Map Reproducibility in Geoscientific Publications: An Exploratory Study

Eftychia Koukouraki[1]
Insitute for Geoinformatics, University of Münster, Germany

Christian Kray
Insitute for Geoinformatics, University of Münster, Germany

Abstract

Reproducibility is a core element of the scientific method. In the Geosciences, the insights derived from geodata are frequently communicated through maps, and the computational methods to create these maps vary in their ease of reproduction. In this paper, we present the results from a study where we tried to reproduce the maps included in geoscientific publications. Following a systematic approach, we collected 27 candidate papers and in four cases, we were able to successfully reproduce the maps they contained. We report on the approach we applied, the issues we encountered and the insights we gained while attempting to reproduce the maps. In addition, we provide an initial set of criteria to assess the success of a map reproduction attempt. We also propose some guidelines for improving map reproducibility in geoscientific publications. Our work sheds a light on the current state of map reproducibility in geoscientific papers and can benefit researchers interested in publishing maps in a more reproducible way.

2012 ACM Subject Classification Human-centered computing → Geographic visualization; Applied computing → Cartography

Keywords and phrases Reproducible Research, Reproduction Assessment, Map Making, Cartography

Digital Object Identifier 10.4230/LIPIcs.GIScience.2023.6

Supplementary Material *Collection (Papers)*: `https://doi.org/10.17605/OSF.IO/P5V2Z`

Acknowledgements We would like to thank the authors who took the time to respond to our queries and helped us to overcome issues we faced in reproducing maps and the anonymous reviewers for their constructive feedback.

1 Introduction

The reproducibility of research results is a critical aspect across all scientific disciplines, and the domain of Geosciences is no exception. It is widely accepted by the scientific community that the ability to reproduce results by other working groups enhances the trust and the reliability of the respective research. In an effort to clarify the different existing terminologies for *reproducibility* and *reproducible research*, Barba [1] defines reproducible research as the case when "authors provide all the necessary data and the computer codes to run the analysis again, re-creating the results", while scientific *replication* can be achieved when a study produces the same results using different methods or different data. This distinction is also known as the Claerbout [3]/ Donoho [4]/ Peng [28] convention.

In recent years, reproducibility has rapidly gained relevance in different areas of Geosciences. In the *Forum on Reproducibility and Replicability in Geography* introduced by Goodchild et al. [8], the discussed topics range from the theoretical dimension of reproducibility and replicability in Geography [34], to more tangible matters, such as the review of current technological solutions [25] in this context.

[1] Corresponding Author

12th International Conference on Geographic Information Science (GIScience 2023).
Editors: Roger Beecham, Jed A. Long, Dianna Smith, Qunshan Zhao, and Sarah Wise; Article No. 6; pp. 6:1–6:16
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Figures are key elements of geoscientific publications and play an important role in the dissemination of scientific findings. Konkol and Kray [15] focused on the role of figures as a means of communicating research results and conducted a survey to identify the nature of the incorporated figures (e.g., maps, time series, histograms, etc.), as well as the frequency that they are used in geoscientific papers. The survey showed that maps were most frequently mentioned, which seems reasonable for this domain. Although there are several definitions for what a map is, most of them agree that maps are abstract representations of the real world [18], in which certain attributes of reality are highlighted, while others are deemphasised or completely left out. In line with fundamental scientific principles and the core role maps play in communicating societally relevant scientific outcomes (e.g., climate change, epidemiological spread), it is essential to be able to reproduce them. This capacity to reproduce results builds trust [10], which is essential for the uptake of scientific outcomes in society.

The aim of the research reported in this paper is thus to investigate the reproducibility of maps that are published as part of scientific publications. For this purpose, we screened 27 open access papers from the Geosciences and tried to reproduce the maps contained therein. In doing so, our objectives were (i) to investigate the current practice of creating maps for scientific publications, (ii) to assess the material availability, (iii) to analyse the complexities of map reproduction, and (iv) determine the degree of reproducibility of recently published maps in geoscientific articles. Our findings show that reproducing maps from recent papers is rarely easy, frequently requires extensive efforts and can even fail despite data and code being available. The fact that currently there are no well-defined success criteria for map reproduction further hampers it. Our main contributions are an initial set of criteria for assessing the success of map reproduction, insights into challenges arising during map reproduction and a set of guidelines for making map reproduction easier. These contributions pave the way for further research into map reproduction and can help researchers in making the maps they publish more reproducible.

The rest of the paper is structured as follows. Section 2 contextualises our work, reviews related studies and initiatives, and defines the term *reproducible map making*. In section 3, we motivate and describe the methodological workflow that we followed for the reproduction study. Section 4 presents the results of the study, including the obstacles that were encountered and the reproduced maps next to the original ones. Section 5 discusses the insights that were revealed and proposes a set of initial guidelines for making maps more reproducible. Section 6 summarises our key findings and outlines future work.

2 Background

In this section, we briefly summarise related work on map creation, open science practices, reproducibility in the Geosciences in general and in map production in particular.

2.1 Map Production

The International Cartographic Association (ICA) defines Cartography as the science, art, and technology of map making and map use [17]. Taylor [35] describes cartography in the context of Geographic Information Systems (GIS) as “the organization, presentation, communication and utilization of geo-information in graphic, digital or tactile form. It can include all stages from data preparation to end use in the creation of maps and related spatial information products”. In the context of map reproduction, it is important to note that this definition also takes into account the technological aspect in the cartographic process and highlights the importance of it in the creation of spatial information products beyond maps.

According to MacEachren [20], apart from their obvious role of visualising geodata, maps also serve as interfaces to the underlying computations, connect human reasoning to complex sources of information and facilitate the understanding of spatial relations. The current map making practices in academia use various software products for different tasks, from spreadsheets to specialized statistical software to specialised cartographic packages and geographic information systems. Frequently, the cartographic process is broken down into multiple steps [6], which are not always connected in a straightforward way. These steps can include data management, statistical analyses, geoprocessing and graphical display of the results. While there is a lot of variability involved – in particular in a scientific context, each step needs to be validated for its objectivity and ideally should be reproducible [7].

