Manuscript Title:

"But I can't preregister my research": Improving the reproducibility and transparency of ecology and conservation with adaptive preregistration for model-based research

Authors and Affiliations

Elliot Gould[1•], Christopher S. Jones[1,2], Jian D.L. Yen[2], Hannah S. Fraser[1], Henry Wootton[2], Megan K. Good[1], David D.H. Duncan[1], Cindy E. Hauser[2], Bonnie C. Wintle[1], Libby Rumpff[1]

• Send correspondence to: Elliot Gould, elliot.gould@unimelb.edu.au

1 School of Agriculture, Food and Ecosystem Sciences, University of Melbourne. The University of Melbourne, Grattan Street, Parkville, Victoria, 3010.

2 Arthur Rylah Institute for Environmental Research, Department of Energy, Environment and Climate Action. 123 Brown Street, Heidelberg, Victoria, 3084.

Acknowledgements

Thank you to Lyndsey Vivian for assisting in data validation and cleaning in the VEFMAP case study. We would also like to thank attendees for participation in our collaborative workshop.

Author contributions (CRediT roles)

Conceptualization: Elliot Gould, Christopher S. Jones, Jian D. L. Yen, Hannah Fraser, and Libby Rumpff.
Data curation: Elliot Gould, Lyndsey Vivian, Christopher S. Jones, Jian D. L. Yen, and Henry Wootton.
Formal analysis: Elliot Gould, Jian D. L. Yen, and Henry Wootton.
Funding acquisition: Elliot Gould and Christopher S. Jones.
Investigation: Elliot Gould, Christopher S. Jones, Jian D. L. Yen, Hannah Fraser, Megan K. Good, Cindy E. Hauser, Henry Wootton, David H. Duncan, and Libby Rumpff.
Methodology: Elliot Gould, Christopher S. Jones, Jian D. L. Yen, Hannah Fraser, Megan K. Good, Henry Wootton, David H. Duncan, and Libby Rumpff.
Project administration: Elliot Gould, Christopher S. Jones, and Jian D. L. Yen.
Resources: Elliot Gould, Christopher S. Jones, Jian D. L. Yen, and Henry Wootton.
Software: Elliot Gould, Christopher S. Jones, Jian D. L. Yen, and Henry Wootton.
Supervision: Bonnie C. Wintle, Hannah Fraser, and Libby Rumpff.
Validation: Elliot Gould, Christopher S. Jones, and Jian D. L. Yen.
Visualization: Elliot Gould, Christopher S. Jones, Jian D. L. Yen, and Bonnie C. Wintle.
Writing - original draft: Elliot Gould.
Writing - review & editing: Elliot Gould, Christopher S. Jones, Jian D. L. Yen, Bonnie C. Wintle, Hannah Fraser, Megan K. Good, Cindy E. Hauser, David H. Duncan, and Libby Rumpff.

Data availability statement

To be provided following peer-review. See *Data for Peer-review statement* under manuscript abstract.

"But I can't preregister my research": Improving the reproducibility and transparency of ecology and conservation with adaptive preregistration for model-based research

Abstract

1. Preregistration is an open-science practice which aims to improve research transparency and mitigate questionable research practices, like cherry-picking results. It helps protect against cognitive biases, like hindsight bias, that can influence how study outcomes are interpreted. There has been little uptake of preregistration in ecology and conservation, arguably because existing pre-registration templates focus on null-hypothesis significance testing whereas ecology and conservation often rely on different types of statistical modelling.
2. We argue that preregistration in model-based research in ecology and conservation is both possible and beneficial, using templates adapted for domain-specific methodologies. We applied a user-centred design approach to translate the concept of preregistration into model-based research practice for ecology and conservation.
3. To better align the internal logic of preregistration with the iterative and non-linear process of ecological modelling, we propose, test and evaluate a methodology for 'adaptive preregistration', using a case study of modelling managed water releases ("environmental flows modelling") in regulated rivers for maintaining riparian vegetation condition in Victoria, Australia.
4. This research provides a template and methodology for implementing adaptive preregistration of ecological models. Although we focus on ecology and conservation in this paper, the concept of adaptive preregistration, and the templates developed here,

could be applied to model-based research in other scientific disciplines within science more broadly. Modelers in ecology and conservation need no longer cry “but I can’t preregister my research.”

Keywords: adaptive preregistration, ecological modelling, environmental flows modelling, good modelling practice, metascience, preregistration, open science, transparency

Data/Code for peer review statement: Data and code hosted on GitHub have been anonymised for peer review. R code for the case study modelling is available in a GitHub repository https://anonymous.4open.science/r/VEFMAP_VEG_Stage6-7B5F/.

Collaborative workshop materials and analysis (Appendix S1 and S2) are also hosted on the OSF, accessed with the anonymised view-only links: https://osf.io/tz5da/?view_only=992aca57db814e9484a603f8e09b349b and https://osf.io/5w4ms/files/osfstorage?view_only=b60796fed71f4cc8abe9e5feac5e9465.

The adaptive preregistration user guide is accessible as a website (URL removed for double-blind peer review, see Appendix S3 for anonymised PDF version guide), with the website code hosted in a GitHub repository (https://anonymous.4open.science/status/EcoConsPreReg-4C05) and Zenodo archive (link suppressed for double-blind peer review, coauthor-1 et al. 2024). Cross-references to the pdf version of the guide (Appendix S3) replace URLs throughout the manuscript.

The preregistration template is available for installation and use as a quarto template extension in either .pdf (Appendix S4), .docx (Appendix S5) or .html format, and is hosted on GitHub at https://anonymous.4open.science/r/EcoConsModPreReg-6FF5 (coauthor-1 2025).

We have anonymised coauthor names / initials throughout the manuscript text and in supporting information. Some material was unable to be anonymised, e.g. the URL to the case study preregistration GitHub commit history, and the URL to the case study GitHub repository releases. These links remain in the manuscript, but the repository is set to private. The repository that the URLs reference is also provided in an anonymised format, described above.

1. Introduction

Over the past decade, a 'replication crisis' in some scientific fields has exposed problems with the reliability of research findings (Camerer et al., 2018; Neoh et al., 2023). One such problem is the prevalence of Questionable Research Practices (QRPs), including p-hacking, Hypothesising After Results are Known (HARKing), and cherry-picking (Wicherts et al., 2016). QRPs increase the probability of false positive results, artificially inflate effect size estimates, and ultimately reduce the replicability of published findings, while encouraging overconfidence in the precision of results and exacerbating the risk of accepting and propagating false facts (Hoffmann et al., 2021). Self-reports from researchers show that QRPs are surprisingly common across disciplines (Agnoli et al., 2017; Makel et al., 2023, Liu et al., 2020), including in ecology (Fraser et al., 2018; Kimmel et al., 2023). Although self-report surveys of QRPs are heavily focused on statistical significance testing research employing p-values, some broader practices, like selective reporting/cherry-picking, occur and operate analogously in other inferential frameworks.

QRPs are precipitated when researchers opportunistically or systematically misuse 'researcher degrees of freedom' without disclosure (Bakker et al., 2018; Wicherts et al., 2016). Researcher degrees of freedom refer to the many decisions and alternative analytic choices researchers make throughout the research process (Wicherts et al., 2016). Analytic decisions are typically subjective but methodologically defensible, however, choices and omissions taken at each 'fork' may collectively and significantly influence the final 'analysis strategy' (*sensu* Hoffmann et al., 2021), or 'modelling path' (*sensu* Hämäläinen & Lahtinen, 2016). The set of all (plausible or reasonable) possible decisions that a researcher could make over the entire analysis from beginning to end has been termed collectively as the 'garden of forking paths' (Gelman &

Loken, 2013). ‘Many-analyst studies,’ where multiple researchers independently investigate the same research question by analysing the same dataset have now demonstrated the consequences of the ‘garden of forking paths’, whereby alternative analytical decisions can lead researchers to make vastly different interpretations of results, with conflicting conclusions (Gould et al., 2025; Silberzahn et al., 2018).

Good intentions and simply being aware of the potential for researcher degrees of freedom to lead to QRPs cannot sufficiently mitigate their risk because of the unconscious nature of their origin (Zvereva & Kozlov, 2021). Preregistration is an open-science practice that aims to distinguish between planned and ad hoc analyses (Parker et al., 2019) thereby restricting opportunities for trialling analytical alternatives (‘researcher degrees of freedom’) and making any deviations from the preregistered plan transparent (Wicherts et al., 2016). Preregistration requires researchers to register their methods and analysis plan in a secure and publicly accessible platform prior to collecting and/or analysing data, which cannot be altered after submission (Parker et al., 2019). Preliminary empirical evidence shows that studies with preregistered analysis plans are more transparently reported and are less likely to contain positive or significant results (Brodeur et al., 2024). This finding in part reflects the researcher commitment to honest, complete and transparent reporting, especially when the preregistration template includes detailed instructions and requirements (Bakker et al., 2018). It also reflects a correction for publication bias, where negative or non-significant results tend not to be published, so would otherwise remain in the 'file drawer'. ‘Registered Reports’ (Koivisto & Mäntylä, 2024), a form of preregistration, contain a direct mechanism for addressing publication bias, whereby the methods and analysis plan are peer-reviewed in advance, and the decision to publish is made *before* the results are known.

One potential objection to preregistration for modelling is that model development is an intrinsically exploratory process (MacEachern & Van Zandt, 2019), while preregistration is for confirmatory research (i.e. hypothesis or theory testing, Prosperi et al., 2019). These conceptions of modelling and preregistration are not entirely wrong, but they are oversimplified. In this iterative and non-linear process of modelling, initial analyses often inform future analyses of the same data whereby the modeler's understanding of the problem and the model itself are incrementally adjusted in light of intermediate results from previous decision points (Dwork et al., 2015; Hämäläinen & Lahtinen, 2016). Modelers are typically uncertain about which model is most appropriate for the data (Popovic et al., 2024) and are reluctant to register a modelling strategy that may turn out to be incompatible with the data (Roettger, 2019). For example, modelers may perform exploratory analyses to validate distributional assumptions (Campbell, 2021), which may result in subsequent changes to the model specification. Also, in response to data, modellers may learn a new approach that they were previously unaware of. We argue that data-dependent decisions like these should not be interpreted as violating a key tenet of preregistration, that is, failing to define the entire analysis protocol ahead of time (Dwork et al., 2015; MacEachern & Van Zandt, 2019). We propose a methodology for implementing an expanded view of preregistration called *Adaptive Preregistration* (*sensu* Srivastava, 2018), which captures decision points and reasoning about modelling choices, but which embraces the data-dependency practices that constitute the modelling development process. This goes beyond the work of others who offer discussion and guidance about when and how to deviate from a preregistration (Lakens, 2024; Willroth & Atherton, 2024), instead providing an approach to preregistration embodying the principles of *registered flexibility* (Roettger, 2019; Srivastava, 2018) and iterative or *interim preregistrations* (Hofman et al., 2023; Ioannidis, 2022).

