

PhD Research Proposal

Elise Gould

2018-05-23

Keywords:

- conservation
- decision-making
- ecology
- reproducibility
- structured decision-making

Introduction

Successful biodiversity conservation and management is underpinned by effective and robust decision-making (Mukherjee et al. 2018). Decision-makers are tasked with allocating limited resources in the face of uncertainty about the effectiveness of alternative management interventions, and incomplete or inadequate scientific information. Moreover, environmental decisions often must be made in complex socio-economic and political contexts, with multiple stakeholders and multiple and/or competing objectives. Formal, structured approaches to decision-making under uncertainty, such as Structured Decision Making, are both implicitly and explicitly espoused as a pathway to more robust and reproducible conservation decisions (e.g. J. L. Moore and Runge 2012; Converse et al. 2013; Schreiber et al. 2004). However, this claim to reproducibility remains untested, and a more nuanced consideration of what exactly reproducibility means in this application is lacking. Given that science is said to be in the grips of a “reproducibility crisis”, and that biased and unfair scrutiny in the broader socio-political climate of “alternative facts” threatens credibility of scientific knowledge (Baillieul, Grenier, and Setti 2018), it is both essential and timely that the reproducibility of decision-support in ecology and conservation is given rigorous research attention.

Failure to reproduce a large proportion of published studies in the fields of psychology and medicine has received considerable attention and provoked heated discussion among researchers in the broader scientific community. Scientific credibility is underpinned by the assumption that findings are both real and replicable (Nakagawa and Parker 2015). As the recent “reproducibility crisis” debate illustrates, this assumption is often incorrect (Baker 2016). For example, the Open Science Collaboration’s Psychology Reproducibility Project reported that less than half of published results could be reproduced (Open Science Collaboration 2015). Although large-scale direct replications are largely absent in ecology (Nakagawa and Parker 2015), an initial assessment of the conditions found to foster reproducibility problems provide evidence that ecology as a discipline is at risk of a “reproducibility crisis” (Fidler et al. 2017).

Research scope: reproducibility *of* decisions versus reproducibility *for* decision decisions

Reproducibility research in ecology has largely focused on the primary evidence-base. I draw on the conservation decision-making (CDM) literature to distinguish between knowledge generated at the level of the individual study level (“primary evidence base”), and the translation of that information into a decision (Dicks, Walsh, and Sutherland 2014), whereby evidence is collated, synthesised, and transformed into decisions, ideally using some formal decision support system (Gardner et al. 2018). Most decision support systems developed using a structured framework utilise one or more “decision tools” at various steps in the decision-making process, comprising conceptual or mechanistic models, representing all or part of the system

in question (Dicks, Walsh, and Sutherland 2014, see discussion point in Chapter 3 for a definition of the distinction between tools and systems).

One exception to the focus on the primary evidence base, is Morrison et al.'s (2016) replication study of the Population Viability Analysis literature. They discuss the potential implications of the non-reproducibility of studies and models on the decision-making process. Firstly, non-reproducible predictions undermine the credibility of the original model, and therefore any decisions informed by that evidence. Any predictions are therefore unreliable, and may result in finite resources being directed inefficiently, and in opportunity costs. Thirdly, the effectiveness of an intervention may not be able to be evaluated against the predictions of the model if they are not reproducible. The ability to compare the outcome of an intervention against model predictions, and hence to measure progress towards the conservation objectives, is especially important in ecology and conservation: true randomized experimental design is often infeasible, such that the performance of a conservation action is measured against the counter-factual, which is often estimated using predictive models (Law et al. 2017). Finally, if the original model predictions cannot be reproduced, the decision-maker is prevented from comparing and evaluating revised models with updated parameters against previous models. This is relevant for contexts where decision-making is informed by ongoing monitoring programmes, such as in adaptive management-based conservation.

This work is a good initial consideration of how reproducibility issues might impact decisions, and it echoes work in the conservation decision-making (CDM) literature underscoring the importance of a 'robust' evidence-base for informing management (Dicks, Walsh, and Sutherland 2014; Law et al. 2017; Sutherland and Wordley 2017; Gardner et al. 2018). While reproducibility for decisions – or the reproducibility of either raw evidence and/or decision-tools generating information informing decisions – is important, it fails to consider the broader decision process in which conservation decisions are made, and explicitly assumes that if the evidence-base is reproducible, then this should translate to reproducible decisions. Yet, decision-making is a "human enterprise" shaped by values and expectations (Mukherjee et al. 2018) and failing to integrate the human elements of decision-making in conservation may lead to sub-optimal outcomes (N. J. Bennett et al. 2017). For this reason, I argue that reproducibility research in ecology and conservation should certainly continue advancing for the primary evidence-base, however, the scope of research should extend beyond evidence-generation, to examine the systematic use and synthesis of that evidence to inform decision-making. The focus of my PhD will therefore be on the reproducibility of decisions: or whether the entire decision-process can be repeated to reproduce the same decision.

Below I propose some examples of 'decision-points' within the broader decision process that may impede the reproducibility of a decision. These are points of departure or variation occurring within the development of a DSS that may result in a different decision-outcome, should the process be repeated – either by the same analyst in the future, or by another independent analyst.

