A GLIMPSE INTO THE REPRODUCIBILITY OF SCIENTIFIC PAPERS IN MOVEMENT ECOLOGY:
HOW ARE WE DOING?

By

JENICCA POONGAVANAN

A THESIS PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER IN SCIENCE

UNIVERSITY OF FLORIDA

2021

To Amaye

ACKNOWLEDGMENTS

I would like to express my sincere gratitude and appreciation to my advisors, Dr. Mathieu Basille and Dr. Rocio Joo Arakawa for always being there to guide, advise and help me throughout this project. I would also like to thank Dr. Victoria Goodall and Dr. Scott Robinson for taking the time to be on my committee. The pandemic did not make it easy! I would also like to thank my parents for their love and support during my studies despite being on the other side of the world. My sincere regards to my friends, Juhi, Simona, Matt, Laura, Tavish and Deepesh for their help, love and fun times when everything was going wrong. Additional thanks to the Human Science Frontier Program for funding this project.

TABLE OF CONTENTS

page

CHAPTER

LIST OF TABLES

LIST OF FIGURES

Abstract of Thesis Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Master in Science

A GLIMPSE INTO THE REPRODUCIBILITY OF SCIENTIFIC PAPERS IN MOVEMENT ECOLOGY:

HOW ARE WE DOING?

By

JENICCA POONGAVANAN
May 2021

Chair: Mathieu Basille

Major: Interdisciplinary Ecology

Reproducibility is the earmark of science and thus Movement Ecology as well. However, studies in disciplines such as biology and geosciences have shown that published work is rarely reproducible. Ensuring reproducibility is not a mandatory part of the research process and thus there are no clear procedures in place to assess the reproducibility of scientific articles. In this study we put forward a reproducibility workflow scoring sheet based on six criteria that lead to successful reproducible papers. The reproducibility workflow can be used by authors to evaluate the reproducibility of their studies before publication and reviewers to evaluate the reproducibility of scientific papers. To assess the state of reproducibility in Movement Ecology, we attempted to reproduce the results from Movement Ecology papers that use behavioral pattern identification methods. We selected 75 papers published in several journals from 2010-2020. According to our proposed reproducibility workflow, sixteen studies reflected at least some reproducibility (scores ≥ 4). In particular, we were only able to obtain the data for 16 out of 75 papers. Out of these, a minority of papers also provided code with the data (6 out of the 16 studies). Out of the 6 studies that made both data and code available, only four studies reflected a high level of reproducibility (scores ≥ 9) owing it to good code annotation and execution. Based on our findings, we proposed guidelines for authors, journals and academic institutions to enhance the state of reproducibility in Movement Ecology.

CHAPTER 1
INTRODUCTION

Reproducibility is a fundamental ingredient in scientific work as it allows researchers to review and re-run studies reported by other scientists. Across different scientific fields the term 'reproducibility' is often used interchangeably with the term 'replicability', which leads to confusion (National Academies of Sciences, Engineering and Medicine, 2019). *Replicability* consists of a study arriving at the same scientific findings as another study by following the same experimental protocols and analytical methods but with new data (Barba, 2018), while reproducibility is obtaining the same results reported in a paper when using the same input data and computational steps to re-run the analysis (Patil et al., 2016). Kitzes et al. (2017) considers results to be computationally reproducible when an independent researcher is able to recreate key quantitative results using the same data and computational code (hereafter code). Generally, researchers can find it difficult to guarantee the replicability of their study but computation plays a big part in deriving results and can usually be the one thing that a researcher can guarantee about a study (Peng, 2015).

Researchers ought to publish reproducible work primarily to: 1) help strengthen scientific claims, 2) maintain public trust, and 3) empower the growth of future scientific research (Kelly, 2006; Way Community et al., 2019). Science relies on trust; researchers rely daily on the work presented by other experts with different areas of expertise. Laypeople have to trust in scientists' findings and counsel to deal with scientific information (Hendriks et al., 2016). Moreover, scientists building on reproducible research reduces the amount of time spent on collecting data themselves and establishing ways to analyze those data when that has already been collected and discovered (Gandrud, 2013). Researchers do not have to work from scratch, they can easily gain time and effort by building upon those established findings and therefore developing new ones.

Further benefit with reproducibility is the increased transparency across the scientific publishing community. A scientific community that works in a transparent environment thrive together. At the same time, when unsound papers are published, researchers will go forward building on faulty research, making use of the same approach or using the results to support their research. Unsound papers go unidentified due to our inability of reproducing those papers which leads to the entire community failing.

In 2016 a survey on reproducibility in science showed that 52% of researchers across numerous fields (biology, chemistry, medicine and others) acknowledged that there is an issue of reproducibility (Baker & Penny, 2016). Despite the advantages, generating reproducible research is still an uncommon practice (Reichman et al., 2011). A key barrier to reproducibility is the scarce availability of data (Lewis et al., 2018), particularly with older publications where data have not been archived online. A recent study in wildlife ecology (Archmiller et al., 2020) was moderately successful at reproducing studies for which the data were available (19 out of 74 studies). Even when data is available there is no guarantee that a study can be fully reproduced. For instance, a second issue is the availability of code underlying the research findings. A study by Culina et al. (2020) showed that code availability is alarmingly low in ecology; only 93 out of 346 assessed articles that were accompanied by code. Note that Culina et al. (2020) did not examine if the code were fully reproducible. Stodden et al. (2018) were able to reproduce the findings of 22 papers out of a sample of 89 published in the journal “Science” for which data and code were available. The state of reproducibility has also shown to differ between fields: in geoscience, reproducibility studies were able to reproduce 33 out of 41 studies (Gil et al., 2016); by contrast in clinical research, more than half of the 168 studies failed to be reproduced.

Other than technical there are also cultural barriers for papers not to be published in a reproducible way (Reichman et al., 2011). Firstly, the credit system does not reward authors enough for the time and effort it would take to fully disclose their work (Heesen, 2018). Secondly, some researchers might be concerned that other people will make use of their data and code to refute or compete with them or even publish before them (Barron, 2018). A contrasting point of view is expressed in Donoho (2002): “True. But competition means that strangers will read your papers, try to learn from them, cite them, and try to do even better. If you prefer obscurity, why are you publishing?”. Published studies can motivate future research, inspire new products and inform government policies (Lewis et al., 2018). So people need to have confidence in published results. If their conclusions are misleading or simple incorrect, we risk time, resources and even our health in the pursuit of false leads.

In this paper, we focus on reproducibility in Movement Ecology. Over recent decades technological advances brought about massive volumes of data to support scientific research (Hampton et al., 2013). In animal movement studies, we moved from Very-High-Frequency (VHF) telemetry tags to Geographical Positioning System (GPS) tags which enabled the collection of accurate and extensive relocation data (Cagnacci et al., 2010; Rocío Joo et al., 2020; Kays et al., 2015). This deluge of data in combination with widely accessible and affordable computing resources enabled scientists to address questions at larger and finer spatial and temporal scales. The needs to address those questions led to an ever increasing body of literature pertaining to the statistical modeling of animal movement which resulted in diverse ecological insights into animal behaviors (Morales et al., 2004; Patterson et al., 2017). Computational reproducibility is crucial to attain a standard of credible research results.