In this study we aim to translate preregistration for application to model-based research, specifically aimed at ecology, conservation and related disciplines. Here, we first identify a generalised ‘modelling workflow’ that captures critical analysis decisions in the model development cycle and use it to build an Adaptive Preregistration template that can work for ecological modelling. We then test and evaluate our template using a case study of modelling managed water releases in regulated rivers (‘environmental flows’) to improve riparian vegetation condition in Victoria, Australia. This case study demonstrates a novel approach to preregistration using a template designed for model development and evaluation.

2. Materials and Methods

We modified preregistration for use in model-based research contexts following a user-centered design framework (Pavelin et al., 2012), where users were engaged at all stages to produce a preregistration template and protocol that is fit for purpose (Figure 1).

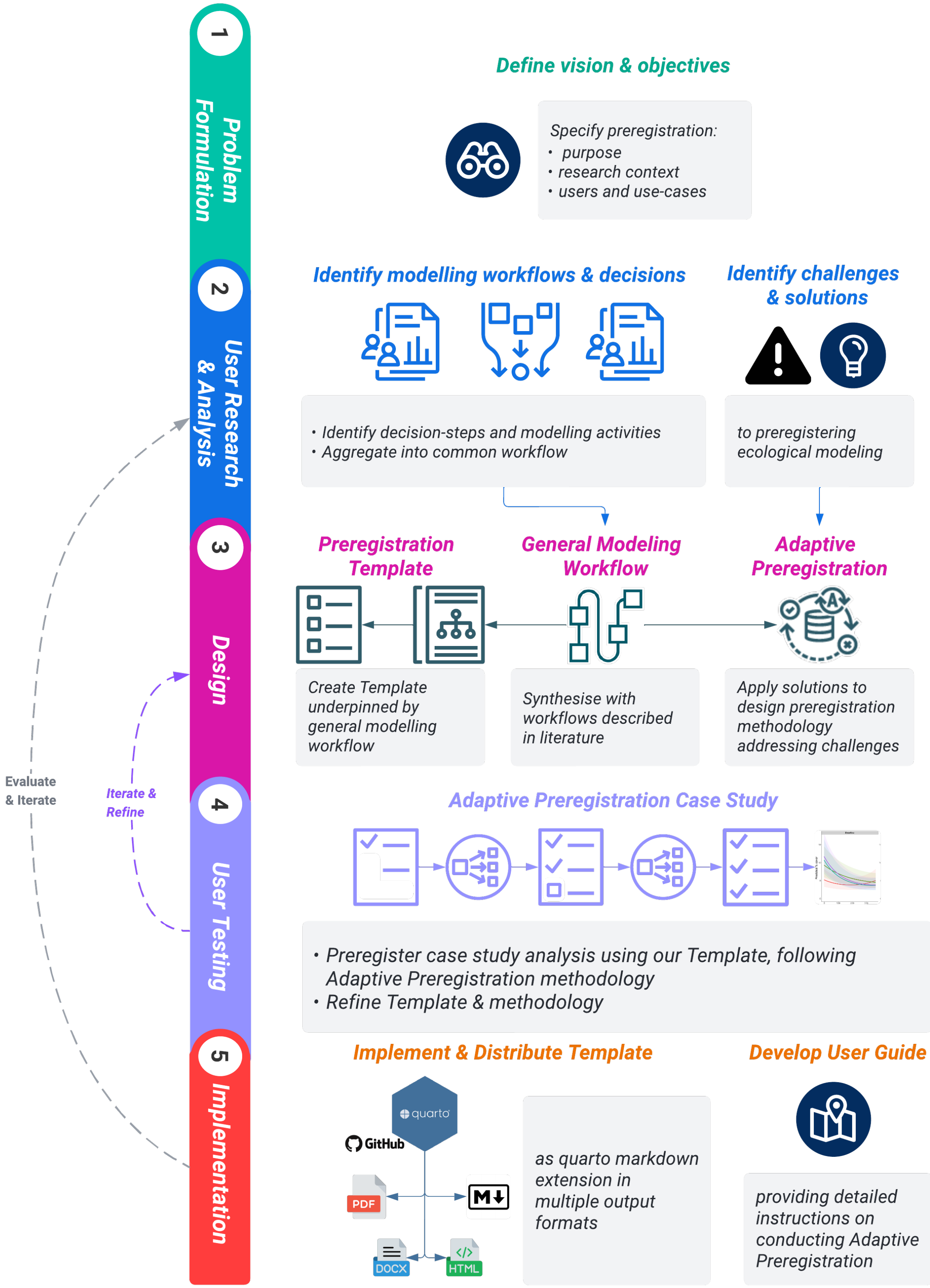

1
Problem Formulation
2
User Research & Analysis
3
Design
4
User Testing
5
Implementation
Evaluate & Iterate
Iterate & Refine
Define vision & objectives
Specify preregistration:
• purpose
• research context
• users and use-cases
Identify modelling workflows & decisions
• Identify decision-steps and modelling activities
• Aggregate into common workflow
Identify challenges & solutions
to preregistering ecological modeling
Preregistration Template
Create Template underpinned by general modelling workflow
General Modeling Workflow
Synthesise with workflows described in literature
Adaptive Preregistration
Apply solutions to design preregistration methodology addressing challenges
Adaptive Preregistration Case Study
• Preregister case study analysis using our Template, following Adaptive Preregistration methodology
• Refine Template & methodology
Implement & Distribute Template
quarto
GitHub
PDF
DOCX
HTML
as quarto markdown extension in multiple output formats
Develop User Guide
providing detailed instructions on conducting Adaptive Preregistration

Figure 1 Process for translating preregistration to ecological modelling research. We followed the user-centred design steps proposed by Pavelin et al. (2012), except for the 'Evaluate and Iterate' phase of the process, which we hope will be taken up by the community when trialling the Adaptive Preregistration methodology and Template developed in this study. Following a collaborative workshop (Step 2), we specified a general modelling workflow that informed the structure and content of a Preregistration Template (Step 3) and proposed a methodology we call 'Adaptive Preregistration' (Step 3) addressing challenges to preregistering model-based research identified in Step 2. We tested and refined the Template and Methodology using a real-world modelling case-study (Step 4). The Template is provided in multiple formats, including as a quarto markdown extension, and instructions for conducting Adaptive Preregistration are detailed in a User Guide (Step 5).

2.1 Problem Formulation: Defining vision & objectives for a preregistration template for model-based research

We first specified the research context, purpose, target users and use-cases, (per Pavelin et al., 2012; Pu et al., 2019, Figure 1, step 1). Our context is model-based research, defined broadly as any research that uses quantitative modelling to answer its research question (Brudvig, 2017). The purposes of the preregistration template and methodology are to:

1. **Delimit 'researcher degrees of freedom'** (Pu et al., 2019), to mitigate the risk of QRPs when conducting model-based research in applied ecology and conservation.
2. **Increase 'research transparency' of ecological modelling,** including both:
    1. *production transparency* — including research artefacts like open-access data and materials, or data collection procedures (Lupia & Elman, 2014),

2. *analytic transparency* — a complete account of how analytic conclusions are drawn from the data (Lupia & Elman, 2014) where model choices, steps in the modelling process, assumptions and expectations about model outputs are clearly articulated (Bodner et al., 2020).

We specified two use-cases for our Adaptive Preregistration template:

1. study authors or modelers seeking to preregister their modelling study,
2. editors, reviewers, and/or readers of the completed preregistration, who are tasked with 'preregistration checking' or verifying that analyses and findings reported in the final manuscript are consistent with the preregistered analysis plan (Pu et al., 2019).

Existing preregistration formats are designed primarily for study authors; they are static text-based documents that do not easily facilitate preregistration checking. We designed our template expand its utility to editors, reviewers and readers.

2.2 Designing the General Modelling Workflow

We conducted a literature review (Appendix S1) to identify a modelling workflow suitable for application to a range of different modelling goals, research contributions, model types and problem contexts; where modelling goals may include exploration, inference and both explanatory and anticipatory predictions (Mouquet et al., 2015; Tredennick et al., 2021), research contributions may include development of new methods or application/extension of existing methods and model transfers (Yates et al., 2018); model types may range from phenomenological to mechanistic (Connolly et al., 2017) and research contexts may include basic and applied modelling exercises across ecology and conservation (Brudvig, 2017). The final output was a workflow that reflected a balance between idealised and actual modelling in practice, aligned with the prescriptive guidelines of 'good modelling practice' (Augusiak et al., 2014).

2.3 Designing the Preregistration Template

We organised and facilitated a workshop with 10 ecologists and ecological modellers (Appendix S2). A key task was to identify common modelling workflows and critical decision-steps in model development (Figure 1, step 2). 'Critical' decision-steps are points in the modelling process where researchers, or modelers, make analytical decisions that could change the output of the analysis. In this activity, workshop participants first individually reflected on their own scientific process for a recent or memorable research project, listing the steps undertaken throughout the process. Next, participants collaboratively mapped the decision steps in their own personal modelling workflows onto a modelling workflow template (https://osf.io/fgd23/?view_only=d3379ed454d84d29b6ef999c0ecc1d83) prepared earlier, informed by our general workflow (Figure 1). We then collaboratively reviewed and refined the modelling workflow into a standard set of modelling phases and commonly implemented decision steps that formed the basis of the preregistration template (Figure 1, step 3).

2.4 Developing an Adaptive Preregistration Methodology

We developed a methodology for application of preregistration to model-based research, or 'adaptive preregistration' (Figure 1, step 3), comprising two key components: registered flexibility and interim preregistrations. We summarise the methodology, and direct readers to our user-guide for more detailed guidance on implementing adaptive preregistration in practice (coauthor-1 et al., 2024, Appendix S3).

Registered Flexibility

Registered flexibility includes preregistration of “plans to deploy flexible strategies” (Srivastava, 2018) whereby the preregistration author specifies flexible heuristics containing alternate analysis or modelling strategies whose execution depends on the outcome of previous decision-points or analyses. For example, when faced with methodological or analytic uncertainty that cannot be resolved without observing parts of the data, conducting preliminary analyses, model checking or other data-dependent decisions, a modeller can preregister a decision-tree that consists of predefined rules about when a particular modelling strategy or decision should be implemented (Baldwin et al., 2022). There are three requirements for adequately registering flexible analyses:

1. Stating what quantity needs to be known to move forward with the modelling and analysis and why.
2. Describing the analysis that will be used to generate this quantity, and which parts of the data will be used.
3. Explaining how the results will be interpreted, listing each alternative decision under consideration, and the analysis result that will trigger each decision respectively.

Interim Preregistrations

The modeller follows an iterative process of preregistration in parallel with modelling, consisting of interim preregistrations that mark key phases of modelling and analysis as different parts of the data are observed (Srivastava, 2018). As the modeller proceeds through the model development process they shift from ideation and preregistration to execution of the preregistered analysis plan, and back again, generating interim preregistrations at different points within the model development process, depending on observed outcomes of the flexible heuristics described in the interim preregistration (Figure 2).The realised modelling path executed from the garden of forking modelling paths will depend on the observed outcomes of the registered flexible analyses at each interim preregistration.