1. Problem formulation is a critical phase of the decision process but is often neglected or at least poorly documented when it comes to published applications of decision support systems in ecology and conservation. Human values shape preferences about the acceptability of decision outcomes in terms of fulfilling conservation objectives and should therefore be properly captured in objectives and performance measures (Conroy and Peterson 2013). Decision modelling outputs may be sensitive to the specification of the performance measure (2014). Failing to properly incorporate decision-maker and stakeholder values may therefore influence the reproducibility of a decision, resulting in sub-optimal decisions that do not adequately meet the true objectives.
2. Scenario analyses are often used to make a case for implementing one management strategy over others. However, Law et al. (2017) emphasise that the selection of scenarios may influence inferences about the appropriate course of action – choosing scenarios that are impossible or highly improbable may yield the impression that a particularly positive or negative outcome is likely. Thus, it is not only the modelling inputs into decisions that should be scrutinised in terms of reproducibility, but also the choice scenarios we wish to run through them.

Research aims and objectives

This major goal of this PhD is to investigate the transparency and reproducibility of decisions generated from structured approaches to conservation decision-making (such as Structured Decision Making, SDM). The first aim of this research is to expand and develop a conceptual understanding of “reproducibility” that is fit for application in conservation decision-making. Existing approaches to reproducibility almost exclusively focus on hypothesis-testing based research, however in applied ecology and conservation technical approaches draw primarily on decision-analytic tools and methods for decision-making. This typology of reproducibility will be sensitive to the particularities of the data, methods and questions commonly encountered within this context. By taking decision-support systems (DSS’s) as its unit of interest, this research aims to identify critical points in the decision process that threaten the reproducibility of ecological decisions and undermine conservation success. Secondly, I aim to systematically review the published literature on decision support systems to evaluate the magnitude and extent of the “reproducibility crisis” in translational research of ecology and conservation. This review will allow the development of a set of guidelines or a check list for improving the transparency and reproducibility of decision support systems in ecology and conservation. Finally, I aim to use real-world decision problem case studies to investigate the replication of decision support systems.

Reproducibility for decision support in ecology and conservation: towards a formal typology

Surveying the reproducibility literature in ecology and conservation

There appear to be two strands of reproducibility research in ecology. The first, termed “reproducible research” by some (Peng 2009; Gandrud 2016), targets the research practices of individuals, and considers reproducibility to be largely computational in nature. It draws on the computational biology and scientific computing literature to proffer technological solutions with an emphasis on software and data management practices. It is closely linked to work within the open data movement in ecology. The second and more recently emerging major strand of reproducibility research draws on meta-research approaches within psychology, medicine, and science more broadly.

“Reproducible research” focuses on the data analysis pipeline for a single analysis or study, taking the starting point for reproducibility to be where analysis is ready to begin (data is on hand and probably digitised). These approaches seem to resonate among ecologists, with much reflection and discussion in the grey literature, particularly in blog and twitter format. Practices for improving computational reproducibility have included version control with incremental changes, functional programming and unit-testing Wilson et al. (2014). Others have emphasised the use of literature programming techniques and containerised analyses that link code to their inferences in a single document (Baumer and Udwin 2015; Gandrud 2016). Noble et al. (2009) describe how to organise computational biology projects while the British Ecological Society have published complete guides to reproducible code and data management (British Ecological Society 2014; Croucher et al. 2017).

The “open data” movement within ecology, and science more broadly, assumes more of a systems-level view of reproducibility: situating the reproducibility of individual studies within the broader life-cycle of scientific research. Some computational ecology research in this context is still targeted at improving the research practices of individuals, however, the overarching aim is to improve the computational reproducibility or research in order to facilitate large-scale meta-analyses (Merali 2010), longitudinal studies and data synthesis (E. White et al. 2013). Solutions arising from the open data movement are a mixture of technological and policy-based. For instance, White et al. (2013) discuss the particularities of ecological data and describe methods for ensuring data is re-usable to others; while Ram (2013) illustrates how git and GitHub can facilitate greater reproducibility and increased transparency. Similarly, Whitlock (2011) describes the technical and cultural issues that must be addressed to increase data-archiving practices in ecology and evolution, including journal policy. For most in the open data movement, the responsibility for reproducibility is shared among

individual researchers, journal editors and reviewers for improving reproducibility via openness practices and policy (Morueta-Holme et al. 2018).

Meta-research approaches include the identification of the types of, and evaluation of the extent of questionable research practices (QRPs) across whole disciplines (e.g. 2018), as well as understanding the broader set of conditions fostering their occurrence and prevalence, such as a lack of transparency and publication bias in a publish-or-perish environment (Parker et al. 2016). Large-scale meta-analyses are a new and emerging technique for evaluating whole fields for reproducibility, but have not been undertaken in ecology and evolution to date (Fidler et al. 2017), even though traditional meta-analyses are not new (G. Stewart 2010). Conceptual understandings of reproducibility extend beyond computational reproducibility, with numerous ‘typologies’ of reproducibility being suggested. Within such typologies, each type of reproducibility is understood as performing its own epistemic function. Solutions to mitigating reproducibility issues tend to target a community of practice, or the incentive structure, rather than individuals. For example, pre-registration of study designs may protect against QRPs including ‘p-hacking’, selective reporting and cherry-picking or HARKing (hypothesizing after results are known) (Parker et al. 2016). Tools for Transparency in Ecology and Evolution (<https://osf.io/g65cb/>) provide a checklist to authors, reviewers and journal editors for improving reporting transparency. There is increasing uptake of checklists, including *Nature Ecology and Evolution* (“A checklist for our community” 2018) and new guidelines for statistical practice for *Conservation Letters* (Fidler et al. 2018).