However compared to other fields, the sub-field of Movement Ecology (specifically behavioral pattern identification studies) has given little attention to reproducibility.

This study aims to address this gap by implementing a reproducibility study using previously published articles in Movement Ecology. We focused on a specific aspect of Movement Ecology, behavioral pattern identification, because 1) behavioral pattern identification methods nowadays are very complex and highly computational and 2) I have an understanding of how behavioral identification methods operate. We also focused on articles that used R (R Core Team, 2020) as it is one of the most used open source software in Movement Ecology (Rocío Joo et al., 2020).

For a study to be reproducible, there are key criteria that need to be fulfilled (National Academies of Sciences, Engineering and Medicine, 2019; Piccolo & Frampton, 2016; Powers & Hampton, 2019; Sandve et al., 2013): We identified the following six criteria: data availability, code availability, the use of open source software, successful execution of the code, code annotation and reproducibility of numerical results. Based on those criteria, we put forward a conceptual workflow to successful reproducible research. The workflow is not only applicable to Movement Ecology but to any code-based analyses and is not only to authors but to reviewers, who can easily assess the reproducibility of other studies.

In contrast to other studies (Archmiller et al., 2020; Konkol et al., 2019; Lewis et al., 2018) that only assessed reproducibility of the reported results, in this study we assign scores to papers based on each aforementioned criteria that encompass most sections of the manuscript. We fully evaluate the reproducibility of 75 studies published between 2010 and 2020. Our goals are to 1) establish a workflow that can be used as a tool by authors and reviewers to evaluate reproducibility in science and 2) evaluate the reproducibility status in the sub-field of Movement Ecology.

CHAPTER 2
METHODOLOGY

Data Collection

To evaluate reproducibility in Movement Ecology, we first screened for some of the most popular methods used in Movement Ecology to identify animal behaviors (Table 1). The list is not considered exhaustive but portrays methods used in behavioral pattern identification reviewed by Gurarie et al. (2016) and Bennison et al. (2018). To assess popularity of each method, we first identified the article that introduced the method in ecology. The sum of the number of citations of this paper 'Google Scholar' and 'Web of Science' (collected by 2020-11-13; Table 1) was used as a proxy of popularity of the method. Based on out popularity proxy, we shortlisted four methods as being the most popular: Hidden Markov models (HMM), Behavioral Change Point Analysis (BCPA), Expectation – Maximisation binary Clustering (EMbC) and First-Passage Time (FPT). We do not review those four methods but used them as a filter to obtain a random sample of relevant papers in Movement Ecology.

Table 2-1: List of behavioral pattern identification methods together with the number of times they have been cited on 'Google Scholar' and 'Web of Science' on 13th November 2020. The four selected methods for this review are highlighted in italics.

Method	Paper introducing the method for behavioral identification in ecology	Year	Number of times cited in		Sum of the number of citations
			Google scholar	Web of Science	
First-Passage Time (FPT)	Using first-passage time in the analysis of area-restricted search and habitat selection (Fauchald & Tveraa, 2003)	2003	558	341	899

Table 2-1. Continued

Method	Paper introducing the method for behavioral identification in ecology	Year	Number of times cited in		Sum of the number of citations
			Google scholar	Web of Science	
Behavioral Change Point Analysis (BCPA)	A novel method for identifying behavioral changes in animal movement data (Gurarie et al., 2009)	2009	335	197	532
Hidden Markov Model (HMM)	Flexible and practical modeling of animal telemetry data: hidden Markov models and extensions (Langrock et al., 2012)	2012	267	158	425
Expected Maximization binary Clustering (EMbC)	Expectation-Maximization binary Clustering for behavioral annotation (Garriga et al., 2016)	2016	81	51	132
K-means clustering	Quantitative classification and natural clustering of Caenorhabditis elegans behavioral phenotypes(Geng et al., 2003)	2003	75	48	123
Bayesian Partitioning of Markov Models (BPMM)	1. Computing the likelihood of sequence segmentation under Markov modelling (Guéguen, 2009)	2009	5	-	34
	2. Segmentation by maximal predictive partitioning according to composition biases (Guéguen, 2001)	2001	29	-	
Kernel Density	Time-in-area represents foraging activity in a wide-ranging pelagic forager (Warwick-Evans et al., 2015)	2015	27	16	43

Once we determined the most cited methods, we used ‘Google Scholar’ and ‘Web of Science’ as search engines to search for eligible papers to be reviewed for the reproducibility analysis. To be eligible, these papers had to 1) use one of the four selected methods to identify animal behaviors, 2) use R to execute the analysis (the most popular open source software in ecology and Movement Ecology (Rocío Joo et al., 2020; Lai et al., 2019), and 3) have a publication date from within the last 10 years (2010-2020). The date criterion was set as a primary filter on both search engines. We considered that more recent papers would have a higher chance to be reproducible because: R and other open software that provides a reproducible environment have become very popular in the ecological community, data sharing platforms (Michener, 2015) and journal demands for sharing data are also recent (Stodden et al., 2013) and half of the selected methods have only been used for animal behavior studies for less than a decade. We ultimately selected 75 papers; they were published in 46 different journals.

A Reproducibility Workflow

Ensuring reproducibility is not a mandatory part of the research process and thus there are no clear procedures in place to assess the reproducibility of scientific articles. Based on guidelines outlined in different scientific publications on reproducibility across several fields (National Academies of Sciences, Engineering and Medecin 2019; R. D. Peng, 2011; Piccolo & Frampton, 2016; Sandve et al., 2013), we were able to establish six criteria that lead to a successful reproducible paper. Those six criteria pertain to data availability, code availability, the use of open source software, correct execution of the code, code annotations and reproducibility of results (Fig 1). Based on these criteria we assigned scores to selected papers in order to measure how reproducible each paper was. Each criterion was rated on a scale of 2 points, for a grand total of 12 points, with 12 indicating an impeccable reproducibility standard and 0 indicating complete lack of reproducibility. We detail the scoring procedure for each criterion below. Although our focus was on the field of Movement Ecology and particularly animal behavior identification, it is important to note that our criteria are of general purpose. Any scientist seeking to make or evaluate reproducibility in research can easily adopt our workflow on other scientific works.