Transparent Documentation of Adaptive Preregistration

Ideally, to transparently document the adaptive preregistration process and facilitate preregistration checking of flexible modelling and analysis strategies (one intended use-case for our methodology and template, Section 2.1) the results of any interim and final analyses must be explicitly linked back to the preregistered analysis strategy. Researchers can take simple steps to achieve this, for example, by retaining separate versions of initial, interim and final preregistration documents; retaining the implementation and results of registered flexible analyses alongside the preregistration documents while recording their filenames in the relevant preregistration document; and utilising existing platforms with basic version control features, such as the Open Science Framework (OSF, https://www.cos.io/products/osf).

Optional: Preregistration With git & GitHub

The use of version-control for transparently documenting code-based analyses is considered best-practice, however, is not yet widespread in ecology and related disciplines (Braga et al., 2023). We propose using git and GitHub as the vehicle for transparent documentation of adaptive preregistration of a modelling study for those researchers already comfortable with these tools (see user-guide for details, coauthor-1 et al., 2024, Appendix S3). The procedure leverages GitHub's tag and release feature (https://help.github.com/en/github/administering-a-repository/releasing-projects-on-github) in conjunction with Semantic Versioning (https:/semver.org/) to track, document and collaborate on changes to the preregistration document and the analysis. Modelling code and results are stored and versioned alongside the preregistration document in the project repository, while *GitHub issues* (see Braga et al., 2023 for explanation and definition) are used to track discrete analysis tasks described in the preregistration. *GitHub issues* are particularly useful for registered flexible analyses, facilitating documentation of any interpretation of the results of registered flexible analyses and their influence on the rationale for the subsequent preregistration; the outcome of the registered flexible analysis is recorded within that issue, hyper-linked to the registered heuristic in the preregistration document, alongside interpretation of those results and documented decisions about which decision alternative is to be selected based on those results.

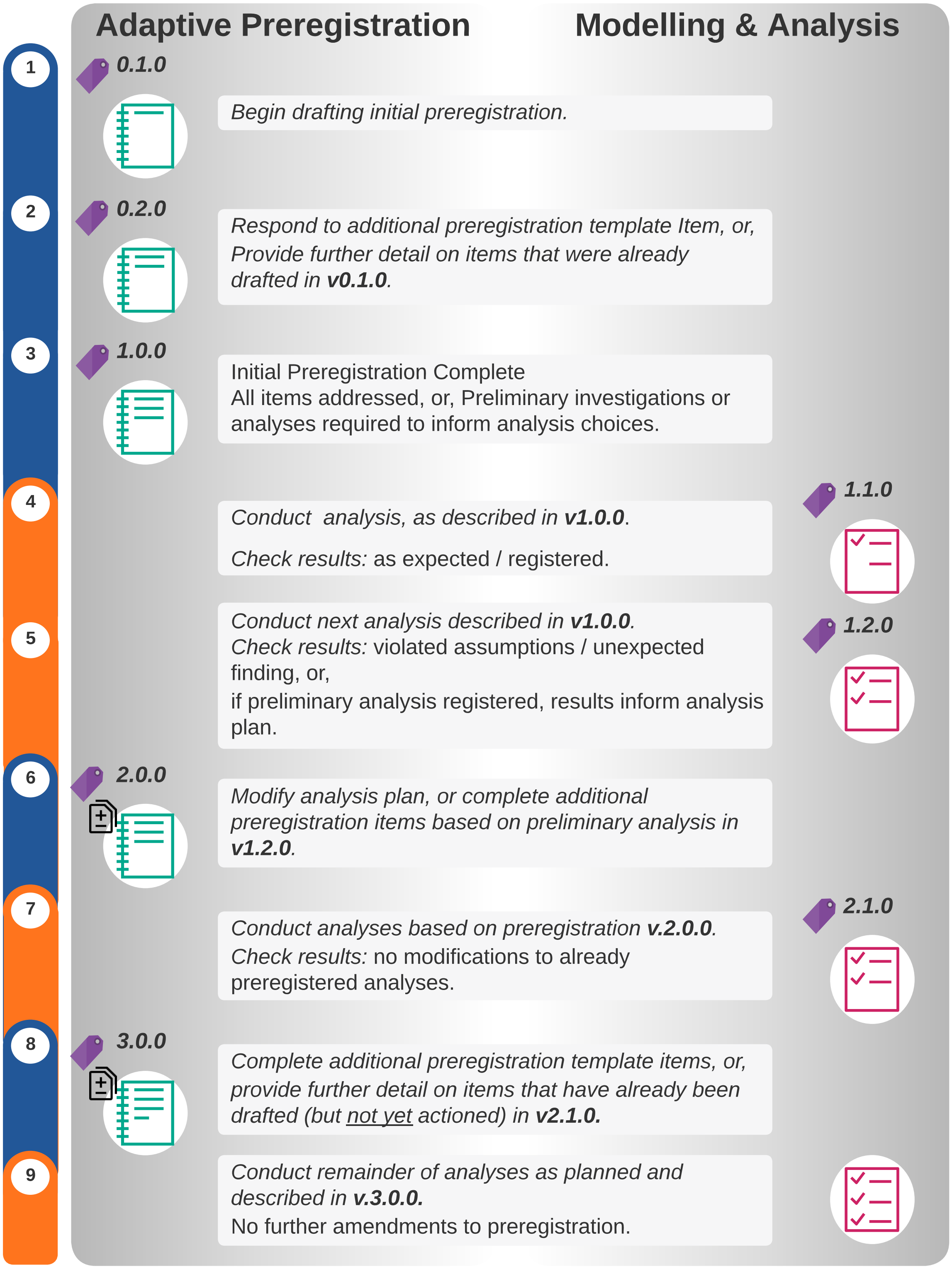

Adaptive Preregistration
Modelling & Analysis
1
0.1.0
Begin drafting initial preregistration.
2
0.2.0
Respond to additional preregistration template Item, or, Provide further detail on items that were already drafted in v0.1.0.
3
1.0.0
Initial Preregistration Complete
All items addressed, or, Preliminary investigations or analyses required to inform analysis choices.
4
1.1.0
Conduct analysis, as described in v1.0.0.
Check results: as expected / registered.
5
1.2.0
Conduct next analysis described in v1.0.0.
Check results: violated assumptions / unexpected finding, or,
if preliminary analysis registered, results inform analysis plan.
6
2.0.0
Modify analysis plan, or complete additional preregistration items based on preliminary analysis in v1.2.0.
7
2.1.0
Conduct analyses based on preregistration v.2.0.0.
Check results: no modifications to already preregistered analyses.
8
3.0.0
Complete additional preregistration template items, or, provide further detail on items that have already been drafted (but not yet actioned) in v2.1.0.
9
Conduct remainder of analyses as planned and described in v.3.0.0.
No further amendments to preregistration.

Figure 2 Implementing Adaptive Preregistration. The preregistration (blue steps) and modelling and analysis processes (orange steps) operate in parallel. Purple tags represent git tags, corresponding to incremental versions of the preregistration. Any changes to the analysis plan receive a GitHub release and major version number increment. Step 3 corresponds to the initial preregistration document, Step 6 corresponds to an interim preregistration, and Step 8 corresponds to the final preregistration. The execution of the Final Preregistration occurs in Step 9.

2.5 Evaluating and Refining Adaptive Preregistration with a Case Study

We preregistered a real-world model-based research problem as a case study for evaluating and testing the Preregistration Template and proposed Adaptive Preregistration methodology (Figure 1, step 4). The case study analysed the effectiveness of environmental flows on maintaining riparian vegetation condition in Victoria, Australia (Jones et al., 2025, *in review*), forming a component of the Victorian Environmental Flows Monitoring and Assessment Program (VEFMAP), a large-scale, long-term monitoring program delivered by the Arthur Rylah Institute for Environmental Research. The VEFMAP research team consisted of; a project lead (coauthor-2), who was responsible for establishing the conceptual framework and research questions that inform the analysis, and a modeller (coauthor-4) who undertook much of the coding and exploratory analysis under direction of a lead modeller (coauthor-3). Modelling experience among the team ranged from moderate to advanced, and the preregistration was collaboratively completed by the case study research team and the study lead (coauthor-1). In addition to following the protocol for adaptive preregistration using git and GitHub, we also utilised the version-control and collaborative project management features of git and GitHub (Braga et al., 2023) to live-develop the preregistration template and capture the case study researchers’ feedback about the template and adaptive preregistration methodology. For example, if it emerged in the process of completing the case study preregistration that we had missed an important step in the modelling process or found that the order and structure of the template should change, the suggested change and its justification was recorded in GitHub discussions. We also conducted follow-up semi-structured interviews with the case study researchers to capture detailed reflections.

3. Results

3.1 The Preregistration Template

The final Preregistration Template refined by application of the draft Preregistration Template developed from the user-research workshops to the case study, is hosted at https://anonymous.4open.science/r/EcoConsModPreReg-6FF5. We provide the template as a quarto markdown extension template (coauthor-1, 2025) which can be installed and rendered to three different formats, html, pdf, and docx (Appendices S4 and S5. Intermediate versions of the template developed through application of the initial template to the case study are available in the case study preregistration document's GitHub commit history (link suppressed for peer-review) or in case study GitHub repository releases (link suppressed for peer-review).

3.2 Findings from the Case Study

The application of our preregistration template and Adaptive Preregistration methodology to the environmental flows modelling case study occurred across three distinct stages (Table 1, Figure 3), with two interim preregistrations (Version 1 and Version 2) preceding the finalised preregistration (Version 3, Appendix S6), each corresponding to distinct phases in both the Preregistration and Modelling & Analysis processes. While preparing Version 1, the modellers decided that they needed to conduct some analysis to specify a set of appropriate candidate models. They preregistered a pilot analysis (Version 2) which used a subset of the full dataset to preserve some degree of data-dependent decision-making while permitting a degree of exploratory analysis. Results from the pilot analysis informed subsequent preregistration of the Main Analysis on the full dataset (Version 3, Appendix S6). Several classes of data-dependent decisions (Liu et al., 2020) emerged during our case study, which were managed using a combination of registered flexibility and interim preregistrations:

- *Data-flow dependencies*: where the output of one decision is the input to another. For example, the models parameterised on the full dataset in the case study's Main Analysis were subject to a suite of model checks as outlined in the decision-trees in Preregistration Version 3, with the results of model checking informing *a priori* specified alternative model structures and model functional forms.
- *Information dependency*: where one decision informs another. For example, outcomes from the exploratory pilot analysis (Interim Preregistration Version 2, Table 1) informed the model structures specified in Preregistration Version 3 (Table 1), as well as the decision heuristics for triggering their fitting and acceptance.
- *Procedural dependency*: where downstream decisions would not exist if some alternative upstream decision was made instead. For example, if preliminary analysis of the data had not revealed over-dispersion or zero-inflation in the data, then the case study modellers would not have fitted zero-inflated Poisson models, nor would they have trialled accounting for over-dispersion or refitting using negative binomial models (Figure 4).

