M., Yue, S., Zhang, B., Zhu, Z., Wen, Y., Lü, G., Lu, M., 2021. Activity-based process construction for participatory geo-analysis. GIScience Remote Sens. 58, 180–198. https://doi.org/10.1080/15481603.2020.1868211.

Ma, Z., Chen, M., Zheng, Z., Yue, S., Zhu, Z., Zhang, B., Wang, J., Zhang, F., Wen, Y., Lü, G., 2022. Customizable process design for collaborative geographic analysis. GIScience Remote Sens. 59, 914–935. https://doi.org/10.1080/15481603.2022.2082751.

McNutt, M., 2014. Reproducibility. Science 343. https://doi.org/10.1126/science.1250475, 229–229.

Mingers, J., 2000. Variety is the spice of life: combining soft and hard OR/MS methods. Int. Trans. Oper. Res. 7, 673–691. https://doi.org/10.1111/j.1475-3995.2000.tb00224.x.

Novère, N.L., Finney, A., Hucka, M., Bhalla, U.S., Campagne, F., Collado-Vides, J., Crampin, E.J., Halstead, M., Klipp, E., Mendes, P., Nielsen, P., Sauro, H., Shapiro, B., Snoep, J.L., Spence, H.D., Wanner, B.L., 2005. Minimum information requested in the annotation of biochemical models (MIRIAM). Nat. Biotechnol. 23, 1509–1515. https://doi.org/10.1038/nbt1156.

Open Science Collaboration, 2015. Estimating the reproducibility of psychological science. Science 349. https://doi.org/10.1126/science.aac4716 aac4716–aac4716.

Orchard, S., Al-Lazikani, B., Bryant, S., Clark, D., Calder, E., Dix, I., Engkvist, O., Forster, M., Gaulton, A., Gilson, M., Glen, R., Grigorov, M., Hammond-Kosack, K., Harland, L., Hopkins, A., Larminie, C., Lynch, N., Mann, R.K., Murray-Rust, P., Lo Piparo, E., Southan, C., Steinbeck, C., Wishart, D., Hermjakob, H., Overington, J., Thornton, J., 2011. Minimum information about a bioactive entity (MIABE). Nat. Rev. Drug Discov. 10, 661–669. https://doi.org/10.1038/nrd3503.

Peng, R.D., Dominici, F., Zeger, S.L., 2006. Reproducible epidemiologic research. Am. J. Epidemiol. 163, 783–789. https://doi.org/10.1093/aje/kwj093.

Phillips, J.D., 2022. The law of scale independence. Spatial Sci. 28, 15–29. https://doi.org/10.1080/19475683.2022.2026466.

Piccolo, S.R., Frampton, M.B., 2016. Tools and techniques for computational reproducibility. GigaScience 5, 30. https://doi.org/10.1186/s13742-016-0135-4.

Pimentel, J.F., Murta, L., Braganholo, V., Freire, J., 2019. A large-scale study about quality and reproducibility of jupyter notebooks. In: 2019 IEEE/ACM 16th International Conference on Mining Software Repositories (MSR). Presented at the 2019 IEEE/ACM 16th International Conference on Mining Software Repositories (MSR). IEEE, Montreal, QC, Canada, pp. 507–517. https://doi.org/10.1109/MSR.2019.00077.

Poldrack, R.A., 2019. The costs of reproducibility. Neuron 101, 11–14. https://doi.org/10.1016/j.neuron.2018.11.030.

Qian, Z., Liu, X., Tao, F., Zhou, T., 2020. Identification of urban functional areas by coupling satellite images and taxi GPS trajectories. Rem. Sens. 12, 2449. https://doi.org/10.3390/rs12152449.

Qian, Z., Chen, M., Yang, Y., Zhong, T., Zhang, F., Zhu, R., Zhang, K., Zhang, Z., Sun, Z., Ma, P., Lü, G., Ye, Y., Yan, J., 2022a. Vectorized dataset of roadside noise barriers in China using street view imagery. Earth Syst. Sci. Data 14, 4057–4076. https://doi.org/10.5194/essd-14-4057-2022.

Qian, Z., Chen, M., Zhong, T., Zhang, F., Zhu, R., Zhang, Z., Zhang, K., Sun, Z., Lü, G., 2022b. Deep Roof Refiner: a detail-oriented deep learning network for refined delineation of roof structure lines using satellite imagery. Int. J. Appl. Earth Obs. Geoinf. 107, 102680 https://doi.org/10.1016/j.jag.2022.102680.