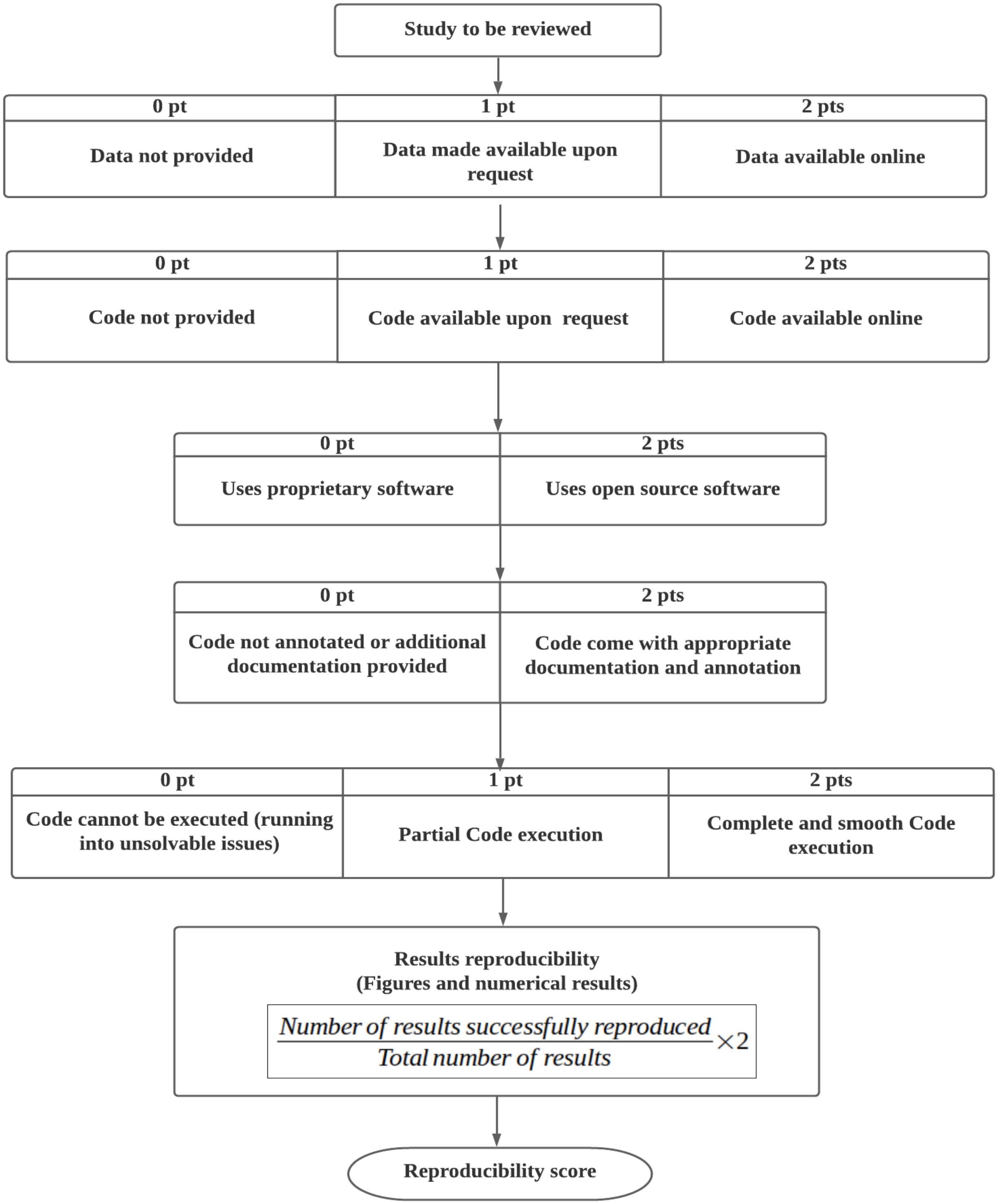

Figure 2-1: A conceptual framework for reproducible research used for scoring scientific papers. The score for reproducibility of results was calculated based on the proportion of results reproduced. For example, if 2 out of 4 tables, figures or paragraphs (i.e. numerical results within paragraphs) presented were successfully reproduced, the score would be : (2/4)×2=1 pt.

Availability of Data and Code

Any analytical process start with data (Figure 2-1). Adopting a reproducible workflow starts with making data and computational code available to the audience to demonstrate the decisions made to generate results. Compared to Archmiller et al. (2020), where they assigned scores only to the reproduced result, in this study we allocated points to papers for data and code availability. Data sharing is the very first step towards reproducibility and should be rewarded.

Therefore, once we had our selection of papers, we investigated whether each paper had their data and code shared online. If so, we downloaded the data and code together with any ancillary information available. In the event where data and code were not readily accessible, we applied the same procedure applied in earlier reproducibility studies (Archmiller et al., 2020): we emailed requests to the corresponding authors, with up to two reminders after two weeks respectively in the absence of response. We were transparent regarding our request for the data and code: our email explained the aim of our study and guaranteed that their identity, paper and data would be kept confidential. If data were available online, a score of 2 was allocated to the paper, if authors had to be contacted to procure the data, a score of 1 was allocated. A score of 0 was allocated if no data were provided or we did not receive any response to our requests after our established deadline. The same scoring approach was applied to the availability of code. Data availability in itself does not ensure the successful reproducibility of a paper (Culina et al., 2020). Similar to data availability, we considered code to be accessible online if they were archived in such a way that they could be readily accessed by anyone. We thus also allocated scores if code were provided along with the data online (score = 2) or upon request (score = 1).

Use of Open Source Software

We also included the use of open source software as a criterion given that for a study to be reproducible we need to be able to freely access any software used during an analysis. We only considered studies that used R and therefore, every study were attributed a score of 2 for making use of open source software. Papers are allocated points for using open source software irrespective or whether they provided data and code.

Code Annotation

Computer code provided alongside research paper have proved to be crucial in reproducing results (Culina et al., 2020). However without proper comments, it may be hard to interpret how the code accomplishes specific tasks and better understand the analysis (Obels et al., 2020). Proper annotations also allows researchers to re-use and adjust the code according to their needs. An approach to address this matter is through detailed code annotations interposed across the computational code, also known as literate programming (Knuth, 1984; Sandve et al., 2013). Thus, 2 points were earned if the code were properly annotated or proper additional documentations were provided for a smooth understanding of the computational code. A score of 0 was allocated if the script consisted only of computer code, with no comments to help understand what each specific task accomplishes.

Code Execution

We next looked at how smoothly we could run the code provided and obtain results. A score of 2 was allocated to the paper if the code were complete and that we were able to run the code as is from start to end without any alteration. A score of 1 was allocated if the code were incomplete and additional code had to be written to obtain some results. For example, if some additional data cleaning were required or if some coding lines required alterations after running into errors. A score of 0 was attributed in cases where we ran into errors that could not be fixed and the authors were unresponsive to our call for help. For example, outdated packages that would not run anymore and no proper guidance were given as a work around, or runtime errors that we attempted to fix to the best of our knowledge but did not lead to similar results.

Results Reproducibility

Like in Archmiller et al. 2020, we evaluated reproducibility in 2 ways. Firstly, whether numerical results cited in the text and tables matched the values stemming from our reproduction attempts (we allowed for differences within the publication's significant digit). Secondly, whether our reproduced figures matched the original figures presented in the paper while allowing for differences in the formatting of figures as well.

The scores to results reproducibility were attributed based on the proportion of results reproduced in the form of numbers and figures. The score was calculated by:

$$\frac{Numberofresultssuccessfullyreproduced}{Totalnumberofresultspresented\text{in}theformofparagraphs, figures\text{and}tables} \times 2 \tag{2-1}$$

where the maximum points a paper could score on reproducibility was 2.