Unpacking “reproducibility of decisions”

The term “reproducibility” broadly encompasses the ability to “replicate an experiment or study and/or its outcomes” (Fidler et al. 2017). In ecology and conservation decision-making, researchers the terms ‘repeatable’, ‘replicable’ and ‘reproducible’ are often used interchangeably and transparency is frequently touted as both a key-feature of environmental decision-support systems, and as central to delivering reproducible decisions (Kim et al. 2016). However, the discussion of reproducibility in these contexts usually stops here, and there is no further discussion as to some definition of the term. McIntosh et al. (2011) elaborate further than most to explain the link between transparency and reproducibility in formal decision support for ecology:

Transparent because rational explanations can be provided to support decisions, and because the user/stakeholder/citizen can reproduce the decision procedure, play with the weights, and perform sensitivity analysis to assess decision strength and robustness.

Examining this statement reveals that, at least for these authors, reproducibility is simply the repeatability of the decision procedure, or is reproducible ‘in-principle’. A study is replicable in principle if its methods, procedures and analysis are described transparently with sufficient detail to be repeated; whereas a study is replicable in practice if the original results can be reproduced exactly or with sufficient similarity (Fidler and Wilcox, 2018 in prep.). The repeatability of a decision procedure does not necessarily ensure that the resultant “decision” from the procedure can be reproduced; the above statement seems to imply that the transparency of structured approaches to decision making guarantees in-practice reproducibility, because it is reproducible in-principle. This assumption needs to be further investigated, especially given how commonly the inherent methodological transparency of SDM and other decision-analytic approaches is used as a justification for their use.

The above use of the term ‘robustness’ in the context of reproducibility of decisions also demands attention. Transparency and in-principle reproducibility of decisions generated from formal approaches to decision-making are commonly associated with delivering ‘robust’ decisions or management approaches (e.g. Schreiber et al. 2004). Few explicitly define what a ‘robust decision is’ – they are implicitly synonymous with ‘good’ or ‘accurate’, or else, are described as being ‘robust to uncertainty’ (J. L. Moore and Runge 2012; Converse et al. 2013). That is, robust management decisions are those that can tolerate some acceptable degree of uncertainty before the optimal decision changes (Regan et al. 2005).

Given the linguistic uncertainty around the properties of “reproducibility”, “transparency” and “robustness” of decision support systems in ecology and conservation, a more thorough treatment should be given to defining the reproducibility of decisions. There are a number of existing typologies of reproducibility (e.g.

Nakagawa and Parker 2015; Patil, Peng, and Leek 2016; Gómez, Juristo, and Vegas 2010) that may serve as the basis for formally defining a conceptual framework for the reproducibility of decisions. Each typology describes the different ways in which a replication study may differ from the original, with some being more finely specified than others. For example, Patil et al.’s (2016) typology is useful for highly computational analyses, outlining points of difference including the population, question, hypothesis, experimental design, experimenter, analysis plan, analyst, code, and estimate and claim. Gomez’ (2010) typology is less finely grained in terms of the classes of difference between a replicate and original; specifying site, experimenter, apparatus, operationalisations, and population properties.

Gomez’ typology systematically considers how varying different subsets of those classes serves different epistemic functions, whereas Pattil et al. give minimal consideration of the epistemic functions of different types of reproducibility, distinguishing only between what they term “reproducible” and “replicable” studies (“reproducible”: *Given a population, hypothesis, experimental design, experimenter, data, analysis plan, and code you get the same parameter estimates in a new analysis*, versus. “replicable”: *Given a population, hypothesis, experimental design, and analysis plan you get consistent estimates when you recollect data and redo the analysis*). In contrast, both Gomez’ and Nakagawa and Parker’s typologies centre on the epistemic functions of generality and validity, which exist on a continuum. Nakagawa provide examples as to how these concepts might apply to ecological phenomena.

What Pattil et al. term ‘reproducible’, has been defined as “reproducible research”, or a set of minimum standards requiring that data and code is made available for others to verify results of the original study and for re-use of the data (Peng 2009). This type of “computational reproducibility” is advocated for as the minimum standard when there is a lack of time or resources for intensive replications (Patil, Peng, and Leek 2016; Williams, Bagwell, and Zozus 2017). As described earlier, “reproducible research” with an emphasis on computational reproducibility comprises the bulk of the reproducibility literature in ecology, perhaps because replications in ecology have typically been considered infeasible due to the inherent spatio-temporal variation in nature, or even unethical for work on threatened species (Nakagawa and Parker 2015).