2.2 Open Science Practices

A fundamental requirement for enabling the reproduction of maps or computational workflows in general is access to all components necessary to carry out the reproduction (e.g., data, code, instructions) [23]. The central importance of the availability of research materials (especially data) is reflected in the several initiatives to standardise the way in which they are publicly shared. The FAIR Data Principles are guidelines for making data Findable, Accessible, Interoperable, and Reusable and have received considerable support by many scientific communities [36]. The Force11 community [2] also introduced the Principles for Data [9] and for Software Citation [33], acknowledging the significance of making these assets accessible. These principles are slowly being adopted by publishers as well.The publisher Copernicus Publications[2], for example, enforces the citation of data and encourages the citation of software, referencing the aforementioned principles.

The Transparency and Openness Promotion (TOP) Committee defined a set of standards with increasing levels of stringency to facilitate and motivate an open culture in scholarly communication [22]. These standards, namely the TOP Guidelines, deal with the topics of data, code and other materials citation and sharing, as well as the transparency of study design (among others). They indicate the extent to which a journal considers them as a requirement for publishing. Based on the TOP Guidelines, the Center for Open Science uses the TOP Factor[3] rating to rank journals [31]. Journals from the disciplines of Geography, Planning and Development, and Earth and Planetary Sciences score seven at maximum in this scale (Cartography and Geographic Information Science, Nature Geoscience, Nature Sustainability), while the highest achieved score across all disciplines is 27.

2.3 Reproducible Research in the Geosciences

Researchers have started to address the issue of reproducibility in the Geosciences and have recognised that open data and methods and open source software are prerequisites for achieving the full potential in transparency [24]. However, Ledermann and Gartner [19] argue that acquiring the source code of an experiment is not enough to reproduce it and rather argue that a well defined and clearly structured programming workflow in an ontological fashion contributes to the transparency, reproducibility and extensibility of scientific experiments. Giraud and Lambert [6] encourage the use of literate programming reports, e.g., Jupyter notebooks or R Markdown, because such programming solutions provide complete instructions from raw data to the cartographic product. In a later work, the same

[2] `https://publications.copernicus.org/`
[3] `https://topfactor.org/`

authors [7] provide an example of reproducible cartographic workflow, which is implemented by combining different geospatial R packages in a single R Markdown script. Knoth and Nüst [14] leverage containers for describing the computational environment of the experiment and for facilitating the reuse of their workflow by others. In more complex visualisation cases, such as terrain representation, Kennelly et al. [13] commented on the usefulness of 3D reference models for evaluating and comparing reproduced visualizations and stated that in the field of cartography such data models do not exist.

Reproducing and replicating visualizations in general is not an easy task. Fekete and Freire [5] consider interactive features as an additional complexity for scientific replication, but emphasize the importance of interactivity in data exploration. A data-centric approach for reproducible visualizations has been proposed by Silva et al. [32], where the authors highlight the significance of a well-defined data flow pipeline. Irrespective of the implementation tools and description mechanisms of the data flow, the purpose of the visualization process is always to gain insights from the data.

2.4 Reproducible Map Making

Even though reproducibility is a topic of rising concern for the scientific community, the term "reproducible cartography" is not frequently encountered in the literature. Giraud and Lambert [6] place reproducible maps on a spectrum, ranging from non-reproducible, when they come as a simple print-out, to fully reproducible, when they are clearly linked with executable code, data and metadata. Although many suggestions have been made regarding the enhancement of map reproducibility in different contexts (terrain representation [13], knowledge graphs [21], etc.), core terms such as "reproducible cartography" and "map reproducibility" so far have not been clearly specified, let alone formally defined.

Since the widely used ICA definition of cartography as a discipline also includes an artistic dimension [17], we prefer to use the terms "map production" and "map making" in the context of this paper and map reproducibility in general. Based on the definition of Barba [1] for reproducible research outlined in Section 1 and on the definition of Taylor [35] for cartography, we propose the following definition:

> **Reproducible map production/making** refers to the provision, organisation, and processing of all materials used in the cartographic process in such a way that a map as a cartographic product can be recreated in an independent experiment.

This process thus aims to create a visual copy of the original map without introducing any significant variations that alter the maps's interpretation. Although the term *map* can refer to a variety of representational artifacts, in the context of this work we focus solely on maps with an apparent geographical reference.

3 Methodological Approach

The main goal of our approach was to assess the degree to which map making practices in scientific publications allow for successful reproduction of the included maps. For this purpose, we collected a set of papers and subsequently attempted to reproduce their map figures. During this process, we screened which tools and programming languages were commonly used for map creation. We also collected information about the effort we had to invest during our reproduction attempts. All the reproductions were run on laptop running Linux Ubuntu 20.04 with 16 GB RAM and an i7-1185G7 @ 3.00GHz with 8 cores.

3.1 Creating the Paper Collection

Selecting a good sample of publications for our study was not an easy task due to the diversity of geoscientific papers and the maps used therein. In order to capture current practices from a broad range of journals, we limited our search to recent articles that were available via Open Access (OA) and were published within the 12 months prior to the beginning of our study (June 2022). The selection workflow that we eventually implemented is illustrated in Figure 1. Although maps can in principle be a part of any paper that reports on research in a geographical context, our investigation targeted journals with an explicit geographical or geoscientific scope. Aligning with the findings outlined in Subsection 2.2, we considered journals that support either the FAIR Data Principles or comply with the TOP Guidelines for Data, Code and Materials Transparency in order to maximise our chances of obtaining all the necessary components for map reproduction. In addition, eligible journals should require a data and/or software availability statement for publishing.

After selecting a journal, the research articles and data description papers were screened chronologically starting from the most recently published and working backwards until the cut-off date (June 2021). More recent works were preferred assuming that the related materials, namely code and data, are more probable to be available and the authors' contacts up-to-date when we ran the study. We excluded technical reports, briefings and reviews to ensure a consistent body of core scientific publications. In order for an article to qualify as suitable candidate for reproduction, we also required it to be OA so that anyone would be able to reproduce the illustrated maps. The next step was scanning the articles to confirm that geographical maps are included. Finally, the data and software availability statement had to clearly mention sources for acquiring the data and the code. If all the aforementioned conditions were met, we added this paper as a reproduction candidate to the list of papers we tried to reproduce.