Global Score

The global score is simply determined by adding all points from the above criteria. We decided to keep equal weight for all of them (i.e. points for each criterion) to make for a simple scale not focusing on any particular step of reproducibility. We considered studies to be highly reproducible if they obtained a score greater or equal to 9. Studies with a score between 4 and 9 were considered to reflect at least some reproducibility. Studies with a score lower than 4 were considered not reproducible.

CHAPTER 3
RESULTS

We selected 75 eligible publications. Only three of these articles contained sufficient information for us to locate both data and code online without the need to contact the authors (2 points for data and another 2 points for code). An additional 11 studies had only their data available online. We thus subsequently emailed 72 authors (Figure 3-1) requesting for data and/or code used in their studies. Three authors responded with the requested material (1 point for each of data and code provided). All others got 0 point on data and code. One author responded asking us to contact someone else who worked on the paper for the data and code, however we had no further response from the other author. Authors of 6 studies opted out for two different reasons: the data were to remain confidential (*n*=2) and they did not have enough time to compile the data and code (*n*=4). Furthermore authors of 7 papers consented to sending their data and code but failed to do so before our established deadline of February 16th. We also had no responses from 35 authors and 15 of our emails returned with undeliverable notes. In some cases we were unable to locate data and code; some papers indicated where the data were situated but we could not find them in those locations (*n*=3). Others papers made use of several datasets and pointed out several sources where they could be found, however it was difficult to identify the correct dataset used and the authors did not respond to any of our requests (*n*=2). Ultimately we were able to obtain data for 16 out of the 75 publications, from which only 6 provided code.

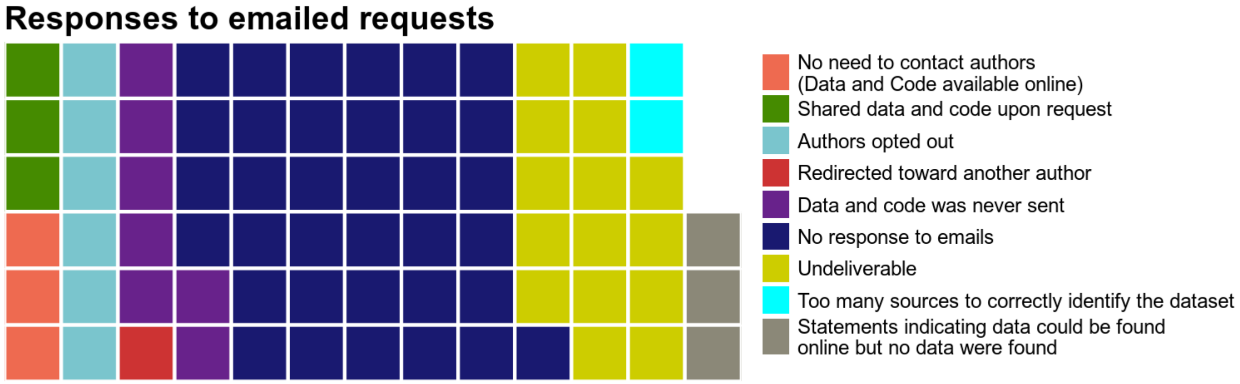

Figure 3-1: Responses to our emailed request from the 72 authors we contacted. 3 studies had data and code readily available online and thus we did email requests to the authors. Each square represents a response we received from the authors and responses are grouped by color.

No reviewed paper obtained a perfect score of 12 based on our proposed scale. Four papers had reproducibility scores greater or equal to 9, reflecting a high level of reproducibility (Figure 3-2). Twelve had reproducibility score between 4 and 9 that reflected at least some reproducibility (Figure 3-2). Fifty-nine of the 75 studies were considered not reproducible with reproducibility scores lower than 4 (Figure 3-2). As a matter of fact, those studies obtained only 2 points for making use of open source software, namely R, which was a selection criterion to start with.

Of the 12 studies that were least reproducible (i.e., reproducibility scores between 4 and 9), we determined that those studies did not provide enough material regarding the analysis to reproduce the results presented in the respective papers. Ten of the 12 studies did not provide any computational code, and only scored 4 points for data and use of open source software (Figure 3-3). Of the remaining two studies in this category (Figure 3-3: BCPA 01 and HMM 01), one study lost points for making the data and code available upon request only, and not online. Moreover, the code were incomplete and poorly annotated and we were therefore unable to reproduce the results presented in the paper to its full extent. The second study lost points due to the fact that despite making their code available, we were unable to execute the code and obtain similar numerical results. We suspect that the analysis code made use of an outdated version of an R package, but package versions were not documented so we can only speculate.

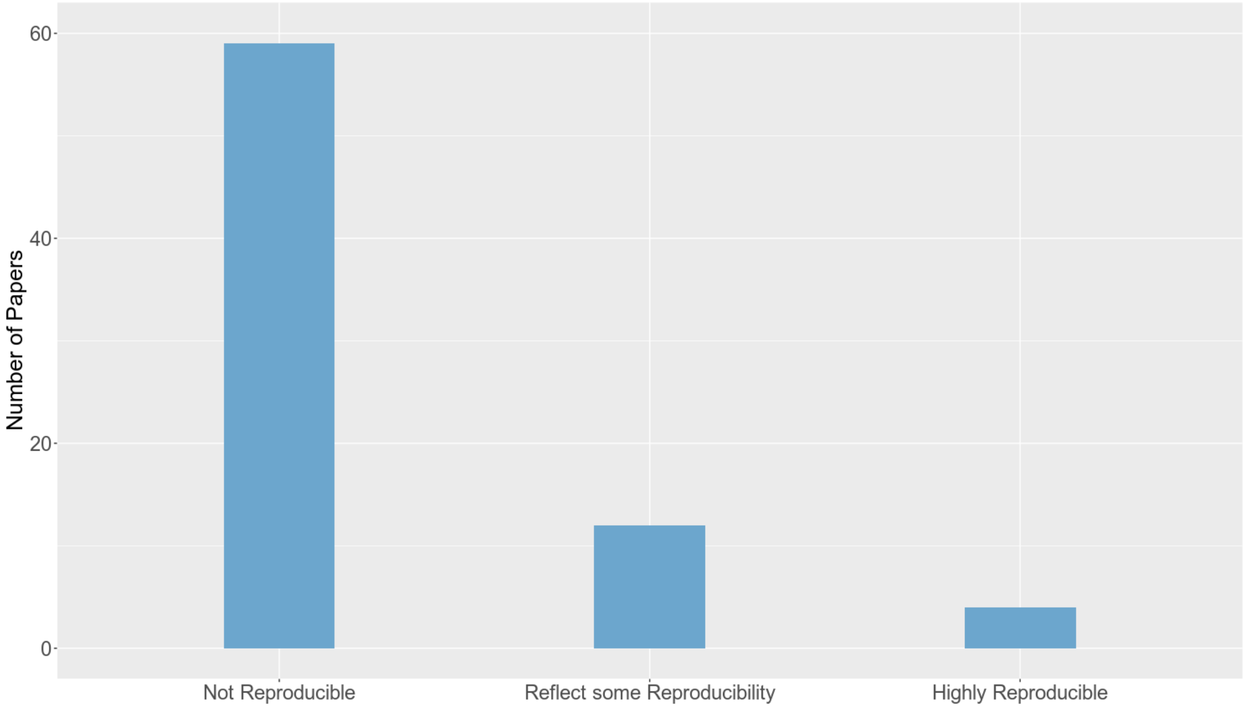

Figure 3-2: Classification of the 75 selected papers according their reproducibility scores. Studies with a score ≥ 9 were considered as highly reproducible. Studies with a score ≥ 4 and < 9 were considered to reflect at least some reproducibility. Studies with a score < 4 were considered not reproducible.