3.2.1 Interim Preregistration (Versions 1 and 2)

Two initial candidate ‘maximal’ models were preregistered in Version 1 (Figure 3), one for each study response variable, however, modellers were unsure if these models would be supported by the dataset due to low sample sizes relative to the complexity and spatio-temporal patchiness of the data. Preregistration Version 2 described an exploratory pilot analysis that aimed to resolve critical uncertainties about how to operationalise the models, given data constraints, and how to specify a candidate set of models that could feasibly be fitted to the full dataset (Table 1). This involved partitioning the dataset into a ‘Pilot Analysis’ subset for making subsequent data-informed modelling decisions, which were then preregistered in a subsequent preregistration of the ‘Main Analysis’ (Version 3, Table 1 and Figure 3). The pilot analysis dataset consisted of observations from a single site with good temporal data coverage, while the main analysis dataset comprised the full dataset, including the pilot dataset. Preliminary analysis informed how to operationalise the models given the underlying structure of the dataset. For example, simple tests of data distributions within hierarchies were required to check for data spread and the presence of zero-inflation, while checks for collinearity among candidate predictor variables were needed to prevent confounding and overfitting, and to ensure relevant interactions were captured in the models. The same procedures for evaluating model fit and model checking were preregistered for both the pilot and main analysis (Figure 4B), however the purpose of model checking in the Pilot Analysis was to identify appropriate model structures, whereas the purpose for the Main Analysis was to ensure that the fitted models were supported by the data.

3.2.2 Final Preregistration (Version 3): Main Analysis

In Version 3 of the Preregistration, we derived and preregistered:

1. A preferred candidate model or 'full model' for each response variable (vegetation richness and vegetation cover), to be fitted to the full dataset (Initial Preregistration, Figure 3 and Appendix S6, p.13).
2. A candidate set of simplified models that separately captured different flow components for each response variable, with two versions of the flow events model for vegetation cover (Flow Events Models A & B, Figure 3 and Appendix S6, pp.20-21).
3. Decision-trees for triggering fitting and acceptance of the candidate simplified models should the preferred full model specifications fail to converge and/or provide appropriate parameter estimates (Figure 3).
4. A second flexible strategy, for determining model functional form and to account for potential over-dispersion (Figure 3). Preregistering this flexibility was motivated by the modellers' concerns that the actual data distributions in the full dataset did not match expectations informed by the pilot analysis. This final decision-point for choosing the model family is subjective and fuzzy, and explicit weightings between the three model performance criteria could not be expressed or preregistered. Instead, modellers explained that the final decision will be guided by the model's overall ability to capture key associations reliably.

Table 1. Key stages of the modelling and Adaptive Preregistration process implemented in the environmental flows modelling case study. Version numbers in parentheses are GitHub tags, marking snapshots of the repository for each Preregistration Version.

Preregistration Version	Preregistration Process	Modelling and Analysis Process
Version 1: (v0.10 – v0.8.1)	Establish problem context, study background and aims, describe data collection and data cleaning process, articulate modelling objectives, and specify candidate models.	Data cleaning and preparation undertaken concurrently by lead and supporting researchers due to limited resourcing and short timelines for project deliverables. Candidate models addressing different components of the research question were preregistered. Data structural properties and potential consequences for model fitting were identified
Version 2: (v0.9.0 – v.0.12.5)	Describe Pilot Analysis on data subset and specify registered flexibility in the form of decision heuristics.	Critical uncertainties in modelling decisions were articulated (e.g. how do we classify best flow regime based on inundation data?). In addition, data partitioning, exploratory data analysis, refined candidate models and model checking procedure preregistered
Version 3: (v0.13.0 – v.0.20.1)	Describe Main Analysis, including specification of decision tree for model selection process based on results of pilot analysis.	Conducted modelling and analysis as outlined in preregistration. Reported deviations and rationale as necessary.

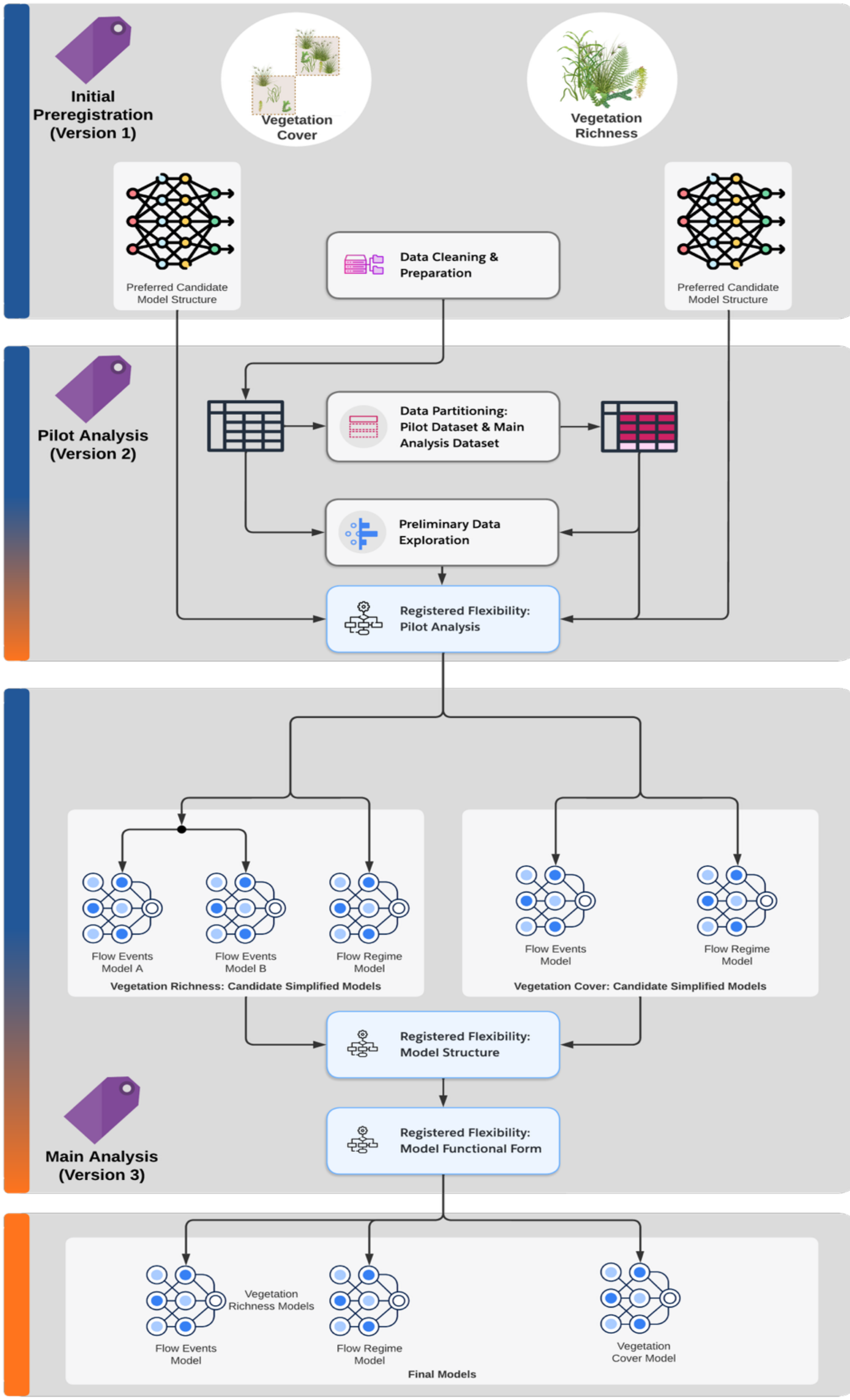

Initial Preregistration (Version 1)
Vegetation Cover
Vegetation Richness
Preferred Candidate Model Structure
Data Cleaning & Preparation
Preferred Candidate Model Structure
Pilot Analysis (Version 2)
Data Partitioning: Pilot Dataset & Main Analysis Dataset
Preliminary Data Exploration
Registered Flexibility: Pilot Analysis
Flow Events Model A
Flow Events Model B
Flow Regime Model
Vegetation Richness: Candidate Simplified Models
Flow Events Model
Flow Regime Model
Vegetation Cover: Candidate Simplified Models
Registered Flexibility: Model Structure
Registered Flexibility: Model Functional Form
Main Analysis (Version 3)
Vegetation Richness Models
Flow Events Model
Flow Regime Model
Vegetation Cover Model
Final Models

Figure 3: The evolution of model development across preregistration iterations in the case study. Each preregistration corresponds to different stages of the model development process, summarised in Table 1. Maximal models for each response variable were specified in the initial preregistration (Version 1). Exploratory and pilot analyses and data partitioning were next preregistered (Version 2), aiming to resolve uncertainties in model specification given data constraints. The pilot analysis identified three candidate simplified models for vegetation richness and two for vegetation cover. These model structures were preregistered in the final preregistration for the Main Analysis (Version 3), in conjunction with registered flexibility for determining model structure and functional form. The two Registered Flexibility steps in the Main Analysis (shown in blue) were informed by preliminary analyses conducted in the Pilot Analysis (Version 2) and are further described in Figure 4. The Main Analysis was preregistered to be conducted on the full dataset. After conducting the preregistered Main Analysis, three final models were identified for model fitting and analysis, one model for vegetation cover, and two vegetation richness models, with one incorporating ‘flow events’ (spring and summer freshes) and the other incorporating ‘flow regime’ (days per year above baseflow and days per year above spring fresh).