Rabe, M., Spieckermann, S., Wenzel, S., 2009. Verification and validation activities within a new procedure model for V&V in production and logistics simulation. In: Proceedings of the 2009 Winter Simulation Conference (WSC). Presented at the 2009 Winter Simulation Conference - (WSC 2009). IEEE, Austin, TX, USA, pp. 2509–2519. https://doi.org/10.1109/WSC.2009.5429641.

Ruscheinski, A., Warnke, T., Uhrmacher, A.M., 2020. Artifact-based workflows for supporting simulation studies. IEEE Trans. Knowl. Data Eng. 32, 15.

Sacks, J., Welch, W.J., Mitchell, T.J., Wynn, H.P., 1989. Design and analysis of computer experiments. Stat. Sci. 4 https://doi.org/10.1214/ss/1177012413.

Sandve, G.K., Nekrutenko, A., Taylor, J., Hovig, E., 2013. Ten simple rules for reproducible computational research. PLoS Comput. Biol. 9, e1003285 https://doi.org/10.1371/journal.pcbi.1003285.

Sargent, R.G., 2010. Verification and validation of simulation models. In: Proceedings of the 2010 Winter Simulation Conference. Presented at the 2010 Winter Simulation Conference - (WSC 2010). IEEE, Baltimore, MD, USA, pp. 166–183. https://doi.org/10.1109/WSC.2010.5679166.

Song, C., Zhang, G., Cheng, C., Chen, F., 2020. Nature and basic issues of Geography. Sci. Geogr. Sin. 40, 6–11.

Steeves, V., Rampin, R., Chirigati, F., 2020. Reproducibility, preservation, and access to research with ReproZip and ReproServer. IASSIST Q. 44, 1–11. https://doi.org/10.29173/iq969.

Stodden, V., McNutt, M., Bailey, D.H., Deelman, E., Hanson, B., Heroux, M.A., Ioannidis, J.P.A., Taufer, M., 2018. Enhancing reproducibility for computational methods. Science 354, 1240–1241.

Tsai, W.-P., Feng, D., Pan, M., Beck, H., Lawson, K., Yang, Y., Liu, J., Shen, C., 2021. From calibration to parameter learning: harnessing the scaling effects of big data in geoscientific modeling. Nat. Commun. 12, 5988. https://doi.org/10.1038/s41467-021-26107-z.

Vasilevsky, N.A., Brush, M.H., Paddock, H., Ponting, L., Tripathy, S.J., LaRocca, G.M., Haendel, M.A., 2013. On the reproducibility of science: unique identification of research resources in the biomedical literature. PeerJ 1, e148. https://doi.org/10.7717/peerj.148.

Wilkinson, M.D., Dumontier, M., Aalbersberg, IjJ., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., 't Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B., 2016. The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3, 160018. https://doi.org/10.1038/sdata.2016.18.

Wolke, A., Bichler, M., Chirigati, F., Steeves, V., 2016. Reproducible experiments on dynamic resource allocation in cloud data centers. Inf. Syst. 59, 98–101. https://doi.org/10.1016/j.is.2015.12.004.

Yue, S., Wen, Y., Chen, M., Lu, G., Hu, D., Zhang, F., 2015. A data description model for reusing, sharing and integrating geo-analysis models. Environ. Earth Sci. 74, 7081–7099. https://doi.org/10.1007/s12665-015-4270-5.

Zhang, M., Jiang, L., Zhao, J., Yue, P., Zhang, X., 2020. Coupling OGC WPS and W3C PROV for provenance-aware geoprocessing workflows. Comput. Geosci. 138, 104419 https://doi.org/10.1016/j.cageo.2020.104419.

Zhang, F., Chen, M., Kettner, A.J., Ames, D.P., Harpham, Q., Yue, S., Wen, Y., Lü, G., 2021. Interoperability engine design for model sharing and reuse among OpenMI, BMI and OpenGMS-IS model standards. Environ. Model. Software 144, 105164. https://doi.org/10.1016/j.envsoft.2021.105164.

Zhang, Z., Qian, Z., Zhong, T., Chen, M., Zhang, K., Yang, Y., Zhu, R., Zhang, F., Zhang, H., Zhou, F., Yu, J., Zhang, B., Lü, G., Yan, J., 2022. Vectorized rooftop area data for 90 cities in China. Sci. Data 9, 66. https://doi.org/10.1038/s41597-022-01168-x.