Two studies were almost completely reproducible (Figure 3-3: HMM 02, HMM 04). Those studies made their data and code available online (Figure 3-3: HMM 02, HMM 04). The code were well annotated and we were able to reproduce the majority of the results. However we suspect that some lines of code were missing given that we failed to reproduce some of the figures presented in the paper. A third paper (Figure 3-3: BCPA03) was classified as highly reproducible with a score of 10. We were able to fully reproduce the numerical results and the code were well annotated (Figure 3-3: BCPA 03). However the paper did not make the data and code readily available online. A fourth study with a score of 9.0 was also classified as highly reproducible (Figure 3-3: EMH 01). We were able to fully reproduce the numerical results, however the code were poorly annotated and we had to discern the corresponding computational lines with some help from the authors (Figure 3-3: EMH 01).

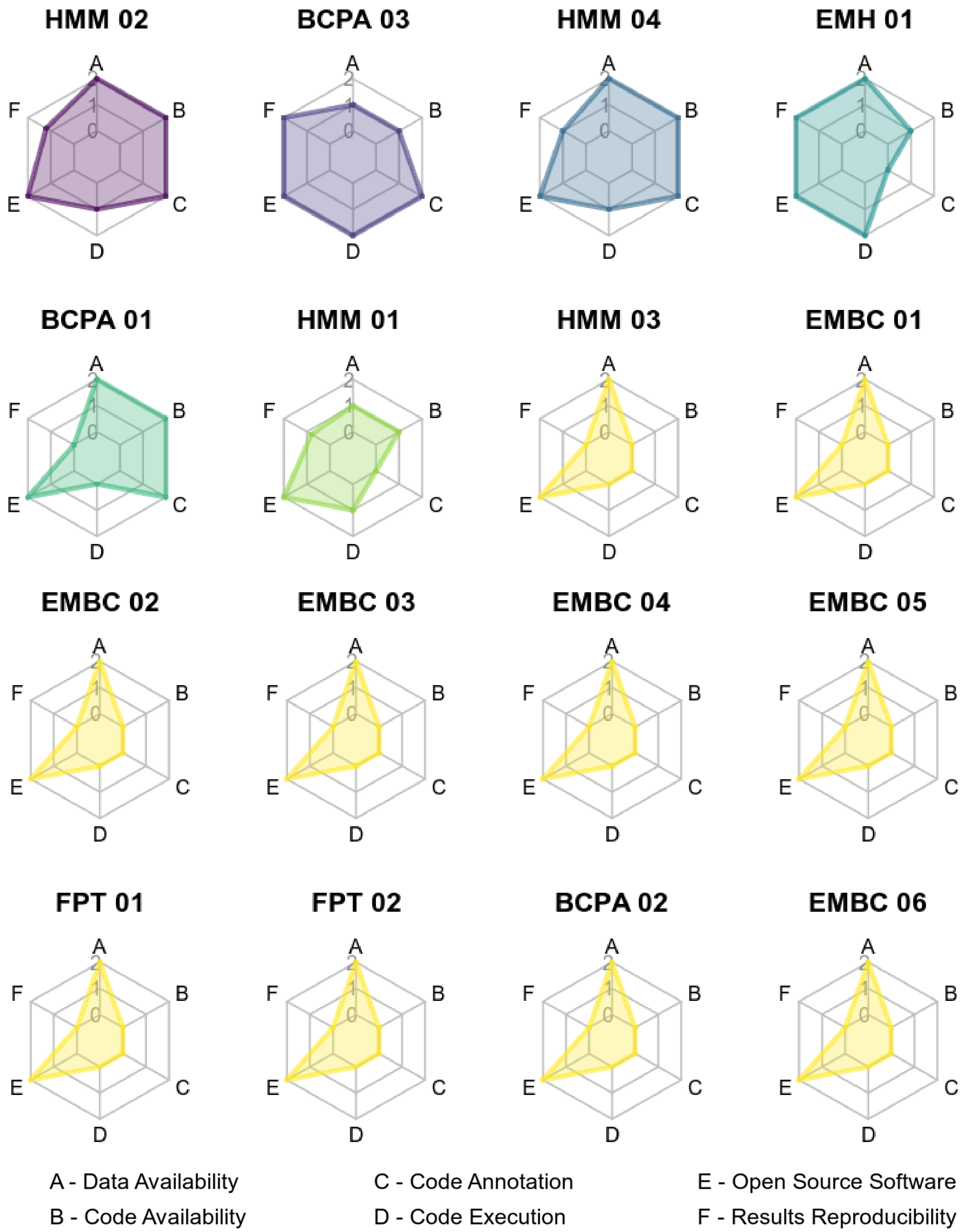

Figure 3-3: Total and detailed scores of the 16 papers with available data for each reproducibility criteria. Each criterion was scored based on a discrete scale {0,1,2}. For the exception of Numerical Reproducibility which was scored based on a continuous scale ranging between 1 and 2. The plots are color-coded based on each respective paper total reproducibility scores.

CHAPTER 4
DISCUSSIONS

The main goal of this study was to provide insight on the extent to which reproducibility is practiced in Movement Ecology. In order to do so, we put forward a workflow based on six criteria that constitute the basis for reproducible research. On a scale from 0 (irreproducible) to 12 (full reproducibility), 59 papers (79%) obtained a score of 2 only, and 16 papers, for which we were able to gather data, obtained scores ranging from 4.0 to 10.2. The workflow we presented can be easily used by scientists as guidance to quantify and evaluate how reproducible their papers are before publishing. It can easily be integrated with ecologists' workflows, provide support for open reproducible research and boost reproducibility in Movement Ecology. The message from this study is clear: the reproducibility of studies in the sub-field of Movement Ecology is low. Despite that we did not assess the reproducibility of studies across the every sub-fields of Movement Ecology, this study provided us with some insight on what might be transpiring across the entire field.

A pattern that characterized the high scoring papers was authors making both data and code readily available online. This seems obvious, but to be re-used, data need to first be found and accessed. Reproducibility starts with open data, one of the basic notions that embodies open science. The latter is a movement that aims at increasing the transparency and accessibility of scientific research through a set of practices (van der Zee & Reich, 2018). Open science encompasses three core concepts: open access, i.e. providing immediate unrestricted access to research articles for re-use (Björk et al., 2014; Swan et al., 2015); open data, whereby researchers make their data freely available to the scientific community (Michener, 2015); and open source, where computational code and software are made freely available to everyone who want to use, change and enhance it (Stodden et al., 2013).