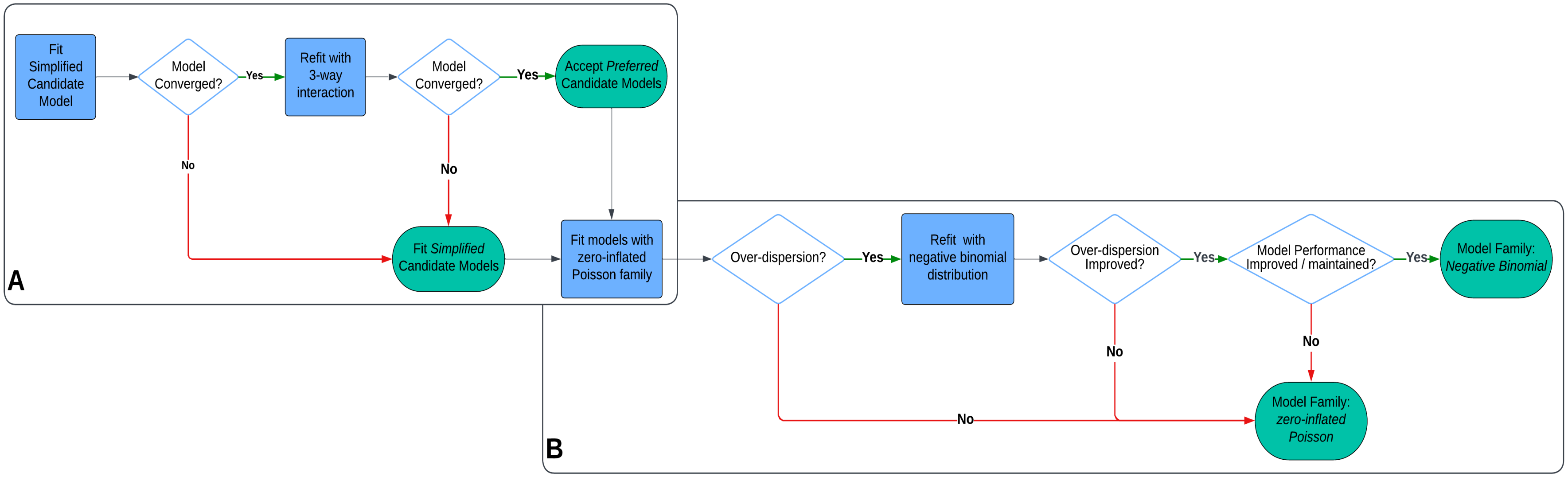


Figure 4: Application of registered flexibility within the case study preregistration, for selecting model structure (A) and selecting an alternative functional form over the default zero-inflated Poisson model (B).

3.3 Adaptive Preregistration in Practice: Benefits and Challenges

At the end of user-testing, case study researchers concluded that preregistration is a valuable tool to facilitate good modelling practice, but also acknowledged several difficulties encountered during Adaptive Preregistration (Table 2).

Table 2: Summary of challenges ⊖, benefits and added value ⊕, as well as tips  for using Adaptive Preregistration for ecological modelling.

Type	Name & Description	Case Study Example	Practitioner Recommendations and Reflections
	**Decide on template resolution.** Time required to make preregistration templates specific, precise, and exhaustive may be off-putting. What is an appropriate resolution?	The preregistration template first used in the case study was unwieldy and overwhelming. Resolution was reduced by restructuring, removing and merging template items, making the task more approachable.	*"Be willing to modify the template if it does not meet the needs of the project. Any attempt to preregister is better than none at all"* –coauthor-3.
	**Decide on template formatting.** The visual design, presentation and formatting affects usability.	Reformatting template with collapsible help and explanatory text boxes, as well as coloured icons to differentiate preregistration items from explanatory text, improved clarity of the initial preregistration and facilitated more efficient completion of the preregistration using the template.	
⊖	**Choose a version-control platform that does not require too much investment in upskilling.** Adaptive Preregistration requires interim versions of the preregistration be deposited in a single location that is ideally version controlled. Some platforms have a steeper learning curve than others, impacting on the efficiency of preregistration.	Using GitHub to track iterations in the preregistration proved difficult to combine with a *GitFlow* approach for simultaneously collaborating on the analysis code and preregistration across branches. The lead modeler (coauthor-3) regularly uses git for version control but was less familiar with the GitHub web GUI. Under broad direction of the lead modeller, a second modeler (coauthor-4) undertook much of the coding and exploratory analysis, but had limited familiarity with version-control systems, and was unfamiliar with collaborative code-base development tools, such as GitHub. Substantial project time was allocated to developing a collaborative GitHub workflow and familiarising both modellers with this approach.	
	**Ensure that modelling decisions and registered flexibility are accurately documented and easily discernible when preregistration checking.** For complex modelling studies, especially where there are multiple models being preregistered and/or where there are multiple models under consideration within a model selection approach, it is difficult to keep track and accurately document analysis decisions, including flexible decision-heuristics.	When checking the preregistration to write the case study report, the lead author (coauthor-1) found it difficult to clearly identify registered decision-triggers for choosing alternative model functional forms over preferred forms, due to information dispersal over multiple sub-sections within the preregistration (4.3.1 Quantitative Model Checking, 3.2.2 Choose Model Family, 3.5 Model Assumptions & Uncertainties, 3.4.2 Estimation Performance Criteria). Moreover, some decisions concerning the model selection process were informed by tacit knowledge of the modelling team and not recorded in the preregistration (Section 3.1.4).	*"I could not figure out why quadratic associations rather than linear associations were fitted at the functional group level when these were not described in the preregistered full model descriptions and pseudo-code. After significant effort, it became apparent that this decision was described across multiple sub-sections"* – coauthor-1. *"In hindsight, we would specify the model selection procedure in both the Pilot Analysis and Main analysis more clearly. The process was more akin to 'make the model more*

			complex again, and make it look more like the full model" - coauthor-3.
⊖	**Adaptive Preregistration may be incompatible with model management workflows in real world applications.** Operational procedures may constrain analytic decision-making too early in the research timeline and requires modellers to be involved earlier and throughout the research. Operational constraints may be easier to overcome in academic rather than government or industry settings.	At the case study institution, similar to many organisations working on applied problems, modellers are typically brought into the research project late in the research programme and are allocated limited time over a discrete period (e.g. 5 days immediately before report-writing and well after project planning) to conduct the modelling with limited knowledge of data provenance, dependency structure and underlying distributions. Following the Adaptive Preregistration procedure – where there were several rounds of iterative planning and preregistration, and implementation and analysis – required modellers to be involved much earlier in the research process than usual. Changes in operational procedure and culture would need to occur at the case study institution to facilitate Adaptive Preregistration becoming commonplace.	
⊖	**Preregistering flexible analyses is difficult and may delay the research process, especially as the analysis increases in complexity.** Formulating alternative decision-pathways and their decision-triggers is challenging under certain conditions: when the underlying dataset is complex, unfamiliar to the modellers, zero-inflated, over-dispersed, and small (low sample size). On the other hand, preregistration of straightforward projects would likely take less time due to increased familiarity with "standard" or "go-to" analyses for simpler data types and structures.	The process of working through and documenting analysis decisions prior to analysing the data delayed completion of the analysis considerably when compared to expectation. The task of mentally forecasting potential analysis scenarios and decision pathways was particularly challenging in this instance, because the dataset was complex and unfamiliar to the modellers. Compared to knowledge and models of processes affecting fish responses to environmental flows, the vegetation models are relatively less developed. The lead modeller has much more familiarity modelling count data on fish and reflected that there would be less analytic unknowns in those instances, which would have made the task of preregistering flexibility much easier.	*"Adaptive Preregistration did delay the research process initially, because we had to carefully consider all the likely scenarios to the end of the process and document those, as opposed to only needing to figure out the first part and then starting. I expected this delay to be about a month; however, the delays were much longer due to the fact that we had never tried preregistration before, we were also feeding back information into the Adaptive Preregistration process, and the model design and structure was very complex and different to things we had done in the past. It was a pretty challenging analysis – far more complex than many ecological studies I have worked on"* – coauthor-2.
⊕⊖	**Adaptive Preregistration is difficult at first and therefore slows the research process** but will become easier and more efficient with familiarity and experience with preregistration.	The preregistration of decision trees and heuristics for triggering the selection of some models over others was the most difficult and time-consuming aspect of Adaptive Preregistration for the case study modellers. However, with practice and experience they expect Adaptive Preregistration will become easier and quicker.	*"The process would definitely be easier and faster the next time, even for a project of the same complexity. Partly this is because of just knowing how it works and what to expect. The first one will always be the most difficult and slow. Going through the Adaptive*

			Preregistration process once or twice will make it much more efficient." – coauthor-2.
⊕	**Planning the analysis at the outset likely saves time in long run**. Using a structured workflow that carefully considers modelling decision points and options in advance reduces scope for time-consuming back-and-forth adjustments later in the process. Having a documented plan to refer back to also saves time on similar, future tasks and report writing.	Typically, the modelling team follows a more trial-and-error based approach to model development, particularly for research and decision-making contexts where conceptual models, variable definitions and model parameters are uncertain (such as for the vegetation components of VEFMAP, compared to fish models). Adaptive Preregistration forced the modellers away from their usual iterative process towards a carefully considered and well-justified a priori construction of the models. The case study modellers thought that the modelling process saved time in the coding of the models and by preventing iterative adjustments in the models and code and saving potential back-and-forth communications between modelling team members when deciding on and implementing and those changes.	*"Undoubtedly using Adaptive Preregistration saved some time at the end of the process, but it is hard to say how much. Shifting the process from a more trial-and-error based approach to a very carefully considered construction of the model a-priori, must surely have saved time in the coding and back-and-forth adjustments stages. The document was very helpful when it came to write the report. We also have a worked example that we are happy with that we can use to leverage off or remind ourselves how we approached a step in the past"* – coauthor-2.
⊕	**Preregistering model structures and predictors in advance strengthens links between methods and the original study aims / concepts.**	Preregistration required careful consideration of the candidate predictors and possible model structures. The case study modellers acknowledged that while the broad approach to model development would likely have followed a similar progression (expanding on a basic *Poisson* model to address zero-inflation and overdispersion, while also reducing a candidate predictor set down to a maximal viable set), in the absence of Adaptive Preregistration they may have proceeded without clear criteria for selecting one model over another (noting that these criteria were still relatively vague in the preregistered study).	*"If these models were developed without preregistration, much of this thinking would have been 'on the fly', which may have resulted in better model fit (under a range of criteria) but perhaps weaker links to the study's aims and conceptual basis"* – coauthor-3.
⊕	**Adaptive Preregistration reduces researcher degrees of freedom** by shifting the nature of the modelling process away from a loosely bounded and iterative, trial and error style strategy towards a more constrained, goal-directed type analysis. Incorporating loosely defined pilot analyses with data subsetting facilitates data	When considering how the case study modelling would have occurred without being preregistered, both the project lead and lead modeler thought that the preregistration process led to similar outcomes to what would be expected under standard practices, but that the modelling process under Adaptive Preregistration may have resulted in a different set of predictors (including entirely different predictor variables) being selected because the model selection process was more constrained compared to usual practice.	*"I do think we probably would have tried more variations on the model structure and variables without the pre-registration. After starting with standard data checks to determine appropriate model structure, I would usually run simple models to see if I could get them to work. I would have then added complexity incrementally adjusting my model as I went and*