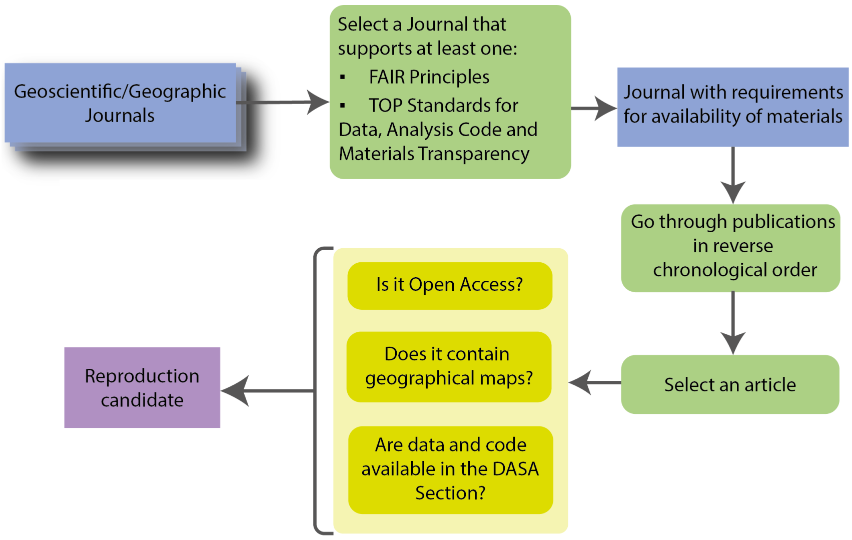


Figure 1 Flowchart of the paper selection process.

3.2 Reproduction Protocol

The procedure that we followed for the actual map reproduction is shown in Figure 2. The process started by reading and following the instructions in the availability statement to obtain the datasets and the software that were used in the paper analysis. This first stage could involve several different steps. In the simplest scenario, the statement contained links where datasets and software could be directly downloaded. The data was found either in a repository maintained by the authors or in a public data portal such as the Copernicus Climate Data Store[4]. In the first case, the data was usually (pre)processed to some extent by the authors, while in the latter, the reproducing researcher had to perform all the data processing stages again. In some cases, the statement briefly explained why the datasets cannot be publicly shared, optionally encouraging the reader to get in contact with the authors or with a third person that is responsible for the distribution. The code was shared in the same repository as the data or in a different one, for example a public website specialising in hosting code such as Github[5]. The repositories occasionally contained directions for connecting the different components to setup the computational environment and re-run the map generation process. After collecting all the necessary materials, we identified the datasets and the parts of the code that corresponded to the creation of every map in the paper. This step required a deeper understanding of the paper and its maps. In particular, this involved reading the figure-related parts of the paper carefully to comprehend the map's purpose in the paper (e.g., exploratory data analysis, presenting results, placeholder map for GUI demonstration) and to determine its role in the computational pipeline.

As a next step, we then attempted to setup the computational environment by installing the required applications, libraries and plugins. If the attempt failed, we investigated the underlying issues and tried to solve them. Potential reasons for failure at this stage were missing data and unclear or incomplete instructions. When all the components were properly assembled and the computing pipeline was set up, we ran the analysis. We considered this step successful if it led to a map output. If running the analysis failed, we repeated the troubleshooting process. Possible reasons could be again missing data files, different library versions and different data processing results in preceding steps.

As a final step in the reproduction process, we assessed how well the generated map reproduced the map published in the original paper. For this purpose, we used a set of assessment criteria, which are described in Subsection 3.2.1. If a generated map was considered a successful reproduction according to these criteria, we added it to the list of successfully reproduced maps. Otherwise, we investigated the reasons for failure and attempted to address them until we ran out of options. If during the entire reproduction process information was unclear, we first tried to infer it from information in the paper and on from online sources, and then contacted the corresponding author to ask for their help. In general, we minimized the amount of intervention in the map production process and avoided adding data transformation steps or visualisation adjustments that were not explicitly stated by the authors of the paper.

3.2.1 Assessing Reproduction Success

To confirm if a map reproduction was successful, initially we used the utility `compare` from the package `ImageMagick`[6] to perform this image-based comparison, but the differences in image resolution, file types and file sizes led to quite large deltas between the original and the reproduced maps, even when they were visually and/or semantically very similar. This was partially due to the fact that the scripts frequently did not generate the same file types

[4] https://cds.climate.copernicus.eu/
[5] https://github.org
[6] https://imagemagick.org/

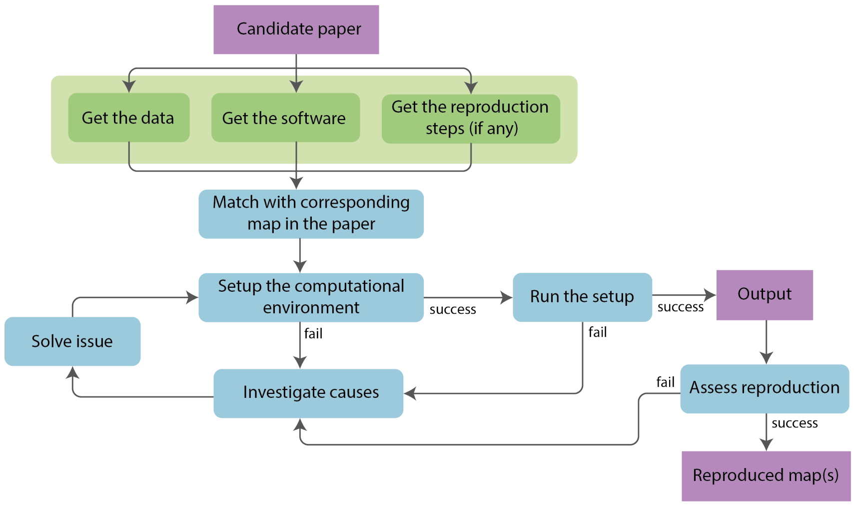


Figure 2 Flowchart of the reproduction process.

as the ones we found in the embedded figures on the publisher's website. For this reason, we avoided the use of any tool for our final assessment and visually compared the reproduced figures with the original ones in a side-to-side fashion.