Ideally, authors should provide their data following the FAIR principles (Wilkinson, 2016). Firstly, data should be *findable*, i.e. by using a globally unique Digital Object Identifiers (DOIs). DOIs are permanent identifiers, in the form of a unique string of numbers, letters and symbols, associated to an address on the web (an URL), which helps eliminate ambiguity among databases. Secondly data should be made *accessible*, as in open, free and available to the world. There are numerous web-based data archiving repositories available to ecologists where movement data can be archived at no or low cost. Thirdly, data should be *interoperable;* data should be available in such a format that it can be integrated with other data. Data and metadata should be presented in standardized formats, so that it can be processed by computers and used by people (Way Community et al., 2019). For instance, for R users this will involve providing data in commonly used formats such as 'Comma seperated values (CSV)' files. Information in CSV is an user-friendly data format and probably the most widely supported across several technological platforms (Mitlohner et al., 2016). Finally, data should be *reusable*, i.e. data can be repurposed for new research. Authors must specify whether the data produced in the project is usable by third parties. In cases where the re-use of some data is restricted, it should be clearly stated and justified.

In this study, 13 out of 75 studies reinforced the FAIR principles by using DOIs and web-based repositories. For example, Movebank is a specialized repository at the disposition of movement ecologists that accepts only animal movement data (https://www.movebank.org/; Kranstauber et al., 2011). Other general-purpose repositories include Dryad (https://datadryad.org/; Greenberg et al., 2009), Zenodo (https://zenodo.org/) and Figshare (https://figshare.com/). Having data in online repositories also eradicate the problem of undeliverable emails and would not require authors to constantly updating their contact information. Archiving repositories are valuable resources as they are typically free to access, assign DOIs, provide licenses (including both proprietary and open source licenses), are long-term and citable (Mislan et al., 2016). The general-purpose repositories also generally accepts data in most common formats. Data should be free of charge and under an open license so it can be reused by other researchers (Way Community et al., 2019).

The provision of data alone does not guarantee the reproducibility of numerical results (Culina et al., 2020). Out of the 16 studies we were able to track down the data, only 3 studies had made their code available in repositories or within the supplementary material of the paper. Authors from 3 additional studies responded to our request for the code. Similarly to data, code can also be stored on repositories and assigned DOIs to provide easy access to the readers. Platforms for code archiving such as Dryad (https://datadryad.org/) and Git Hub (https://github.com/; GitHub, 2016) where scientists can upload their code at a no or low cost, have seen their popularity growing in recent years (Culina et al., 2020). Mislan et al. (2016) further presents a comprehensive list of repositories for archiving code.

Another criteria that was significant in determining high scoring papers was code annotations and documentation. Code annotations are important pieces of text placed within the code to explain it, and help other researchers understand the code. Literate programming helps scientists understand the logic behind the code and allows scientists to adjust the analysis accordingly. With the deluge of data in Movement Ecology, came along complex statistical and computational models to investigate this data (Hampton et al., 2013). To concisely write about every step performed during an analysis in the method section of a paper can be hard. However, clearly stating analytical methods through annotations or additional code documentation helps the reader better understand the data and analysis. The lack of incentive for movement ecologists to make their code available also stems from the fact that appropriately documenting code can be time consuming and authors often do not receive enough recognition for the effort they put into it.

Nowadays, the existence of literate programming tools allow users to combine analyses and presentation of results into one document. For instance, *Rmarkdown* is a literate programming tool that keeps code and words together, and can be used to produce presentation documents from one script (see Gandrud, 2013 for guidelines to using Rmarkdown). The communication aspect of an analysis should not be recorded after the fact but as you are going through it. Annotating code is a gift you do not only give to others but to yourself as well. Coming back to code after a long time and recall the decisions made at the time can be confusing. Annotations can save your future self time.

We also came across a software related issue. For instance, in one particular case, the package available for download was newer version of the indicated package in the analysis, and we obtained numerically different results. R is one of the most used software in Movement Ecology (Rocío Joo et al., 2020; Lai et al., 2019) and has a vibrant package ecosystem (Rocio Joo et al., 2020). R packages are constantly changing and becoming deprecated (Ellison, 2010). To ensure that a project can be recomputed again at another time or by someone else, the version of the software and packages used need to be appropriately documented.

Every computer has its own unique computational environment, from the operating system being used to the versions of software packages installed (Piccolo & Frampton, 2016). Executing an analysis on a future date or on different computer can cause code not to run at all, let alone reproduce comparable results (Way Community et al., 2019). There are several approaches that are available to capture the computational environment in which an analysis was conducted such as *Renv, Docker* and *Binder*.

Renv is an R package that carefully records versions of the packages and their dependencies you have installed on your computer. By creating per-project libraries, Renv ensures that updating a package version later on for a different project will not affect the version of the package used for your initial project. Secondly, Docker (https://www.docker.com/) is an open source tool that works with containers. A container provides a virtual environment that wraps up code and all its dependencies so that the analysis can be executed on an different computing environment (Boettiger & Eddelbuettel, 2017). Containers make full reproducibility actually feasible. For example, in the event that an analysis was carried out using an older version of a package, a container will allow you to run the analysis using the older version of the package while still keeping the up-to-date version of the package on your computer (see Peikert & Brandmaier, 2019; Way Community et al., 2019 for a detailed description on Docker). Thirdly, Binder (https://mybinder.org/) is a web-based service that enables users to upload and share fully-functioning versions of projects online which can be accessed and interacted with by others via a web browser.

On the other hand, many researchers in ecology who resort to computational methods are self taught and are often unfamiliar with the best practices that support reproducibility (Poisot, 2015). In the field of ecology, there are initiatives such as the Carpentries (https://carpentries.org/) that provide ecologists with fundamental coding and data science skills needed to conduct reproducible research. While technical issues need to be addressed, proving training and supportive tools to ecologists are not sufficient to eliminate the practice of irreproducible research. As a matter of fact, the necessary tools are now all freely available as open source software, and there is ample documentation on the web to use them. Instead, we argue that the main challenge is more cultural than technical.

Publishing their research findings through scientific journals is one of the ultimate objective of researchers. Thus, journals can have a substantial influence on increasing reproducibility within the ecology community (Stodden et al., 2013). Across fields, an increasing number of funding agencies and journals now require researchers to make their data and code publicly available (Mislan et al., 2016; Stodden et al., 2013). Mislan et al. (2016) evaluated data and code sharing policies amongst 96 ecological journals and found that only 14 journals encouraged code sharing. In 2020, Culina et al. (2020) found that the number increased to 72 journals. However, it is difficult to determine whether a journal mandates or merely encourages data and code sharing (Culina et al., 2020). Scientific societies in particular have a role to play: (Stodden et al., 2013)

For instance, as of February 1st 2021, the Ecological Society of America (ESA) now requires that whosoever submits a paper to an ESA journal needs to disclose all underlying data and statistical code relevant to research findings. Raw data, metadata, code and additional documentations are to be submitted with the initial manuscript for peer review and editorial approval. This type of policy advocates for open research and at the same time strengthens scientific claims through indisputable evidence.