	exploration being formally incorporated into the analytic decision-making process.		*troubleshooting problems as they arose. I would have adjusted the definition of the regime variable, in particular, based on ecological expectations and model outputs (if required)"* – coauthor-2.
⊕	**Adaptive Preregistration increases research transparency.** Specifying modelling and analysis plans, in a preregistration, together with justification of choices, ensures decisions are documented and reported in greater detail. Typically, only the final models are reported.	The use of interim preregistrations of the pilot analysis ensured that preliminary analyses of exploratory nature were documented when they otherwise would not have been, the results of which were expressly linked to candidate model structure selection in the preregistered main analysis. In particular, the project lead noted that early stages of the analysis were documented when typically, decisions and analyses at these early stages, such as data cleaning and processing decisions, would not have been reported.	*"We would have conducted an extensive data review in the same way as with the preregistration, but the process would not have been as well documented and that would have made it much more likely that we would have had gaps in the process and therefore more potential for errors"* – coauthor-2.
⊕	**Adaptive Preregistration increases research accountability.** Documenting the process helps demonstrate accountability to funders and decision makers.	The project timeline included several reporting milestones and review processes. At early to middling timepoints, the modellers were able to share the preregistration document to funders to indicate how they intended to conduct the analysis, communicate rationale for particular analysis choices, and demonstrate time investment.	*"I was very glad to have the preregistration on hand to share or refer to for reporting milestones and review processes. The project was also long and complex with several delays, so it was nice to have a clear document outlining the process so that we could go back and check if we forgot an element of our plans"* – coauthor-2.
⊕	**Adaptive preregistration encourages regular communication between modellers and others involved in the research.** Clear communication among team members improves the chances that model outputs align with the project aims.	*"The Adaptive Preregistration process (like static preregistration) requires clear communication between all people involved in collecting and modelling the data. Frequent communication provides more clarity about the purpose and scope of analyses, which may lead to model outputs being more closely aligned with project aims than would otherwise occur"*– coauthor-3.	
⊕	**Preregistration helps clarify good practice and process for less experienced researchers.**	The lead and assistant modellers worked closely together during the case study, with the assistant modeller implementing part of the model development and coding of the models under the supervision of the lead modeller. The iterative nature of Adaptive Preregistration and its interim preregistrations could facilitate less experienced researchers to document their analysis	*"The process helps to avoid future problems, improves transparency and quality of research practices, and is a very useful tool for less experienced researchers wanting to develop their skills. I strongly encourage everyone to*

		choice rationale and validate this with experts or more experienced modellers before proceeding, so they could progress with confidence.	*go through it to understand how it works and how it could be of benefit to themselves and/or their colleagues"* – coauthor-2
	**Develop the preregistration and analysis over a short time period** so that institutional or organisational shifts do not disrupt the research process.	*"The case study presented here was developed over several years, during which time there was staff turnover and changes in project constraints (including budgets). This complicated the process."* – coauthor-3.	
	**Adaptive Preregistration provides a paper trail and archive of the current state of the project,** facilitating swift continuation of progress in case of unexpected loss of key personnel.	*"It would have been extremely important to have the preregistration document on hand if something happened to our biometrician (e.g. sickness or other absence, which is common). If we had their code and the pre-registration document, another biometrician would have been able to very quickly get everything they needed to progress things."* – coauthor-2.	

4. Discussion

The potential benefits of preregistration are clear: it encourages careful and considered analysis, restricts researcher degrees of freedom, and mitigates the Questionable Research Practices that can derail research. Preregistration can improve the reproducibility and reliability of scientific research (Koivisto & Mäntylä, 2024).

Existing preregistration templates are best suited to hypothesis testing research, and less suited to the style of analysis often taken by ecologists: that is, iterative modelling of complex relationships, often switching between exploratory and confirmatory practices (Alspaugh et al. 2019; Connolly et al. 2017; Prosperi et al. 2019). Although iteration makes preregistration less approachable for model-based studies, it need not preclude it entirely. Here, we show how 'Adaptive Preregistration' can facilitate best practice via a structured, principled and flexible approach to analysis.

Despite fears that preregistration may limit creativity and flexibility (Pu et al., 2019) and lessen researchers' engagement with data (MacEachern & Van Zandt, 2019), our case study shows how adaptive preregistration can improve reproducibility and transparency of process, research outputs and analytic decision-making (Liu et al 2020). By separating the planning stage from the analysis, our lead case study researcher found the structured approach helped focus modelling and analysis activities, maintaining clear links between methodological decisions and the study aims and conceptual models (Table 2).

By facilitating data exploration in the modelling through registered flexibility, data-subsetting and interim preregistrations, Adaptive Preregistration trades unbounded explorations for more goal-directed analyses, constraining researcher degrees of freedom and ‘fishing expeditions.’ In the absence of Adaptive Preregistration, the case study modellers felt they would have started with a much larger decision-space, trialling more variations on model structure and predictor variable operationalisations along the way, risking overfitting (Lewis et al., 2023).

Our findings also show that Adaptive Preregistration increases both analytic and model transparency (Bodner et al., 2020; Lupia & Elman, 2014) by ensuring that modelling choices, steps, assumptions, and expectations about outputs are articulated in greater detail than would typically occur (Schmolke et al., 2010). Documenting intermediate models, analyses and results avoids reporting only the final models and exposes the influence of exploratory and data-contingent analyses on modelling choices. In our case study, the use of explicit prompts for choosing between and transparently justifying alternative model functional forms served to illuminate the analytic garden of forking paths, which too often remains opaque in ecological modelling (Schmolke et al. 2010, Fitzpatrick et al. 2024).

Modelling is considered an ‘art form’ as well as a science (Smaldino, 2020). Unpacking ingrained tacit knowledge may take time, especially the first time around, but there is clear value to capturing this expertise for future researchers working on the same or similar problems. It took several attempts for our case study modellers to fully articulate registered flexibility when selecting model structure and functional form, perhaps because these procedures were second-nature in the modeller’s practice, but by directing users to specific modelling tasks, preregistration prompted modellers to explicitly recognise, document and justify tacit modelling choices.

4.1 Difficulties encountered using Adaptive Preregistration in our case study

Our case study researchers reported that the most challenging aspects of Adaptive Preregistration involved anticipating analysis decisions in advance, in addition to changes to model management and timelines. Registering flexibility requires careful thought and time to adequately capture analytic decision-points and describe the process for resolving those decisions. In our case study, while the decision heuristic for choosing model functional form was accurately preregistered, the process for determining model structure was underspecified in the 'Main Analysis' preregistration. This occurred because the preregistration did not specify that the heuristic should only be applied to some models and not others (i.e. Flow event models and not flow regime models, Figure 3).

The case study researchers noted that anticipating potential analysis scenarios and decision pathways was complicated by large uncertainties about the underlying system dynamics in a complex and patchy dataset. The preregistration, while detailed, did not fully capture the modellers' intended and implemented model selection strategy, where after fitting the set of simpler models identified from the pilot study (preregistration version 2, Table 1), the modellers iteratively trial different combinations of model terms to generate models closer to the more complex 'maximal' models identified in the initial preregistration (preregistration version 1, Table 1). The preregistered specification did not explain which combinations of model terms would be trialled, and in what order they would be trialled, rendering aspects of the model selection process unregistered. Project milestones with tight project turn-around times left limited time to iteratively review analysis specification decisions, resulting in some decisions and components of the modelling process being omitted from the preregistration. In an ideal world, following our Adaptive Preregistration user guide more closely may have avoided underspecificity in the case study preregistration, but our case study faced the same resource constraints typical of the real world, for example, costs required to upskill some members of the research team with git and GitHub.

Our second aim for the template and adaptive preregistration methodology was to facilitate ‘preregistration checking,’ — the process of verifying that the analyses reported in a study matched those in the preregistration (Pu et al., 2019). We approached preregistration checking through the adaptive preregistration methodology itself, specifically, through the 'registered flexibility' mechanism and 'interim preregistrations', together with our git and GitHub implementation protocol. These facilitated comparison of the data-dependent modelling pathways realised in the course of the study with the full set of decision alternatives at each decision-point, providing explicit rationale for any data-dependent preregistered analysis. By version-controlling the preregistration document with git and GitHub, the genesis of the preregistration from one version of the next is made explicit, for example, GitHub’s ‘diff view’ illustrates exactly how and where a document has changed between versions (Appendix S3, Figure 1). Also, the use of GitHub 'issues' in our protocol tracks the findings of each analysis and links them to the corresponding preregistration item. By viewing each preregistration item’s GitHub 'issue', the implemented analyses and findings, together with any subsequent analyses, is clear to anyone checking the preregistration. Due to limitations in our own case study application (discussed above), the analysis paths and rationale were only transparent for some preregistration items (e.g. preregistered data cleaning and processing tasks https://anonymous.4open.science/r/VEFMAP_VEG_Stage6-7B5F/issues/23) or only partially transparent, with discussion of some changes to the analysis plan being recorded transparently (e.g. https://anonymous.4open.science/r/VEFMAP_VEG_Stage6-7B5F/issues/23issues/71 - issuecomment-1990420515), but not explicitly linked to changes to the modelling and analysis code itself due to the failure to use GitHub issue tags for changes to the relevant modelling code. Consequently, we only partially achieved our aim of facilitating preregistration checking.

These aspects combined may result in researchers feeling like Adaptive Preregistration is slowing the research process because preregistration front-loads the decision-making and planning in the project timeline, delaying analysis and implementation of the preregistration (Evans et al., 2023), which may be incongruent with existing institutional approaches to the governance, administration and operational support of modelling (i.e. “model management” Arnold et al., 2020, p.2). In hindsight, the case study preregistration process may have been more efficient and better specified if modellers and/or biostatisticians were brought earlier into the project’s planning and decision-making process to allow sufficient time to review, record and better connect analysis decisions to project aims. Due to the data constraints and uncertainty in determining adequate model structures for the data, the case study was a particularly complex analysis with which to apply and test adaptive preregistration. For simpler analyses, adaptive preregistration would likely not be as difficult once the process is broadly understood. We recognise that learning and implementing a new approach is unappealing at first, and although adaptive preregistration may save time later on, it may require a rearrangement of project timelines and resources.

Adaptive preregistration is likely to become easier and more efficient as researchers gain experience and familiarity with preregistration. When implementing Adaptive Preregistration for the first time, we recommend the following:

1. Start with simple descriptions or dot points in preregistration responses and build detail and complexity in analysis specification incrementally, as well as in thorough and ongoing collaborative review.

2. Refer to existing preregistrations to guide researchers in completing their own. Although there are few examples in ecology now, with future uptake and an increasing number of journals offering the registered reports publication pathway, e.g. Conservation Biology, Biological Conservation, Journal of Ecology and Nature Ecology & Evolution, readers should expect an expanding pool of preregistrations to draw on. Existing registration repositories, such as OSF registries, can also be searched.

3. Developing refined templates for specific applications (e.g. prediction versus inference, or specialised complex methods and study designs) may be necessary for adequately capturing analytical decisions. In contrast, with time and community uptake, a set of decision-points common across modelling methods and applications may naturally evolve over time through practice. Our template offers a starting point from which researchers can use, modify and add to as they see fit.

4. For some researchers, using git and GitHub for transparently documenting the adaptive preregistration process may be too challenging. In such cases, or for simpler analyses, separate sequential analysis plans could be preregistered for the

same study with metadata linking the chain of preregistrations without tracking associated analysis files. The OSF currently allows for preregistration updating after submission (but not updating of project files, https://help.osf.io/article/410-registration-files#ViewingPrevious).