Chapter One: Defining a typology of reproducibility for decision support systems in conservation and ecology

A major aim of this PhD is to devise a typology of reproducibility that is applicable to decision support systems in ecology and conservation. A formal definition of the type, measure and purpose of reproducibility of decisions is necessary for evaluating the reproducibility of decision support systems. I aim to draw on work emerging within the second more recent strand of reproducibility research in ecology and conservation, advancing conceptual understandings beyond “computational reproducibility” and “reproducible research”. I propose that the typology will synthesise some combination of existing typologies, such as the three described above. For example, Pattil et al.’s (2016) points of difference may prove useful for considering phases of the SDM process relying on complex mathematical models; while Gomez (2010) and Nakagawa and Parker (2015) may aid in understanding the epistemic functions of different types of reproducibility of decisions.

Most formal decision-support systems for ecology and conservation lie outside of the typical hypothetico-deductive model of science. Although the concept of reproducibility is not inherently specific to that model, the causes and solutions to the reproducibility crisis have focused exclusively on this mode. For instance, QRP’s including p hacking, cherry picking, and HARKing are specific to Null-hypothesis significance testing (NHST) (Fraser et al. 2018). Solutions to the reproducibility crisis, such as preregistration of analysis plans, are specifically targeted at addressing the cognitive biases that foster these QRPs, and so also adhere to this model of science. Measures of reproducibility, such as those employed by the COS reproducibility project have also focused on the hypothetico-deductive model of science, but not all.

The first step in building the typology is to define what constitutes a ‘successful’ replication of a decision support system. An obvious candidate is the final decision outcome, or the decision recommended at the end of the decision process. However, this will need careful consideration, depending on the type of replication being undertaken. For instance, suppose a ‘conceptual replication’ be undertaken to reproduce a decision support system for a particular decision-context. If the problem formulation phase is not constrained, it

is possible that the set of potential decision alternatives might not even match those in the original study. This may occur because analysts vary in their ability to properly elicit the problem specification from decision-makers. The fundamental objective(s) might not be properly specified, or else the analyst might be subject to ‘evidence complacency’ (see Sutherland and Wordley 2017) or be anchored by the alternatives proposed by the decision-maker, ignoring existing evidence and knowledge about the full range of potential decision alternatives for a given problem. The replicate and the original study would not be comparable in this situation, if the decision outcome is the measure of reproducibility.

Following that, key points of divergence in the decision process should be identified. I suspect these will map closely onto the different phases of the decision-framework being used. This is to facilitate the identification of different types of reproducibility, each serving different epistemic functions. For each type of reproducibility deemed relevant to conservation decision-making, the set of constraints and conditions constituting a successful replication must be carefully specified, in reference to its epistemic function. Some thought will also need to be given to how the epistemic functions are relevant to the context of conservation decision-making. For example, the importance of validity seems obvious – a replication supporting the validity of a decision recommended by the original DSS would also lend credibility to the decision process that led to that decision. However the relevance of generality to this context is not so obvious – does a conceptual replication that converges on the same decision as the original DSS give validity to the original decision, and / or does it tell us something about the applicability of the particular suite of tools built into the DSS for a given decision-context?

At present, the format this work will assume is unclear. It could be embedded in other chapters (Chapter 3; systematic review, or Chapters 4 and 5: the replication case study), or constitute a standalone chapter.

Identifying reproducibility issues for decision support in ecology and conservation: non-hypothesis testing based research

The reproducibility literature has focused exclusively on hypothesis-testing, whether that be Bayesian or frequentist. This also applies to initial research focusing on ecology and evolution. However, Fidler (2016) correctly identifies that in applied ecological research, particularly in conservation science, non-hypothesis testing methods, such as decision-theory, cost-effectiveness analysis, optimization and other scientific computing methods are common. These approaches come with their own set of reproducibility issues. However, a full understanding of the types of reproducibility issues, as well as their impact on either the evidence-base, or the decisions informed by the evidence-base, is yet to emerge.

In what is likely the first reproducibility study in conservation, Morrison et al. (2016) evaluate the Population Viability Analysis (PVA) literature by performing direct tests of repeatability and reproducibility: they found that poor model parameter reporting practices meant a large number could not even be reproduced in-principle, let alone in-practice. This work marks an important step towards identifying reproducibility issues relevant not only to PVA-based research, but also to CDM more broadly. Outside of ecology, where reproducibility research has received much more attention and focus, I could only identify a single paper that considered non-NHST based work in the reproducibility discussion. Crutzen and Peters (2017) suggest re-terming “power” and “underpowered” studies as being “undersamplesized”, in a move to embrace the disciplinary shift away from NHST towards inferences based on confidence intervals for effect size estimates. They exemplify such research as studies aiming to obtain accurate power estimates. Although their treatment of the issue is rather cursory, it is promising to see the scope of reproducibility research expanding to non-NHST approaches in broader science. However, it is pertinent that reproducibility research in ecology and CDM: we cannot evaluate the reproducibility of this body of work without first identifying relevant types of reproducibility issues.