Journals might not be liable for the science being published but they are responsible for reporting it. Peer-review processes act as safety checks where experts examine submitted papers for potential shortcomings (Smith, 2006). The field of Movement Ecology is becoming highly computational (Hampton et al., 2013) and providing data and code will help reviewers follow the decision making process and identify erroneous discoveries. For instance, a rigid review of data would have perhaps prevented an incident that happened in Canada. In early 2020, an article (Pennisi, 2020) that shocked the science community emerged; a well-known behavioral ecologist was accused of data manipulation. Following the incident, several scientific papers that used his data were retracted from prominent scientific journals. Mandating policies such as making data, code and documentations available during the peer review process can deter such ethical misconduct.

Over the course of the study, I have learned that science should not be a race. Scientists are so much focused on publishing, that quantity has become more important than quality. Science is like a building, what is the use of a beautiful building if it is built on weak foundations? At the same time, the irony behind this reproducibility study is that in itself it is not fully reproducible. Data from the original studies that we collected during the review process are to be kept confidential and cannot be shared, as the goal of this work is not to denounce particular studies and scientists, but rather to criticize and find commonalities in an entire field. Instead of the raw data, we can only provide a fully reproducible account of the statistics and figures presented in this study, in a GitHub repository (**work in progress** at https://github.com/Jenicca).

In order to bring a more accountable and productive scientific culture, academic institutions, journals, funding organizations and policymakers can all play a role in improving open science and reproducibility in research results. The goal with reproducibility is not to point out errors or point fingers at somebody’s hard work but to make sure as a scientific community we all grow together and stronger so as to leave a bullet proof legacy behind.

LIST OF REFERENCES

Archmiller, A. A., Johnson, A. D., Nolan, J., Edwards, M., Elliott, L. H., Ferguson, J. M., Iannarilli, F., Vélez, J., Vitense, K., Johnson, D. H., & Fieberg, J. (2020). Computational Reproducibility in The Wildlife Society's Flagship Journals. The Journal of Wildlife Management, 84(5), 1012–1017. https://doi.org/10.1002/jwmg.21855

Barba, L. A. (2018). Terminologies for Reproducible Research.

Barron, D. (2018). How freely should scientists share their data. Scientific American Blog Network. Https://Blogs. Scientificamerican. Com/Observations/How-Freely-Should-Scientists-Share-Their-Data.

Bennison, A., Bearhop, S., Bodey, T. W., Votier, S. C., Grecian, W. J., Wakefield, E. D., Hamer, K. C., & Jessopp, M. (2018). Search and foraging behaviors from movement data: A comparison of methods. Ecology and Evolution, 8(1), 13–24. https://doi.org/10.1002/ece3.3593

Björk, B.-C., Laakso, M., Welling, P., & Paetau, P. (2014). Anatomy of green open access. Journal of the Association for Information Science and Technology, 65(2), 237–250. https://doi.org/10.1002/asi.22963

Boettiger, C., & Eddelbuettel, D. (2017). An introduction to rocker: Docker containers for R. ArXiv Preprint ArXiv:1710.03675.

Cagnacci, F., Boitani, L., Powell, R. A., & Boyce, M. S. (2010). Animal ecology meets GPS-based radiotelemetry: A perfect storm of opportunities and challenges. Philosophical Transactions of the Royal Society B: Biological Sciences, 365(1550), 2157–2162. https://doi.org/10.1098/rstb.2010.0107

Culina, A., van den Berg, I., Evans, S., & Sánchez-Tójar, A. (2020). Low availability of code in ecology: A call for urgent action. PLOS Biology, 18(7), e3000763. https://doi.org/10.1371/journal.pbio.3000763

Ellison, A. M. (2010). Repeatability and transparency in ecological research Published by : Wiley on behalf of the Ecological Society of America Stable URL : http://www.jstor.org/stable/27860827 Repeatability and transparency in ecological research. Ecology, 91(9), 2536–2539.

Fauchald, P., & Tveraa, T. (2003). Using first-passage time in the analysis of area-restricted search and habitat selection. Ecology, 84(2), 282–288. https://doi.org/10.1890/0012-9658(2003)084[0282:UFPTIT]2.0.CO;2

Gandrud, C. (2013). Reproducible research with R and R studio. CRC Press.

Garriga, J., Palmer, J. R. B., Oltra, A., & Bartumeus, F. (2016). Expectation-Maximization Binary Clustering for Behavioural Annotation. PLOS ONE, 11(3), e0151984. https://doi.org/10.1371/journal.pone.0151984

Geng, W., Cosman, P., Baek, J.-H., Berry, C. C., & Schafer, W. R. (2003). Quantitative Classification and Natural Clustering of Caenorhabditis elegans Behavioral Phenotypes. Genetics, 165(3), 1117–1126. https://doi.org/10.1093/GENETICS/165.3.1117

Guéguen, L. (2009). Computing the likelihood of sequence segmentation under Markov modelling. Journal of Animal Ecology, 69–84. http://arxiv.org/abs/0911.3070

Guéguen, L. (2001). Segmentation by maximal predictive partitioning according to composition biases. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2066, 32–44. https://doi.org/10.1007/3-540-45727-5_4

Gurarie, E., Andrews, R. D., & Laidre, K. L. (2009). A novel method for identifying behavioural changes in animal movement data. Ecology Letters, 12(5), 395–408. https://doi.org/10.1111/j.1461-0248.2009.01293.x

Hampton, S. E., Strasser, C. A., Tewksbury, J. J., Gram, W. K., Budden, A. E., Batcheller, A. L., Duke, C. S., & Porter, J. H. (2013). Big data and the future of ecology. In Frontiers in Ecology and the Environment (Vol. 11, Issue 3, pp. 156–162). John Wiley & Sons, Ltd. https://doi.org/10.1890/120103

Heesen, R. (2018). Why the Reward Structure of Science Makes Reproducibility Problems Inevitable. The Journal of Philosophy, 115(12), 661–674. https://doi.org/10.5840/jphil20181151239

Joo, Rocio, Boone, M. E., Clay, T. A., Patrick, S. C., Clusella-Trullas, S., & Basille, M. (2020). Navigating through the r packages for movement. Journal of Animal Ecology, 89(1), 248–267.

Joo, Rocío, Picardi, S., Boone, M. E., Clay, T. A., Patrick, S. C., Romero-Romero, V. S., & Basille, M. (2020). A decade of movement ecology. ArXiv Preprint ArXiv:2006.00110.