It is important to emphasise that even when two maps (the original and the reproduced one) are not exactly identical, they still can manage to convey the same message. Any criteria for assessing map reproduction should therefore go beyond simple pixel-based comparison. Based on these considerations, previous work on the reproducibility of figures in scientific publications [16] and the experiences gained during our reproduction study, we used a three by two matrix to assess the success of map reproduction (see Table 1). In this matrix, we categorised the observed differences based on the map element which they concern: the map body, the legend and other elements. The *map body* refers to the actual depiction of a geographic region in the map, e.g. a visual abstraction of a country. The *legend* contains complementary information that define the meaning of entities visualised in the map body. This includes, for example, a table explaining the meaning of symbols shown in the map body. All other visual elements, which are neither part of the map body nor the legend, were grouped under the *other elements* category. This includes, e.g., the North arrow, scale bars, or map titles.

For each of the three categories, we distinguish between two types of differences: aesthetic and semantic ones. *Aesthetic differences* refer to the way in which the map elements are visually expressed or styled, e.g., the colour used to depict certain aspects or which font was used. *Semantic differences* include any changes that affect the substance of the map, e.g., the absence of relevant elements or factual differences. When assessing whether a map reproduction was successful, we used the matrix to classify any differences we observed. If the observed differences were aesthetic in nature and were consistent with the context of the paper, we considered the reproduction a success. For example, a different colour scheme in the reproduced map is an aesthetic difference and, if reflected in the legend as well, it does not break the success of the reproduction. If not, though, it constitutes a semantic difference and the map reproduction does not qualify as successful. It should be mentioned that the success of map reproduction also depends on the purpose, the context and the target audience of the initial map.

Table 1 Map elements and categories used to assess reproduction success with general examples (in italics) and examples from reproduction study (in bold); numbers in brackets indicate how often a difference type occurred during our reproduction study.

	Aesthetic	**Semantic**
Map body	**colour** (4), **annotation style** (3)	**geographic extent** (4), **annotation placement** (14), **spatial distribution** (2)
Legend	**font** (1), **placement** (1), *colour*	**measurement scale** (1), *text content, mismatch with map body*
Other elements	**font** (3), **placement** (10), **stroke width** (1), *size, colour*	*measurement units*

4 Results

The paper collection process described in Section 3 resulted in 27 reproduction candidates. The maps in two papers were demonstrating user interfaces and did not visualise any analysis results related to the paper content. They were thus not further considered. Ten papers on the list of reproduction candidates produced their maps using proprietary GIS or statistical software that was not freely distributed for educational or research purposes. In addition, some of these applications were unavailable for our (open source) operating system, i.e. Linux Ubuntu. This prevented us from reproducing the maps depicted in these papers in the way they were originally generated. However, three of these papers shared the scripts that they developed for their studies. One paper elaborated the data analysis and the relevant map creation using a proprietary GIS in detail, and shared these step-by-step instructions. This enabled us to attempt reproducing the map using a free and open source GIS. Six of the remaining 15 papers did not disclose any code at all, which also prevented us from reproducing the maps they contained.

4.1 Reproduction Outcomes

In total, we initiated the reproduction process as described in Figure 2 for nine out of the 27 reproduction candidates. Four of these reproductions were stalled after the phase of getting the data and the software. The reproduction of three candidate papers could not proceed because no part of the code was associated with creating the maps shown in the paper. For one candidate paper, reproduction was not possible because the code was compressed as a .rar file that produced a "corrupt header" error when unpacking. For the remaining six reproduction candidates, we were able to eventually generate a visual output. For two of these papers, we assessed the reproduction attempts of all the included maps as failed. We therefore managed to successfully reproduce maps from four papers.

The time required for reproduction varied greatly. One paper pointed to an online interactive notebook written in JavaScript that visualised the map figures of the paper in a transparent way. In this case, the successful reproduction required only a few clicks. For another paper, setting up the computational environment, going through several processing stages and resolving the issues we stumbled upon in cooperation with the authors, led to a successful reproduction only after seven weeks. The effort required for the other papers (both with failed and successful reproduction) fell between those two extremes. In the following, we describe the obstacles that we faced during the reproduction process in more detail.

4.2 Encountered Obstacles

In addition to issues preventing us from reproducing maps at all as outlined in 4.1, we mainly encountered data-related issues and issues in configuring the computational environment.

4.2.1 Data-related Issues

As discussed in Section 3, we targeted only papers with available data. Despite this, we still occasionally ran into missing files while executing the workflow. Since we could overcome such obstacles only with the help of the authors, we approached them via e-mail asking for information about where to find the missing files.

As mentioned before, some papers pointed to organisation websites for downloading the data. Possible changes in the organisation's data cleaning and pre-processing practices since the elaboration of the study led to different datasets, and eventually failed reproduction. The same was true for data processing methods that contain stochastic operations, which were neither preserved nor documented. Since substantial differences in the underlying datasets entail different analysis results and consequently maps, we stopped our efforts in such cases and considered the map reproduction for the corresponding paper as failed.

In the case of one paper, the data was shared as a database dump. Importing it into a database turned out to be complicated, as there was no accompanying information regarding the database name and log-in credentials. We eventually discovered the necessary information in the code files. A further issue was the lack of documentation of the installed plugins in the database where the dump was extracted from. To solve this issue, we searched the Internet for the error/warning log we received when trying to work with the database dump to identify the missing plugins and to import the dump into a properly configured database.

4.2.2 Computational Environment Configurations

While data-related issues caused problems for our reproduction attempts, by far the most considerable impediment that we faced in almost every reproduction attempt was related to the configuration of the computational environment. This includes in particular the lack of documentation regarding a) the versions of the software packages that make up the computational environment, b) the intended usage of the code, and c) the connections between the scripts and the data. This was especially true when there were multiple data processing steps involved in the map making process and several data sources that needed to be accessed individually and then combined. In order to resolve these issues, we had to resort to thoroughly scrutinising the source code itself and to experimenting with various configurations in the hope to identify the correct one. Frequently, we also had to change the (hard coded) file paths in the code in order to match our file system structure.