Chapter Two: QRPs for non-hypothesis testing research

The first output for this research is to generate a “roadmap” of reproducibility issues, biases and questionable research practices (QRPs) that are encountered when developing Decision Support Systems in ecology and conservation. In this chapter I will attempt to identify where exactly in the DSS development process particular biases or ‘decision-points’ are likely to occur. This work should serve as a launching point for proposing standards and technical solutions targeted at individual (or teams of) researchers / analysts, with the broader objective of minimising the extent and magnitude of questionable research practices for DSS.

This chapter will constitute (from what I am aware of) the first attempt to investigate reproducibility issues for non-hypothesis testing based research. To this extent, the knowledge resulting from this chapter should also be applicable to other fields of science beyond ecology and conservation where translational research and non-hypothesis testing methods are prevalent. It will also build on the work of Fraser et al. (2018) in advancing research on reproducibility and transparency in ecology.

Methods for generating the roadmap

1. Sketch out the SDM / DSS development process, but breakdown into modelling steps if necessary. As a starting point, I will take the Structured Decision Making process as the overarching framework for building a Decision Support System. a. Does this process differ for different tools / decision frameworks?
2. Identify sources of bias AND QRPs that others have identified in ecology and evolution, but also in other scientific disciplines and translational research fields a. at the individual DSS level b. at a higher level, e.g. in the evidence-base d. Are these biases / QRPs applicable to DSS’s? e. are there other biases unique to DSS’s that haven’t been considered? f. Is their occurrence specific to the type of tool or application under consideration? i.e. are some tools more robust to biases / QRPs than others?
3. Map these biases / QRPs onto 1, where do they occur at the various decision-points?

I plan on utilising the outputs of the reproducibility discussion session planned for the Qaecera Retreat. The aim of the session is to initiate awareness and discussion among Qaecologists and Cebranalysts about reproducibility in our research practices, with the subsequent aim of equipping people with solutions to minimise or overcome these issues. The format for the session will involve structured / facilitated discussion in small break-out groups. Given the collaborative nature of this work, participants will be invited to co-author the paper after the retreat. This might also incentivise people to attend the session!

Evaluating the Transparency and likely reproducibility of decisions in ecology and conservation

Fidler et al. (2017) call for an assessment of the completeness and transparency of methodological and statistical reporting in journals for ecology. Such assessments should take the form of extensive journal surveys, with the aim of highlighting deficiencies in journal’s reporting policies.

Why is it important: Incomplete reporting impedes direct replication, meta-analysis, and also direct re-analysis projects. From personal experience, I have reason to believe that incomplete reporting is rife in environmental decision support systems. In working with Bayesian Belief Networks for catchment management, it was impossible to re-build all but one system model from the published literature. At best, the causal structure of the model is reproducible, but only one published study I encountered contained sufficient information so as to reproduce its parameterisation. Importantly, there is also the issue of understanding modeller choices about how and why variables were parameterised. This information is rarely recorded, and often the source of empirical data used in parameterisation is not reported. The next chapter of my research will take up this call, where I will conduct a systematic review aiming to systematically evaluate the decision support literature in ecology and conservation for its transparency and likely reproducibility.

Chapter Three: Systematic Review

The roadmap of QRPs developed in Chapter Two will inform both the scoping and evaluation criteria for the systematic review. The results of the review will be used to generate a ‘reproducibility check-list’ or set of criteria for DSS in ecology and conservation. The check-list will be aimed at individual researchers, but with the hope of being adopted by relevant journals. The check-list will need to balance a sufficiently acceptable set of standards against resource constraints on individual researchers in order to prevent the check-list from deterring important DSS work from being published.

Methods:

I will follow guidelines for undertaking a systematic review in the format and procedure established by Collaboration for Environmental Evidence (CEE) Evidence Synthesis (Centre for Environmental Evidence, n.d.). I chose this method of systematic review partially because the topic area is relevant (environmental management), and partially because of the comprehensiveness of the guidelines: the guidelines provides detailed and systematic methods for developing the search strategy, inclusion / exclusion criteria, and coding criteria for evaluating the literature, including pilot searches for refining criteria. The other draw card was the enforced development and pre-registration of the protocol for conducting the review in order to prevent ‘mission creep’ as the review progresses.

In addition, I will use the PRISMA (Preferred Reporting Items for Systematic Reviews, (Moher et al. 2015)) statement for reporting, as recommended by Nakagawa and Poulin (2012). The PRISMA check-list contains a check-list of 27 reporting items, as well as a flow diagram that visualizes the database searching procedure as well as decisions for including and evaluating studies. The aim of the statement is to increase the transparency of the literature review process.

Review scope:

I distinguish between ‘decision *frameworks*’ and ‘decision *tools*’: frameworks may be considered the structured set of procedures governing the approach to an entire decision process - i.e. from problem formulation, to modelling the consequences of decision alternatives, to evaluating alternatives. Structured Decision Making (SDM) is an example of a decision framework (Bower et al. 2017). Decision tools may be understood as the particular procedures or methods used to solve or provide inputs to inform a particular phase in the overarching decision framework. For example, Bayesian Networks, Population Viability Analyses and Multi- Criteria Decision Analysis constitute decision tools. This review will investigate the reproducibility of decisions generated from ‘decision *frameworks*’, rather than just the reproducibility of ‘decision *tools*’.