Kays, R., Crofoot, M. C., Jetz, W., & Wikelski, M. (2015). Terrestrial animal tracking as an eye on life and planet. Science, 348(6240), aaa2478–aaa2478. https://doi.org/10.1126/science.aaa2478

Kelly, C. D. (2006). The Quarterly Review of Biology- Replicating empirical research in behavioral ecology: how and why it should be done but rarely ever is. In The Quarterly Review of Biology (Vol. 81, Issue 3).

Kitzes, J., Turek, D., & Deniz, F. (2017). The practice of reproducible research: case studies and lessons from the data-intensive sciences. Univ of California Press.

Knuth, D. E. (1984). Literate programming. The Computer Journal, 27(2), 97–111.

Konkol, M., Kray, C., & Pfeiffer, M. (2019). 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. https://doi.org/10.1080/13658816.2018.1508687

Lai, J., Lortie, C. J., Muenchen, R. A., Yang, J., & Ma, K. (2019). Evaluating the popularity of R in ecology. Ecosphere, 10(1), e02567. https://doi.org/10.1002/ecs2.2567

Langrock, R., King, R., Matthiopoulos, J., Thomas, L., Fortin, D., & Morales, J. M. (2012). Flexible and practical modeling of animal telemetry data: Hidden Markov models and extensions. Ecology, 93(11), 2336–2342. https://doi.org/10.1890/11-2241.1

Lewis, K. P., Vander Wal, E., & Fifield, D. A. (2018). Wildlife biology, big data, and reproducible research. Wildlife Society Bulletin, 42(1), 172–179. https://doi.org/10.1002/wsb.847

Michener, W. K. (2015). Ecological data sharing. In Ecological Informatics (Vol. 29, Issue P1, pp. 33–44). Elsevier B.V. https://doi.org/10.1016/j.ecoinf.2015.06.010

Mislan, K. A. S., Heer, J. M., & White, E. P. (2016). Elevating The Status of Code in Ecology. In Trends in Ecology and Evolution (Vol. 31, Issue 1, pp. 4–7). Elsevier Ltd. https://doi.org/10.1016/j.tree.2015.11.006

Morales, J. M., Haydon, D. T., Frair, J., Holsinger, K. E., & Fryxell, J. M. (2004). Extracting more out of relocation data: Building movement models as mixtures of random walks. Ecology, 85(9), 2436–2445. https://doi.org/10.1890/03-0269

National Academies of Sciences, Engineering, and M. (2019). Reproducibility and Replicability in Science. In Reproducibility and Replicability in Science. National Academies Press. https://doi.org/10.17226/25303

Obels, P., Lakens, D., Coles, N. A., Gottfried, J., & Green, S. A. (2020). Analysis of Open Data and Computational Reproducibility in Registered Reports in Psychology. Advances in Methods and Practices in Psychological Science, 3(2), 229–237. https://doi.org/10.1177/2515245920918872

Patil, P., Peng, R. D., & Leek, J. T. (2016). Title: A statistical definition for reproducibility and replicability Authors: Prasad Patil, Roger D. Peng, Jeffrey T. Leek. BioRxiV, 8–13.

Patterson, T. A., Parton, A., Langrock, R., Blackwell, P. G., Thomas, L., & King, R. (2017). Statistical modelling of individual animal movement: an overview of key methods and a discussion of practical challenges. AStA Advances in Statistical Analysis, 101(4), 399–438. https://doi.org/10.1007/s10182-017-0302-7

Peikert, A., & Brandmaier, A. M. (2019). A reproducible data analysis workflow with R Markdown, Git, Make, and Docker.

Peng, R. (2015). The reproducibility crisis in science: A statistical counterattack. Significance, 12(3), 30–32. https://doi.org/10.1111/j.1740-9713.2015.00827.x

Piccolo, S. R., & Frampton, M. B. (2016). Tools and techniques for computational reproducibility. In GigaScience (Vol. 5, Issue 1, p. 30). BioMed Central Ltd. https://doi.org/10.1186/s13742-016-0135-4

Poisot, T. (2015). Best publishing practices to improve user confidence in scientific software. Ideas in Ecology and Evolution, 8(1).

Powers, S. M., & Hampton, S. E. (2019). Open science, reproducibility, and transparency in ecology. Ecological Applications, 29(1), e01822. https://doi.org/10.1002/eap.1822

R Core Team. (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.r-project.org/

Reichman, O. J., Jones, M. B., & Schildhauer, M. P. (2011). Challenges and opportunities of open data in ecology. Science, 331(6018), 703–705.

Sandve, G. K., Nekrutenko, A., Taylor, J., & Hovig, E. (2013). Ten Simple Rules for Reproducible Computational Research. PLoS Computational Biology, 9(10), e1003285. https://doi.org/10.1371/journal.pcbi.1003285

Smith, R. (2006). Peer review: a flawed process at the heart of science and journals. In J R Soc Med (Vol. 99). www.rsmpress.co.uk],

Stodden, V., Guo, P., & Ma, Z. (2013). Toward Reproducible Computational Research: An Empirical Analysis of Data and Code Policy Adoption by Journals. PLoS ONE, 8(6), e67111. https://doi.org/10.1371/journal.pone.0067111

Stodden, V., Seiler, J., & Ma, Z. (2018). An empirical analysis of journal policy effectiveness for computational reproducibility. Proceedings of the National Academy of Sciences of the United States of America, 115(11), 2584–2589. https://doi.org/10.1073/pnas.1708290115

Swan, A., Gargouri, Y., Hunt, M., & Harnad, S. (2015). Open Access Policy: Numbers, Analysis, Effectiveness. http://arxiv.org/abs/1504.02261

Warwick-Evans, V., Atkinson, P., Gauvain, R., Robinson, L., Arnould, J., & Green, J. (2015). Time-in-area represents foraging activity in a wide-ranging pelagic forager. Marine Ecology Progress Series, 527, 233–246. https://doi.org/10.3354/meps11262

Way Community, T. T., Arnold, B., Bowler, L., Gibson, S., Herterich, P., Higman, R., Krystalli, A., Morley, A., O'Reilly, M., Whitaker, K., Way Community, T. T., Arnold, B., Bowler, L., Gibson, S., Herterich, P., Higman, R., Krystalli, A., Morley, A., O'Reilly, M., & Whitaker, K. (2019). The Turing Way: A Handbook for Reproducible Data Science. Zndo. https://doi.org/10.5281/ZENODO.3233986

Wilkinson, M. D. (2016). Comment: The FAIR Guiding Principles for scientific data management and stewardship. Nature Publishing Group. https://doi.org/10.1038/sdata.2016.18

BIOGRAPHICAL SKETCH

Jenicca is originally from Mauritius, an island found on the east of Madagascar in the Indian Ocean. She started with a Bachelor’s degree in South Africa at the Nelson Mandela University, then made her way to the University of Cape Town for a Master’s in Ecological Statistics where she joined the SEEC - Centre for Statistics in Ecology, Environment and Conservation: a research group within the Department of Statistical Sciences at the University of Cape Town. She then switched to behavioral ecology and is now graduating with a second Master’s in Interdisciplinary Ecology (2021) from the University of Florida. Her work focuses on reproducibility in movement ecology and she is the researcher behind this study.