Such issues could be addressed by containerisation, for example, but we did not come across any paper that preserved the computational environment – neither in the form of a container nor as a deployed application. One of the screened papers shared the package list for its Python environment as .yml file, which saved us much time. However, we had to manually change the version of one package, since there were conflicts that could not be automatically resolved by the package manager (conda). Two papers listed the names of the Python packages they considered most significant, but without mentioning the corresponding versions. This caused substantial extra effort during the reproduction process as some of the packages we had to install in our system conflicted with each other during the execution of the scripts, resulting in an abrupt termination of the workflow. After an extensive Internet research on various Python forums, we tracked down which packages caused the error and tried several combinations of different versions until we could execute the scripts successfully.

Although the aforementioned issues mainly refer to interpreted languages that rely on package availability, such as Python and R, similar problems were encountered in the case of applications that use compiled languages, such as Java. Changes in the version or linkage to the repositories of the application's dependencies (libraries, plugins) not only mandate re-configuration in part of the code, but also may result in different visualised output.

4.3 Communication with the Authors

Communicating with the authors proved to be a valuable resource in addressing many of the issues outlined in Subsection 4.2. We contacted the correspondence authors via the e-mail address provided in the paper, stating that we are running a reproduction study that is focused on maps and expressing our interest in their work. In the same e-mail we described the technical issues that we were facing and asked for their help. Most frequently we reached out for help in case the code required more data files than those pointed out by the availability section. When we came across broken links, we requested the correct URLs from the authors. Other reasons for contacting the authors were to obtain more detailed instructions for executing the code and to clarify which library versions had to be used. It is worth noting that the authors almost always replied to us within a few days and were willing to advise on how to proceed with the reproduction. In one occasion, this communication resulted in an update of the corresponding data repository of the paper. We are very grateful for the extensive and helpful support we received from the authors during this study.

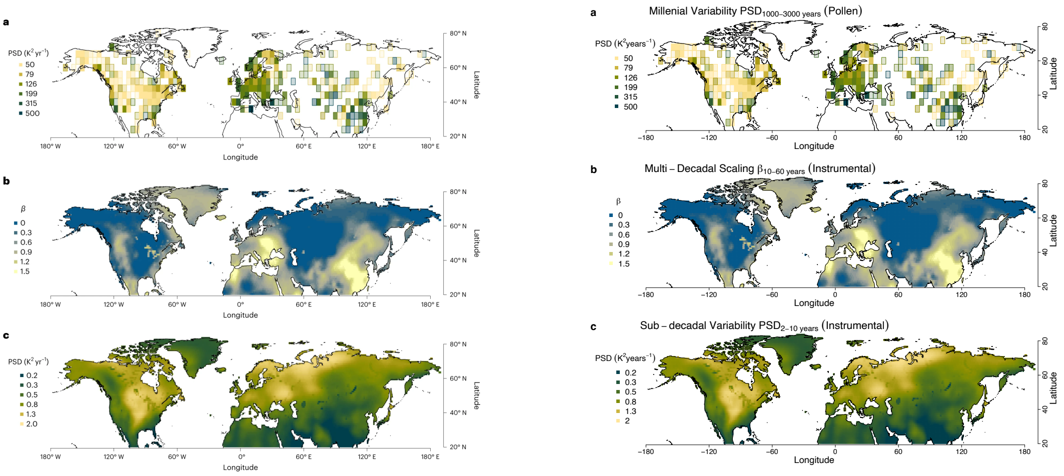
4.4 Reproduced maps

Successful Reproduction

The maps in Figure 3 were reproduced using R scripts [12]. We can observe that the reproduced map has titles for all the subfigures, while the original does not. The fonts of the axes labels and units and the legend are not the same. The units in the axes that show the longitude and latitude in degrees are represented in a different way. The orientation of the vertical axis labels is also different. Finally, the strokes appear bolder and the colors more saturated in the reproduced map. As we can see, the original and the reproduced maps are not identical in the pixel level, but they visualise the same data in a way that the intended message is still conveyed within the context of the paper. Therefore we considered this reproduction to be successful.

Unsuccessful Reproduction

In this example, the analysis of the paper was elaborated in ArcGIS Pro. Taking advantage of the instructions for reproducing the analysis that was provided by the authors [29], we tried to reproduce these steps in QGIS. Evidently, the reproduced maps differ greatly from the original ones. We attribute these deviations to different implementations in the underlying functionalities of the two GUI applications and to possible missing steps in the instruction sheet. Apart from the color scheme, the fonts, the basemap and the geographic extent, which we attempted to approximate with manual configurations in the GUI, we ended up with deep semantic differences that change drastically how the maps are read in the context of the paper. The reproduced maps differ in the shape of the tiled area, the measurement scale (as observed in the legend), and eventually, in the spatial distribution that is illustrated. As the maps diverge considerably in semantic level, we considered this reproduction attempt unsuccessful.



(a) Original. **(b)** Reproduced.

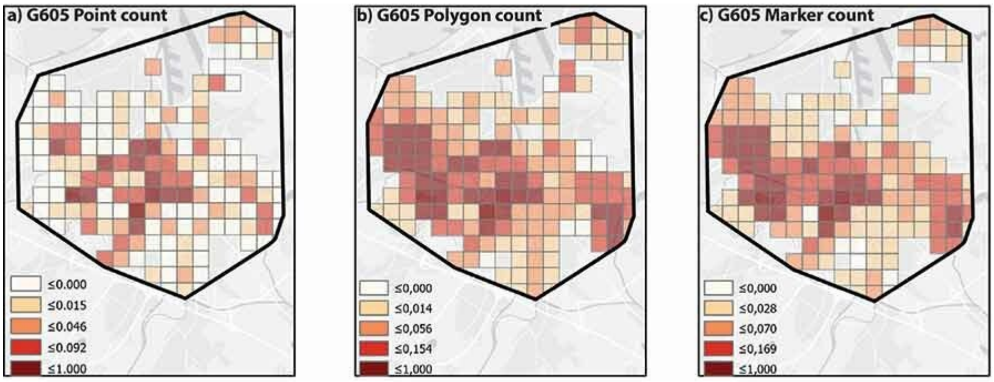
Figure 3 Example of successful reproduction with differences of aesthetic nature. Original (a) and reproduced (b) maps were created with R script. Original Figure (a) is extracted from Herbert et al. [11] as is, under CC BY 4.0 (`https://creativecommons.org/licenses/by/4.0/`).