Analysis and Reporting:

By focusing on decision frameworks more broadly – rather than a particular framework, like SDM – I hope to be able to evaluate whether different frameworks might influence the transparency and likely reproducibility of resultant decisions. This should also provide a means of testing the claim that SDMs increase transparency, and therefore the reproducibility of decisions. Importantly, in not restricting the scope too heavily, this provides a way for dealing with linguistic ambiguity around the language of decision support in ecology and conservation. For example, there is considerable slippage between the terms ‘decision support system’ and ‘decision support tools.’ Although SDM approaches are very common in conservation, restricting the inclusion of studies to nominally SDM studies will likely omit much research that might be decision-analytic in approach, and therefore still constitute a comparable structured or formal framework for decision-making. Following a pilot test of the searching and inclusion criteria, the scope might be revised (i.e. if too many papers are returned).

I am presently unclear on what analyses to conduct. But they will probably include the presentation of some descriptive statistics based around critical evaluation criteria.

Chapter Four: case studies in replicating DSS’s

The final chapters will involve the use of case studies of real-world ecological and conservation decisions. The objective of the case studies is to demonstrate an application of one or several types of reproducibility proposed in the typology in Chapter One. It would also serve as a proof-of-concept for replicating a decision

support system. Depending on the degree of complexity and size of the case studies, I could compare a selection of different DSS problems. The exact form of the case studies / experiments will depend on the results from previous chapters, so their specification here remains purposefully vague.

Environmental water allocations

Chris Jones has been tasked with demonstrating the benefit of environmental water allocations across Victoria and is seeking collaboration to develop this project. It encompasses multiple river systems, each with multiple management agencies, and so presents the opportunity for a replication experiment, perhaps taking a ‘red card’ approach (Fidler et al. 2017). This could involve giving a problem brief with 2 - 3 objectives, for example, to different teams of researchers, asking them to develop a decision support system, choosing which ever approach or tools that they like. They could be surveyed afterwards, asking them to describe what they did and why. Alternatively, they could write up their methods section for submission afterwards. One issue is that we need to entice people to participate. This could be using funds (which are available). Another approach would be to offer co-authorship.

Importantly, because this is a real-world problem, I will need to ensure that Chris also benefits from this work. Given that Chris’ goals are currently unclear, it is difficult to specify now. However, it could potentially serve as a form of sensitivity analysis, providing evidence about the robustness of different models, but also about the processes used to derive those models. This type of replication analysis would be able to test whether the decision is reasonably robust to different processes, or tools.

Threatened Woodland Recovery Planning

This project aims to develop recovery plans for threatened woodland communities by generalising across communities without losing critical aspects of the structure and functioning of each community. The form that this work might take is unclear to me, however, the decision problem poses an interesting case study for work on reproducibility in both ecology and conservation decision-making, because it allows us to explore questions around the role of replications in terms of generality.

Timeline of activities and goals

Attached.

References

- “A checklist for our community.” 2018. *Nature Ecology & Evolution* 2 (6). Nature Publishing Group: 913–13. doi:10.1038/s41559-018-0574-7.
- Baillieu, John, Gerry Grenier, and Gianluca Setti. 2018. “Reflections on the Future of Research Curation and Research Reproducibility [Point of View].” *Proceedings of the IEEE* 106 (5): 779–83. doi:10.1109/JPROC.2018.2816618.
- Baker, Monya. 2016. “Is There a Reproducibility Crisis?” *Nature.com*. https://www.nature.com/polopoly_fs/1.19970.1469695948!/menu/main/topColumns/topLeftColumn/pdf/533452a.pdf?origin=ppub.
- Baumer, Benjamin, and Dana Udwin. 2015. “R Markdown.” *Wiley Interdisciplinary Reviews: Computational Statistics* 7 (3): 167–77. doi:10.1002/wics.1348.
- Bennett, Nathan J, Robin Roth, Sarah C Klain, Kai Chan, Patrick Christie, Douglas A Clark, Georgina Cullman, et al. 2017. “Conservation social science: Understanding and integrating human dimensions to improve conservation.” *Biological Conservation* 205 (C). The Authors: 93–108. doi:10.1016/j.biocon.2016.10.006.
- Bower, Shannon D, Jacob W Brownscombe, Kim Birnie-Gauvin, Matthew I Ford, Andrew D Moraga, Ryan