5. Any attempt to implement adaptive preregistration is unlikely to work perfectly the first time, there will be mistakes and details that are omitted. Being upfront about this in study reporting is still better than avoiding the preregistration entirely. See Figure 1 in Lakens (2024) and Table 1 in Wilroth & Atherton (2024) for guidance on and examples of reporting preregistration deviations.

4.2 What's the right resolution for a modelling preregistration template?

We aimed to design a preregistration template that was 'parsimonious', in that 1) template items should be specific and exhaustive enough to adequately constrain researcher degrees of freedom (Wicherts et al., 2016) so as to; 2) facilitate transparent documentation of the modelling paths and decisions conducted during research, while; 3) not requiring a prohibitive level of detail. Increasing the completeness and resolution of the template in capturing decisions throughout the modelling process may improve the ability of preregistration to restrict researcher degrees of freedom and transparently report results, however it also makes the preregistration more challenging to complete. Refining the template during user-testing, greatly reduced the perceived difficulty of preregistration. However, our case study researchers conceded that some modellers may still consider the final template to be too cumbersome to invest the time in preregistration.

The writing and reporting phase of the case study also highlighted the need for ongoing refinement of the template content and structure in pursuit of our second use-case for the adaptive preregistration template: ‘preregistration checking’. Information about the registered flexibility for determining alternative model functional forms was spread over multiple sub-sections within the case study’s preregistration template, hampering our ability to conduct preregistration checking post-analysis (see Table 2). We simplified the template during testing to ameliorate this issue, but it might be further improved by building modularity into the template and preregistration platform when preregistering multiple or complex models, such as in our case study.

4.3 Future Work

While we believe we have partly addressed the problem of preregistering model-based research with the adaptive preregistration protocol, further work remains, particularly around developing cyberinfrastructure to effectively and transparently handle registering and reporting flexible analyses. Improvements to the template content, structure and visual presentation are limited by the infrastructure at existing registries. A registry platform that can accommodate adaptive preregistration requires a dedicated instrument for capturing flexible analyses, as well as reporting their outcomes and linking them to downstream decisions in subsequent preregistrations, which cannot be accommodated by existing registries, such as the Open Science Foundation (OSF, https://help.osf.io/article/330-welcome-to-registrations), OSF registries (https://www.cos.io/products/osf-registries), As Predicted (https://aspredicted.org/), and EcoEvoArXiv (https://ecoevorxiv.org/). Similarly, a preregistration platform with modular content that is shown to the user conditionally based on their previous preregistration processes, such that the preregistration ‘expands’ as items are completed, may streamline the preregistration process.

As open-science practices are becoming increasingly embedded in the publication process, e.g. reproducibility checklists at Conservation Biology and data- and code-checking in the peer-review workflow at Ecology Letters (Thrall et al., 2023), preregistration should further streamline the publication process by enhancing the transparency and rigour of the study. We encourage researchers to preregister their research using our template and adaptive preregistration protocol flexibly, adopting elements that work and experimenting with alternative implementations that work better. New preregistration templates for expanding application of preregistration specific modelling approaches and methods could be developed by integrating our template with existing reporting checklists and guidelines, such as ODMAP for reporting Species Distribution Models (Fitzpatrick et al., 2021). Alternately, researchers may identify a reduced version of our template with a minimum set of items that balances trade-off between an exhaustive and specific template and the resource burden of pre-specifying the analysis.

5. Conclusions

Although challenges remain in preregistering ecological modelling, it is important to remember the costs and risks of not doing so. Ecological modelling has long been plagued by poor transparency. Incomplete reporting of the modelling process risks models being used in inappropriate applications or decisions and may mask poor model design or serious flaws in the model, resulting in adverse or irreversible outcomes when informing decision-making. Moreover, a lack of transparency in describing ecological models and reporting the modelling process provides significant opportunity for undisclosed researcher degrees of freedom, and hence, the possibility of Questionable Research Practices, with researchers routinely executing alternative analyses, and selectively reporting them when they do. Reproducible, transparent and reliable models are essential for sound conservation decision-making, and adaptive preregistration can improve documentation of modelling decisions, which is helpful for remembering and explaining the study's method at the time of report-writing, to aid others in replicating and understanding model-based research, and to communicate to funders and stakeholders.

We provide a template and methodology for adaptive preregistration that we hope will extend the benefits of preregistration to model-based research in a way that strengthens rather than inhibits that research. Although we focus on ecological applications in this paper, many elements of the template are relevant across fields, and the template is open source and can be readily adapted to new fields and purposes.

Conflict of Interest Statement

The authors declare no conflict of interest.

Supporting Information

Filename	**Description**
Appendix_S1-ModelingWorkflowIdentification_Method.docx	**Appendix S1:** Description of the method for identifying the modelling workflow used in the workshop.

Figure-S1-coded-workshop-outputs.pdf	**Figure S1:** Figure of workshop output coding and analysis, described in Appendix S1.
Appendix_S2-UserResearchWorkshopAnalysis_MaterialsMethods.docx	**Appendix S2:** Additional details of the user research and analysis process, including workshop materials and analysis.
Appendix_S3-AdaptivePreregistration_UserGuide.pdf	**Appendix S3:** Detailed user guide to implementing adaptive preregistration transparently using git and GitHub. .pdf reproduction of coauthor-1 et al. (2024), https://anonymous.4open.science/status/EcoConsPreReg-4C05.
Appendix_S4-EcoConsPreReg_template.pdf	**Appendix S4:** Preregistration template in .pdf format, generated from https://anonymous.4open.science/r/EcoConsModPreReg-6FF5..
Appendix_S5-EcoConsPreReg_template.docx	**Appendix S5:** Preregistration template in .docx format, generated from https://anonymous.4open.science/r/EcoConsModPreReg-6FF5.
Appendix_S6-case_study_preregistration.pdf	**Appendix S6:** Final version of the case study preregistration.

Appendix_S7-case_study_pilot_analysis_summary_report_2024_01_24.docx	**Appendix S7:** Summary report of the case study pilot analysis.

References

Addison, P. F., Rumpff, L., Bau, S. S., Carey, J. M., Chee, Y. E., Jarrad, F. C., Jarrad, F. C., McBride, M. F., & Burgman, M. A. (2013). Practical solutions for making models indispensable in conservation decision-making. *Diversity and Distributions*, *19*(5-6), 490–502. https://doi.org/10.1111/ddi.12054

Agnoli, F., Wicherts, J. M., Veldkamp, C. L. S., Albiero, P., & Cubelli, R. (2017). Questionable research practices among Italian research psychologists. *PLoS One*, *12*(3), e0172792. https://doi.org/10.1371/journal.pone.0172792

Alspaugh, S., Zokaei, N., Liu, A., Jin, C., & Hearst, M. A. (2019). Futzing and Moseying: Interviews with Professional Data Analysts on Exploration Practices. *IEEE Transactions on Visualization and Computer Graphics*, *25*(1), 22–31. https://doi.org/10.1109/TVCG.2018.2865040

Arnold, T., Joseph H. A. Guillaume, Guillaume, J. H. A., Tuomas J. Lahtinen, Lahtinen, T. J., W. Vervoort, & Vervoort, R. W. (2020). From ad-hoc modelling to strategic infrastructure: A manifesto for model management. *Environmental Modelling and Software*, *123*, 104563. https://doi.org/10.1016/j.envsoft.2019.104563

Augusiak, J., Van den Brink, P. J., & Grimm, V. (2014). Merging validation and evaluation of ecological models to “evaludation”: A review of terminology and a practical approach. *Ecological Modelling*, *280*, 117–128. https://doi.org/10.1016/j.ecolmodel.2013.11.009

Bakker, M., Veldkamp, C. L. S., van Assen, M. A. L. M., Crompvoets, E. A. V., Ong, H. H., Nosek, B. A., Soderberg, C. K., Mellor, D. T., & Wicherts, J. M. (2018). Ensuring the quality and specificity of preregistrations. *PsyArXiv*. https://doi.org/10.31234/osf.io/cdgyh

Baldwin, J. R., Pingault, J. B., Schoeler, T., Sallis, H. M., & Munafò, M. R. (2022). Protecting against researcher bias in secondary data analysis: Challenges and potential solutions. *Eur J Epidemiol*, *37*(1), 1–10. https://doi.org/10.1007/s10654-021-00839-0

Bodner, K., Fortin, M.-J., & Molnár, P. K. (2020). Making predictive modelling ART: Accurate, reliable, and transparent. *Ecosphere*, *11*(6). https://doi.org/10.1002/ecs2.3160

Braga, P. H. P., Hébert, K., Hudgins, E. J., Scott, E. R., Edwards, B. P. M., Sánchez Reyes, L. L., Grainger, M. J., Foroughirad, V., Hillemann, F., Binley, A. D., Brookson, C. B., Gaynor, K. M., Shafiei Sabet, S., Güncan, A., Weierbach, H., Gomes, D. G. E., & Crystal-Ornelas, R. (2023). Not just for programmers: How GitHub can accelerate collaborative and reproducible research in ecology and evolution. *Methods in Ecology and Evolution*, *14*(6), 1364–1380. https://doi.org/10.1111/2041-210X.14108

Brodeur, A., Cook, N. M., Hartley, J. S., & Heyes, A. (2024). Do Preregistration and Preanalysis Plans Reduce *p* -Hacking and Publication Bias? Evidence from 15,992

Test Statistics and Suggestions for Improvement. *Journal of Political Economy Microeconomics*, *2*(3), 527–561. https://doi.org/10.1086/730455

Brudvig, L. A. (2017). Toward prediction in the restoration of biodiversity. *Journal of Applied Ecology*, *54*(4), 1013–1017. https://doi.org/10.1111/1365-2664.12940

Camerer, C. F., Dreber, A., Holzmeister, F., Ho - Nature Human, T. H., & 2018. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature.com*.