5 Discussion

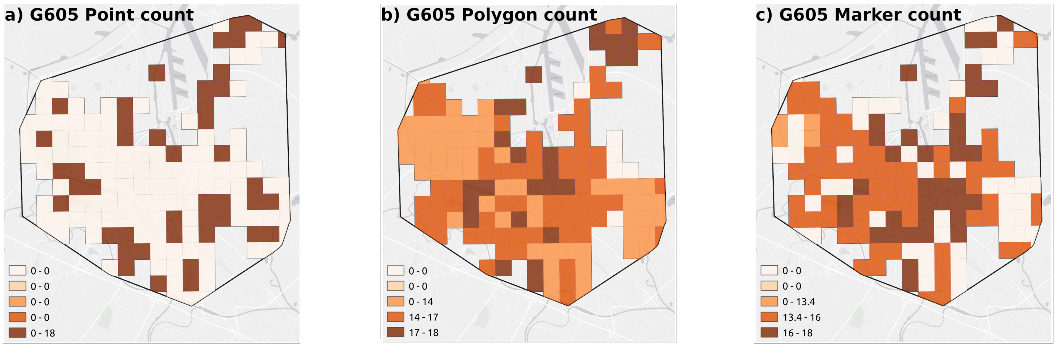
5.1 The State of Map Reproducibility

Based on our results and the insights gained through the reproduction study, it is evident that map reproducibility is in a dire state: out of 27 candidate papers from the last two years, we were able to only reproduce maps from four papers. However, it should be noted that reproducibility in general is not yet fully established in the Geosciences as a general requirement. In addition, half of the papers used commercial, proprietary software, which we excluded since it would require those who want to reproduce the contained maps to purchase the software. It is possible that the maps of those papers could have been reproduced, had we accepted to use the corresponding software. In one such case, we attempted to reproduce the maps using free and open software, which was not successful. Standardising map production functionalities and its description could be a way to bridge this gap, as it would enable different software to generate the same maps. Overall, we were positively surprised to be able to reproduce any maps at all, given the obstacles we faced.

Conceptualising and operationalising map reproduction success was also a challenge, as it became clear quite quickly that a pixel-based comparison is not rather useful. The initial set of criteria we defined enabled us to systematically assess reproduction success but constitute only a first step towards defining and formalising this concept and the underlying process. On a more practical level, poor documentation or a complete lack thereof required the most effort and frequently led to reproduction failure. While this is understandable given the current reward mechanisms in scientific publishing, it would also be easy to overcome and greatly reduce the effort involved in map reproduction. The data related issues we encountered also highlighted the importance of archival data repositories with persistent links for map reproduction.



(a) Original.



(b) Reproduced.

Figure 4 Example of unsuccessful reproduction with semantic differences. Original (a) was created with ArcGIS Pro and shows a subset of Figure 5 of Ramírez Aranda et al. [30], under CC BY 4.0 (`https://creativecommons.org/licenses/by/4.0/`). The reproduction (b) was created with QGIS.

5.2 Recommendations for Improving Map Reproducibility

From the experiences we gathered during our reproduction, we can infer a number of recommendations for improving map reproducibility. The most important is to *fully document all steps and intermediate results*, so that people unfamiliar with the research can execute them independently and assess whether the intermediate outcomes they produce correspond to those produced by the authors, in line with what was proposed by Silva et al. [32]. In the simplest scenario, this could include providing a flowchart that shows which data was used at what step using which analysis component. This way, it is much easier to identify what might have gone wrong when the final result differs from the original map. A second recommendation for improving map reproducibility is to *publish data, code and configurations with reproduction in mind*. Using persistent links and data repositories avoids issues related to dead links and changed datasets. When using public repositories, researchers should assess their suitability for reproduction and make use of features that facilitate reproduction (e.g., archiving repositories in Github so that the snapshot at map production time is preserved). Furthermore, it is helpful to provide relative file paths rather than absolute ones as well as to preserve the structure of the file system. Containerisation is one promising option that can help overcome many issues related to computational environment configurations, which has been successfully used for reproduction in previous work [24]. A third recommendation is to *keep the map production process as simple as possible*. When we tried to reproduce maps, we generally observed that the more steps were involved (from data pre-processing to the final map), the more difficult it became to successfully reproduce the final map. In some cases, complex production processes cannot be avoided but should then be very well documented. A fourth recommendation is to *educate researchers on how to conduct reproducible research*. This has been suggested before [16, 26] and is important not only for making them aware of the importance of reproducibility, but also to enable them to publish their research in a reproducible way, including the maps they develop. Our final recommendation is to *define reproducible map standards* so that assessing successful reproduction becomes easier and researchers can easily estimate whether their published work meets those standards. There are suggestions for reproducibility standards in general [22, 23, 27], but as outlined in Section 3.2.1, a more nuanced approach might be needed for maps as a simple pixel-based comparison does not capture well whether two maps convey the same message. The criteria we outlined in that section are a starting point towards defining such standards. Finally, a change of reward schemes in academic publishing (also strongly recommended by Ostermann et al. in [27]) could greatly benefit map reproducibility, as at the moment the extra effort required for making maps reproducible is not rewarded.

5.3 Limitations

Our study was exploratory in nature and thus is subject to a number of limitations. This includes the relatively small amount of papers (27), which was reduced further after deeper analysis. While this limits the generalisability of our findings, we were still able to identify many relevant obstacles and gain insights into how maps from academic papers can be reproduced. In addition, the reproduction study was carried out without a rigid, predefined protocol, which was due to the novel domain (map reproducibility). Hence, later reproduction attempts benefited from lessons learnt in earlier ones. Furthermore, the study was carried out by a single researcher and their abilities and knowledge affected the reproduction process, e.g., in terms of expertise in certain programming languages. It can be expected that reproducing researchers vary in their abilities and knowledge as well, so while having only one person do the reproductions limited generalisability, it also provided a realistic test for map reproducibility. Finally, we used an initial simple set of assessment criteria to determine the success of map reproduction. While clearly further research is needed here, we consider the proposed set a good starting point and the presented study can serve as an initial evaluation of those criteria as well.