- J P Pusiak, Eric D Turenne, Aaron J Zolderdo, Steven J Cooke, and Joseph R Bennett. 2017. “Making Tough Choices: Picking the Appropriate Conservation Decision-Making Tool.” *Conservation Letters* 11 (2): e12418–7. doi:10.1111/conl.12418.
- Centre for Environmental Evidence. n.d. **Guidelines and Standards for Evidence Synthesis in Environmental Management**. Edited by ANDREW S PULLIN, G K Frampton, B Livoreil, and G Petrokofsky. Volume 5. www.environmentalevidence.org/information-for-authors.
- Conroy, Michael J, and J T Peterson. 2013. *Decision Making in Natural Resource Management: A Structured, Adaptive Approach*. Wiley Blackwell.
- Converse, Sarah J, Clinton T Moore, Martin J Folk, and Michael C Runge. 2013. “A matter of tradeoffs: Reintroduction as a multiple objective decision.” *Journal of Wildlife Management* 77 (6). Wiley-Blackwell: 1145–56. <https://search.ebscohost.com/login.aspx?direct=true&db=eih&AN=89305806&site=ehost-live>.
- Croucher, Mike, Laura Graham, Tamora James, Anna Krystalli, and Francois Michonneau. 2017. *A Guide to Reproducible Code in Ecology and Evolution*. Edited by Natalie Cooper and Pen-Yuan Hsing. British Ecological Society.
- Crutzen, Rik, and Gjalt-Jorn Y Peters. 2017. “Targeting Next Generations to Change the Common Practice of Underpowered Research.” *Frontiers in Psychology* 8 (July): 67–64. doi:10.3389/fpsyg.2017.01184.
- Dicks, Lynn V, Jessica C Walsh, and William J Sutherland. 2014. “Organising evidence for environmental management decisions: a 4S hierarchy.” *Trends in Ecology & Evolution* 29 (11). Elsevier Ltd: 607–13. doi:10.1016/j.tree.2014.09.004.
- Fidler, Fiona, Yung En Chee, Bonnie Wintle, Mark A Burgman, Michael A McCarthy, and Ascelin Gordon. 2016. “Why and how to evaluate reproducibility of ecological research.” *BioScience*, June, 1–27. <https://osf.io/g7vha/>.
- Fidler, Fiona, Yung En Chee, Brendan A Wintle, Mark A Burgman, Michael A McCarthy, and Ascelin Gordon. 2017. “Metaresearch for Evaluating Reproducibility in Ecology and Evolution.” *BioScience*, January, biw159–8. doi:10.1093/biosci/biw159.
- Fidler, Fiona, Hannah Fraser, Michael A McCarthy, and Edward T Game. 2018. “Improving the transparency of statistical reporting in Conservation Letters.” *Conservation Letters* 11 (2). Wiley/Blackwell (10.1111): e12453–4. doi:10.1111/conl.12453.
- Fraser, Hannah Stephanie, Timothy H Parker, Shinichi Nakagawa, A Barnett, and Fiona Fidler. 2018. “Questionable Research Practices in Ecology and Evolution,” March, 1–24. doi:10.17605/OSF.IO/AJYQG.
- Gandrud, Christopher. 2016. *Reproducible Research with R and R Studio, Second Edition*. CRC Press. http://books.google.com.au/books?id=Ce35CQAAQBAJ&pg=PA10&dq=intitle:Reproducible+Research+with+R+and+RStudio+Second+Edition&hl=&cd=1&source=gbs_api.
- Gardner, Charlie J, Patrick O Waeber, Onja H Razafindratsima, and Lucienne Wilmé. 2018. “Decision complacency and conservation planning.” *Conservation Biology*, May, 1–10. doi:10.1111/cobi.13124.
- Giljohann, K M, K M Giljohann, M A McCarthy, M A McCarthy, L T Kelly, and L T Kelly. 2014. “Choice of biodiversity index drives optimal fire management decisions.” *Ecological Modelling*. <http://www.esajournals.org/doi/abs/10.1890/14-0257.1>.
- Gómez, Omar S, Natalia Juristo, and Sira Vegas. 2010. “Replications Types in Experimental Disciplines.” In *Proceedings of the 2010 Acm-Ieee International Symposium on Empirical Software Engineering and Measurement*, 3. ACM.
- Kim, Milena Kiatkoski, Louisa Evans, Lea M Scherl, and Helene Marsh. 2016. “Applying Governance Principles to Systematic Conservation Decision-Making in Queensland.” *Environmental Policy and Governance* 26 (6): 452–67. doi:10.1002/eet.1731.
- Law, Elizabeth A, Paul J Ferraro, Peter Arcese, Brett A Bryan, Katrina Davis, Ascelin Gordon, Matthew H Holden, et al. 2017. “Projecting the performance of conservation interventions.” *BIOC* 215 (September).

Elsevier: 142–51. doi:10.1016/j.biocon.2017.08.029.

McIntosh, B S, J C Ascough II, M Twery, J Chew, A Elmahdi, D Haase, J J Harou, et al. 2011. “Environmental decision support systems (EDSS) development - Challenges and best practices.” *Environmental Modelling & Software* 26 (12). Elsevier Ltd: 1389–1402. doi:10.1016/j.envsoft.2011.09.009.

Merali, Z. 2010. *Error: why scientific programming does not compute*. Nature. http://scholar.google.com/scholar?q=related:0nzms2QeCT0J:scholar.google.com/&hl=en&num=20&as_sdt=0,5.

Moher, David, Larissa Shamseer, Mike Clarke, Davina Gherzi, Alessandro Liberati, Mark Petticrew, Paul Shekelle, and Lesley A Stewart. 2015. “Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement.” *Systematic Reviews* 4 (1): 1. doi:10.1186/2046-4053-4-1.

Moore, Joslin L, and Michael C Runge. 2012. “Combining Structured Decision Making and Value-of-Information Analyses to Identify Robust Management Strategies” 26 (5): 810–20. doi:10.1111/j.1523-1739.2012.01907.x.