5.4 Future Work

Our work brought to light several gray areas and loosely defined concepts regarding reproducible research, which also depend on individual interpretation. An important area for future work is thus to further investigate the diverse perspectives surrounding the notion of successful map reproduction. This applies in particular to the question when a map is considered reproducible, which factors affect this decision, and how significant various semantic and aesthetic differences are in this context. Another interesting question for further research is to what extent can the reproducing researcher modify the map making process to create an identical copy of the original map without compromising its integrity. Understanding and integrating the different viewpoints will contribute to a more comprehensive evaluation of map reproducibility. The clarification of these concepts will pave the way for more systematic approaches to map comparisons, which can be useful beyond the reproducibility of the map making processes as well. Examining these factors will contribute to advancing the understanding and implementation of reproducible research in the Geosciences.

6 Conclusion

In this study, we explored how reproducible maps are that are included in recent scientific publications. We collected a total of 27 papers, attempted to reproduce the maps contained therein, and managed to successfully reproduce them in four cases. We report on the process that we followed and the obstacles that we encountered. Our key contributions are an initial definition of *reproducible map making*, an inceptive set of criteria to assess the success of a map reproduction attempt and a set of guidelines for improving map reproducibility in geoscientific publications. Our work – while exploratory in nature – provides a first systematic analysis of map reproducibility. These outcomes are aligned with previous work regarding reproducibility in the domain of Geosciences but the particularities of maps as representational artifacts introduce further challenges and require further research for conceptualising and operationalising map reproduction. As a next step in this line of work, we are therefore planning to consult researchers in the Geosciences to develop a deeper understanding of what constitutes a successful map reproduction.

References

1 Lorena A. Barba. Terminologies for Reproducible Research, February 2018. `doi:10.48550/arXiv.1802.03311`.

2 Philip E. Bourne, Timothy W. Clark, Robert Dale, Anita de Waard, Ivan Herman, Eduard H. Hovy, and David Shotton. Improving The Future of Research Communications and e-Scholarship (Dagstuhl Perspectives Workshop 11331). *Dagstuhl Manifestos*, 1(1):41–60, 2012. Place: Dagstuhl, Germany Publisher: Schloss Dagstuhl — Leibniz-Zentrum fuer Informatik. `doi:10.4230/DagMan.1.1.41`.

3 Jon F. Claerbout and Martin Karrenbach. Electronic documents give reproducible research a new meaning. In *SEG Technical Program Expanded Abstracts 1992*, SEG Technical Program Expanded Abstracts, pages 601–604. Society of Exploration Geophysicists, January 1992. `doi:10.1190/1.1822162`.

4 David L. Donoho, Arian Maleki, Inam Ur Rahman, Morteza Shahram, and Victoria Stodden. Reproducible Research in Computational Harmonic Analysis. *Computing in Science & Engineering*, 11(1):8–18, January 2009. Conference Name: Computing in Science & Engineering. `doi:10.1109/MCSE.2009.15`.

5 Jean-Daniel Fekete and Juliana Freire. Exploring Reproducibility in Visualization. *IEEE Computer Graphics and Applications*, 40(5):108–119, September 2020. `doi:10.1109/MCG.2020.3006412`.

6 Timothée Giraud and Nicolas Lambert. Reproducible Cartography. In Michael P. Peterson, editor, *Advances in Cartography and GIScience*, Lecture Notes in Geoinformation and Cartography, pages 173–183, Cham, 2017. Springer International Publishing. `doi:10.1007/978-3-319-57336-6_13`.

7 Timothée Giraud and Nicolas Lambert. Reproducible Workflow for Cartography – Migrants Deaths in the Mediterranean. *Proceedings of the ICA*, 2:1–7, July 2019. `doi:10.5194/ica-proc-2-38-2019`.

8 Michael F. Goodchild, A. Stewart Fotheringham, Peter Kedron, and Wenwen Li. Introduction: Forum on Reproducibility and Replicability in Geography. *Annals of the American Association of Geographers*, 111(5):1271–1274, July 2021. `doi:10.1080/24694452.2020.1806030`.

9 Data Citation Synthesis Group. Joint Declaration of Data Citation Principles. Technical report, Force11, 2014. `doi:10.25490/A97F-EGYK`.

10 Michael A. Heroux, Lorena Barba, Manish Parashar, Victoria Stodden, and Michela Taufer. Toward a Compatible Reproducibility Taxonomy for Computational and Computing Sciences. Technical Report SAND2018-11186, Sandia National Lab. (SNL-NM), Albuquerque, NM (United States), October 2018. `doi:10.2172/1481626`.

11 R. Hébert, U. Herzschuh, and T. Laepple. Millennial-scale climate variability over land overprinted by ocean temperature fluctuations. *Nature Geoscience*, 15(11):899–905, November 2022. Number: 11 Publisher: Nature Publishing Group. doi:10.1038/s41561-022-01056-4.

12 Raphaël Hébert. Hhl2022, September 2022. doi:10.5281/zenodo.7062762.

13 Patrick J. Kennelly, Tom Patterson, Bernhard Jenny, Daniel P. Huffman, Brooke E. Marston, Sarah Bell, and Alexander M. Tait. Elevation models for reproducible evaluation of terrain representation. *Cartography and Geographic Information Science*, 48(1):63–77, January 2021. doi:10.1080/15230406.2020.1830856.

14 Christian Knoth and Daniel Nüst. Reproducibility and Practical Adoption of GEOBIA with Open-Source Software in Docker Containers. *Remote Sensing*, 9(3):290, March 2017. Number: 3 Publisher: Multidisciplinary Digital Publishing Institute. doi:10.3390/rs9030290.

15 Markus Konkol and Christian Kray. In-depth examination of spatiotemporal figures in open reproducible research. *Cartography and Geographic Information Science*, 46(5):412–427, September 2019. doi:10.1080/15230406.2018.1512421.