Morrison, Clare, Cassandra Wardle, and J Guy Castley. 2016. “Repeatability and Reproducibility of Population Viability Analysis (PVA) and the Implications for Threatened Species Management.” *Frontiers in Ecology and Evolution* 4 (August): 875–7. doi:10.3389/fevo.2016.00098.

Morueta-Holme, Naia, Meagan F Oldfather, Rachael L Olliff-Yang, Andrew P Weitz, Carrie R Levine, Matthew M Kling, Erin C Riordan, et al. 2018. “Best practices for reporting climate data in ecology.” *Nature Climate Change*, January. Springer US, 1–3. doi:10.1038/s41558-017-0060-2.

Mukherjee, Nibedita, Aiora Zabala, Jean Huge, Tobias Ochieng Nyumba, Blal Adem Esmail, and William J Sutherland. 2018. “Comparison of techniques for eliciting views and judgements in decision-making.” *Methods in Ecology and Evolution* 9 (1): 54–63. doi:10.1111/2041-210X.12940.

Nakagawa, Shinichi, and Timothy H Parker. 2015. “Replicating research in ecology and evolution: feasibility, incentives, and the cost-benefit conundrum.” *BMC Biology*, October. BMC Biology, 1–6. doi:10.1186/s12915-015-0196-3.

Nakagawa, Shinichi, and Robert Poulin. 2012. “Meta-analytic insights into evolutionary ecology: an introduction and synthesis.” *Evolutionary Ecology* 26 (5): 1085–99. doi:10.1007/s10682-012-9593-z.

Noble, William Stafford. 2009. “A Quick Guide to Organizing Computational Biology Projects.” *PLoS Computational Biology* 5 (7): e1000424. doi:10.1371/journal.pcbi.1000424.g001.

Open Science Collaboration. 2015. “Estimating the reproducibility of psychological science.” *Science (New York, N.Y.)* 349 (6251): aac4716–aac4716. doi:10.1126/science.aac4716.

Parker, Timothy H, Wolfgang Forstmeier, Julia Koricheva, Fiona Fidler, Jarrod D Hadfield, Yung En Chee, Clint D Kelly, Jessica Gurevitch, and Shinichi Nakagawa. 2016. “Transparency in Ecology and Evolution: Real Problems, Real Solutions.” *Trends in Ecology & Evolution* 31 (9). Elsevier: 711–19. doi:10.1016/j.tree.2016.07.002.

Patil, P, R D Peng, and J Leek. 2016. “A statistical definition for reproducibility and replicability.” *bioRxiv*. <http://www.biorxiv.org/content/early/2016/07/29/066803.abstract>.

Peng, R D. 2009. “Reproducible research and biostatistics.” *Biostatistics* 10 (3): 405–8.

Ram, Karthik. 2013. “Git can facilitate greater reproducibility and increased transparency in science.” *Source Code for Biology and Medicine* 8 (1). BioMed Central: 7. doi:10.1186/1751-0473-8-7.

Regan, H M, Y Ben-Haim, B Langford, and et al. 2005. “Robust decision-making under severe uncertainty for conservation management.” *Ecological Applications* 15 (4). Eco Soc America: 1471–7. <http://www.esajournals.org/doi/abs/10.1890/03-5419>.

Schreiber, E Sabine G, E Sabine G Schreiber, Andrew R Bearlin, Andrew R Bearlin, Simon J Nicol, Simon J Nicol, Charles R Todd, and Charles R Todd. 2004. “Adaptive management: a synthesis of current

- understanding and effective application.” *Ecological Management & Restoration* 5 (3). Wiley Online Library: 177–82. <http://onlinelibrary.wiley.com/doi/10.1111/j.1442-8903.2004.00206.x/full>.
- Stewart, G. 2010. “Meta-analysis in applied ecology.” *Biology Letters* 6 (1): 78–81. doi:10.1098/rsbl.2009.0546.
- Sutherland, William J, and Claire F R Wordley. 2017. “Evidence complacency hampers conservation.” *Nature Ecology & Evolution* 1 (9). Springer US: 1–2. doi:10.1038/s41559-017-0244-1.
- White, Ethan, Elita Baldrige, Zachary Brym, Kenneth Locey, Daniel McGlinn, and Sarah Supp. 2013. “Nine simple ways to make it easier to (re)use your data.” *Ideas in Ecology and Evolution* 6 (2): 1–10. doi:10.4033/iee.2013.6b.6.f.
- Whitlock, MC. 2011. “Data archiving in ecology and evolution: best practices.” *Elsevier*. doi:10.1016/j.tree.2010.11.006.
- Williams, Mary, Jacqueline Bagwell, and Meredith Nahm Zozus. 2017. “Data management plans, the missing perspective.” *Journal of Biomedical Informatics* 71 (July): 130–42. doi:10.1016/j.jbi.2017.05.004.
- Wilson, Greg, D A Aruliah, C Titus Brown, Neil P Chue Hong, Matt Davis, Richard T Guy, Steven H D Haddock, et al. 2014. “Best Practices for Scientific Computing.” *PLoS Biology* 12 (1): e1001745. doi:10.1371/journal.pbio.1001745.