16 Markus Konkol, Christian Kray, and Max Pfeiffer. Computational reproducibility in geoscientific papers: Insights from a series of studies with geoscientists and a reproduction study. *International Journal of Geographical Information Science*, 33(2):408–429, February 2019. doi:10.1080/13658816.2018.1508687.

17 Menno-Jan Kraak. Strategic Plan for 2019–2027. Technical report, International Cartographic Association, Enschede, April 2019. URL: https://icaci.org/strategic-plan/.

18 Menno-Jan Kraak and Sara Irina Fabrikant. Of maps, cartography and the geography of the International Cartographic Association. *International Journal of Cartography*, 3(sup1):9–31, October 2017. doi:10.1080/23729333.2017.1288535.

19 Florian Ledermann and Georg Gartner. Towards Conducting Reproducible Distributed Experiments in the Geosciences. *AGILE: GIScience Series*, 2:1–7, June 2021. doi:10.5194/agile-giss-2-33-2021.

20 Alan M MacEachren. Cartography as an Academic Field: A Lost Opportunity or a New Beginning? *Cartographic Journal*, 50(2):166–170, May 2013. doi:10.1179/0008704113Z.00000000083.

21 Gengchen Mai, Weiming Huang, Ling Cai, Rui Zhu, and Ni Lao. Narrative Cartography with Knowledge Graphs. *Journal of Geovisualization and Spatial Analysis*, 6(1):4, February 2022. doi:10.1007/s41651-021-00097-4.

22 B. A. Nosek et al. Promoting an open research culture. *Science*, 348(6242):1422–1425, June 2015. Publisher: American Association for the Advancement of Science. doi:10.1126/science.aab2374.

23 Daniel Nüst, Carlos Granell, Barbara Hofer, Markus Konkol, Frank O. Ostermann, Rusne Sileryte, and Valentina Cerutti. Reproducible research and GIScience: an evaluation using AGILE conference papers. *PeerJ*, 6:e5072, July 2018. Publisher: PeerJ Inc. doi:10.7717/peerj.5072.

24 Daniel Nüst, Markus Konkol, Edzer Pebesma, Christian Kray, Marc Schutzeichel, Holger Przibytzin, and Jörg Lorenz. Opening the Publication Process with Executable Research Compendia. *D-Lib Magazine*, 23, January 2017. doi:10.1045/january2017-nuest.

25 Daniel Nüst and Edzer Pebesma. Practical Reproducibility in Geography and Geosciences. *Annals of the American Association of Geographers*, 111(5):1300–1310, July 2021. doi:10.1080/24694452.2020.1806028.

26 Frank O. Ostermann. Peer assessment to improve reproducibility of computational project work. In Hans-Ulrich Heiß, Hannu-Matti Järvinen, Annette Meyer, and Alexandra Schulz, editors, *SEFI 49th Annual Conference*, SEFI Proceedings, pages 1080–1090, Berlin, Germany, 2021. Technische Universität Berlin. URL: https://www.sefi.be/wp-content/uploads/2021/12/SEFI49th-Proceedings-final.pdf.

27 Frank O. Ostermann, Daniel Nüst, Carlos Granell, Barbara Hofer, and Markus Konkol. Reproducible Research and GIScience: An Evaluation Using GIScience Conference Papers. In Krzysztof Janowicz and Judith A. Verstegen, editors, *11th International Conference on Geographic Information Science (GIScience 2021) – Part II*, volume 208 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 2:1–2:16, Dagstuhl, Germany, 2021. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi:10.4230/LIPIcs.GIScience.2021.II.2.

28 Roger D. Peng. Reproducible Research in Computational Science. *Science*, 334(6060):1226–1227, December 2011. Publisher: American Association for the Advancement of Science. doi:10.1126/science.1213847.

29 Nohemí Ramirez Aranda, Jeroen De Waegemaeker, Viktor Venhorst, Wim Leendertse, Eva Kerselaers, and Nico Van de Weghe. Point, polygon, or marker? In search of the best geographic entity for mapping Cultural Ecosystem Services using the online PPGIS tool, "My Green Place.", December 2020. doi:10.5281/zenodo.4347404.

30 Nohemí Ramírez Aranda, Jeroen De Waegemaeker, Viktor Venhorst, Wim Leendertse, Eva Kerselaers, and Nico Van de Weghe. Point, polygon, or marker? In search of the best geographic entity for mapping cultural ecosystem services using the online public participation geographic information systems tool, "My Green Place". *Cartography and Geographic Information Science*, 48(6):491–511, November 2021. doi:10.1080/15230406.2021.1949392.

31 Center for Open Science. New Measure Rates Quality of Research Journals' Policies to Promote Transparency and Reproducibility. URL: https://www.cos.io/about/news/new-measure-rates-quality-research-journals-policies-promote-transparency-and-reproducibility.

32 Claudio T. Silva, Juliana Freire, and Steven P. Callahan. Provenance for Visualizations: Reproducibility and Beyond. *Computing in Science & Engineering*, 9(5):82–89, September 2007. doi:10.1109/MCSE.2007.106.

33 Arfon M. Smith, Daniel S. Katz, and Kyle E. Niemeyer. Software citation principles. *PeerJ Computer Science*, 2:e86, September 2016. Publisher: PeerJ Inc. doi:10.7717/peerj-cs.86.

34 Daniel Sui and Peter Kedron. Reproducibility and Replicability in the Context of the Contested Identities of Geography. *Annals of the American Association of Geographers*, 111(5):1275–1283, July 2021. doi:10.1080/24694452.2020.1806024.

35 D. R. FRASER Taylor. CHAPTER 1 – Geographic Information Systems: The Microcomputer and Modern Cartography. In Fraser Taylor, editor, *Modern Cartography Series*, volume 1 of *Geographic Information Systems*, pages 1–20. Academic Press, January 1991. doi:10.1016/B978-0-08-040277-2.50009-X.

36 Mark D. Wilkinson et al. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1):160018, March 2016. Number: 1 Publisher: Nature Publishing Group. doi:10.1038/sdata.2016.18.