Campbell, H. (2021). The consequences of checking for zero-inflation and overdispersion in the analysis of count data. *Methods in Ecology and Evolution*, *12*(4), 665–680. https://doi.org/10.1111/2041-210x.13559

Connolly, S. R., Keith, S. A., Colwell, R. K., & Rahbek, C. (2017). Process, Mechanism, and Modelling in Macroecology. *Trends Ecol Evol*, *32*(11), 835–844. https://doi.org/10.1016/j.tree.2017.08.011

Dwork, C., Feldman, V., Hardt, M., Pitassi, T., Reingold, O., & Roth, A. (2015). The reusable holdout: Preserving validity in adaptive data analysis. *Science*, *349*(6248), 636–638. https://doi.org/10.1126/science.aaa9375

Evans, T. R., Branney, P., Clements, A., & Hatton, E. (2023). Improving evidence-based practice through preregistration of applied research: Barriers and recommendations. Accountability in Research, 30(2), 88–108. https://doi.org/10.1080/08989621.2021.1969233

Fitzpatrick, M. C., Lachmuth, S., & Haydt, N. T. (2021). The ODMAP protocol: A new tool for standardized reporting that could revolutionize species distribution modelling. *Ecography*. https://doi.org/10.1111/ecog.05700

Fraser, H., Parker, T., Nakagawa, S., Barnett, A., & Fidler, F. (2018). Questionable research practices in ecology and evolution. *PLoS One*, *13*(7), e0200303. https://doi.org/10.1371/journal.pone.0200303

Gelman, A., & Loken, E. (2013). The garden of forking paths: Why multiple comparisons can be a problem, even when there is no “fishing expedition” or “p-hacking” and the research hypothesis was posited ahead of time. *Department of Statistics, Columbia University*.

coauthor-1, et al.(2024). *EcoConsPreReg: Adaptive Preregistration for Model-Based Research in Ecology and Conservation*. Zenodo [link suppressed for peer-review]

Gould, E., Fraser, H. S., Parker, T. H., Nakagawa, S., Griffith, S. C., Vesk, P. A., Fidler, F., Hamilton, D. G., Abbey-Lee, R. N., Abbott, J. K., Aguirre, L. A., Alcaraz, C., Aloni, I., Altschul, D., Arekar, K., Atkins, J. W., Atkinson, J., Baker, C. M., Barrett, M., … Zitomer, R. A. (2025). Same data, different analysts: variation in effect sizes due to analytical decisions in ecology and evolutionary biology. *BMC Biology*, *23*(1), 35. https://doi.org/10.1186/s12915-024-02101-x

coauthor-1 (2025) *EcoConsModPreReg: A Quarto Template for Preregistering Modelling Studies*. Zenodo [link suppressed for peer-review]

Hämäläinen, R. P., & Lahtinen, T. J. (2016). Path dependence in Operational Research—How the modelling process can influence the results. *Operations Research Perspectives*, *3*, 14–20. https://doi.org/10.1016/j.orp.2016.03.001

Hoffmann, S., Schönbrodt, F., Elsas, R., Wilson, R., Strasser, U., & Boulesteix, A. L. (2021). The multiplicity of analysis strategies jeopardizes replicability: Lessons learned across disciplines. *R Soc Open Sci*, *8*(4), 201925. https://doi.org/10.1098/rsos.201925

Hofman, J. M., Chatzimparmpas, A., Sharma, A., Watts, D. J., & Hullman, J. (2023). *Pre-registration for Predictive Modelling* (arXiv:2311.18807). arXiv. https://arxiv.org/abs/2311.18807

Ioannidis, J. P. A. (2022). Pre-registration of mathematical models. *Mathematical Biosciences*, *345*, 108782. https://doi.org/10.1016/j.mbs.2022.108782

Jones, C., Vivian, L., Wootton, H., Gould, E., Yen, J., Mole, B., Thomas, F., Backstrom, A., and Just, K. (2025, in review). Vegetation responses to managed river flow events and regimes. Arthur Rylah Institute for Environmental Research Unpublished Report for the Department of Energy, Environment and Climate Action, Victoria.

Kimmel, K., Avolio, M. L., & Ferraro, P. J. (2023). Empirical evidence of widespread exaggeration bias and selective reporting in ecology. *Nature Ecology & Evolution*, *7*(9), 1525–1536. https://doi.org/10.1038/s41559-023-02144-3

Koivisto, E., & Mäntylä, E. (2024). Are Open Science instructions targeted to ecologists and evolutionary biologists sufficient? A literature review of guidelines and journal data policies. *Ecology and Evolution*, *14*(7), e11698. https://doi.org/10.1002/ece3.11698

Lahtinen, T. J., Guillaume, J. H. A., & Hämäläinen, R. P. (2017). Why pay attention to paths in the practice of environmental modelling. *Environmental Modelling &Amp; Software*, *92*, 74–81. https://doi.org/10.1016/j.envsoft.2017.02.019

Lakens, D. (2024). When and How to Deviate From a Preregistration. *Collabra: Psychology*, *10*(1), 117094. https://doi.org/10.1525/collabra.117094

Lewis, A. S. L., Rollinson, C. R., Allyn, A. J., Ashander, J., Brodie, S., Brookson, C. B., Collins, E., Dietze, M. C., Gallinat, A. S., Juvigny-Khenafou, N., Koren, G., McGlinn, D. J., Moustahfid, H., Peters, J. A., Record, N. R., Robbins, C. J., Tonkin, J., & Wardle, G. M. (2023). The power of forecasts to advance ecological theory. *Methods in Ecology and Evolution*, *14*(3), 746–756. https://doi.org/10.1111/2041-210x.13955

Liu, Y., Althoff, T., & Heer, J. (2020). Paths Explored, Paths Omitted, Paths Obscured: Decision Points & Selective Reporting in End-to-End Data Analysis. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3313831.3376533

Lupia, A., & Elman, C. (2014). Openness in Political Science: Data Access and Research Transparency. *PS: Political Science & Politics*, *47*(01), 19–42. https://doi.org/10.1017/s1049096513001716

MacEachern, S. N., & Van Zandt, T. (2019). Preregistration of Modelling Exercises May Not Be Useful. *Comput Brain Behav*, *2*(3-4), 179–182. https://doi.org/10.1007/s42113-019-00038-x

Makel, M. C., Hodges, J., Cook, B. G., & Plucker, J. A. (2023). *Both Questionable and Open Research Practices Are Prevalent in Education Research*.

Mouquet, N., Lagadeuc, Y., Devictor, V., Doyen, L., Duputié, A., Eveillard, D., Faure, D., Garnier, E., Gimenez, O., Huneman, P., Jabot, F., Jarne, P., Joly, D., Julliard, R., Kéfi, S., Kergoat, G. J., Lavorel, S., Le Gall, L., Meslin, L., … Loreau, M. (2015). REVIEW: Predictive ecology in a changing world. *Journal of Applied Ecology*, *52*(5), 1293–1310. https://doi.org/10.1111/1365-2664.12482

Neoh, M. J. Y., Carollo, A., Lee, A., & Esposito, G. (2023). Fifty years of research on questionable research practises in science: quantitative analysis of co-citation patterns. *Royal Society Open Science*, *10*(10), 230677. https://doi.org/10.1098/rsos.230677

Parker, T., Fraser, H., & Nakagawa, S. (2019). Making conservation science more reliable with preregistration and registered reports. *Conservation Biology.*

Pavelin, K., Cham, J. A., de Matos, P., Brooksbank, C., Cameron, G., & Steinbeck, C. (2012). Bioinformatics meets user-centred design: A perspective. *PLoS Comput Biol*, *8*(7), e1002554. https://doi.org/10.1371/journal.pcbi.1002554

Popovic, G., Mason, T. J., Drobniak, S. M., Marques, T. A., Potts, J., Joo, R., Altwegg, R., Burns, C. C. I., McCarthy, M. A., Johnston, A., Nakagawa, S., McMillan, L., Devarajan, K., Taggart, P. L., Wunderlich, A., Mair, M. M., Martínez-Lanfranco, J. A., Lagisz, M., & Pottier, P. (2024). Four principles for improved statistical ecology. *Methods in Ecology and Evolution*, 2041–210X.14270. https://doi.org/10.1111/2041-210X.14270

Prosperi, M., Bian, J., Buchan, I. E., Koopman, J. S., Sperrin, M., & Wang, M. (2019). Raiders of the lost HARK: A reproducible inference framework for big data science. *Palgrave Communications*, *5*(1). https://doi.org/10.1057/s41599-019-0340-8

Pu, X., Zhu, L., Kay, M., & Conrad, F. (2019). *Designing for preregistration in practice: Multiple norms and purposes*. *Preprint*. https://doi.org/https://doi.org/10.31219/osf.io/w5ked

Roettger, T. B. (2019). Researcher degrees of freedom in phonetic research. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *10*(1). https://doi.org/10.5334/labphon.147

Schmolke, A., Thorbek, P., Deangelis, D. L., & Grimm, V. (2010). Ecological models supporting environmental decision making: A strategy for the future. *Trends in Ecology & Evolution*, *25*(8), 479–486. https://doi.org/10.1016/j.tree.2010.05.001

Silberzahn, R., Uhlmann, E. L., Martin, D. P., Anselmi, P., Aust, F., Awtrey, E., Bahník, Š., Bai, F., Bannard, C., Bonnier, E., Carlsson, R., Cheung, F., Christensen, G., Clay, R., Craig, M. A., Dalla Rosa, A., Dam, L., Evans, M. H., Flores Cervantes, I., … Nosek, B. A. (2018). Many Analysts, One Data Set: Making Transparent How Variations in Analytic Choices Affect Results. *Advances in Methods and Practices* *in Psychological Science*, *1*(3), 337–356. https://doi.org/10.1177/2515245917747646

Smaldino, P. (2020). *How to translate a verbal theory into a formal model.*

Srivastava, S. (2018). *Sound Inference in Complicated Research: A Multi-Strategy Approach*. https://doi.org/10.31234/osf.io/bwr48

Tredennick, A. T., Hooker, G., Ellner, S. P., & Adler, P. B. (2021). A practical guide to selecting models for exploration, inference, and prediction in ecology. *Ecology*, *102*(6), e03336. https://doi.org/10.1002/ecy.3336

Wicherts, J. M., Veldkamp, C. L., Augusteijn, H. E., Bakker, M., van Aert, R. C., & van Assen, M. A. (2016). Degrees of Freedom in Planning, Running, Analyzing, and Reporting Psychological Studies: A Checklist to Avoid p-Hacking. In *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2016.01832

Willroth, E. C., & Atherton, O. E. (2024). Best Laid Plans: A Guide to Reporting Preregistration Deviations. *Association for Psychological Science*. https://doi.org/https://doi.org/10.1177/25152459231213802

Yates, K. L., Bouchet, P. J., Caley, M. J., Mengersen, K., Randin, C. F., Parnell, S., Fielding, A. H., Bamford, A. J., Ban, S., Barbosa, A. M., Dormann, C. F., Elith, J., Embling, C. B., Ervin, G. N., Fisher, R., Gould, S., Graf, R. F., Gregr, E. J., Halpin, P. N., … Sequeira, A. M. M. (2018). Outstanding Challenges in the Transferability of Ecological Models. *Trends Ecol. Evol. (Amst.)*, *33*(10), 790–802. https://doi.org/10.1016/j.tree.2018.08.001

Zvereva, E. L., & Kozlov, M. V. (2021). Biases in ecological research: Attitudes of scientists and ways of control. *Sci Rep*, *11*(1), 226. https://doi.org/10.1038/s41598-020-80677-4

Appendix S1: Modelling Workflow Identification

Task 1: Identify Scientific Workflows for Ecological Modelling

The goal of this analysis task was to identify scientific workflows for ecological modelling, both in practice, and idealised, with the broader aim of identifying a workflow that describes the modelling cycle that should underpin the structure and content of a preregistration template for ecological modelling.

We conducted a literature review on 'good modelling practice' and structured decision making in ecological management, coding distinct tasks in the modelling workflow into 'phases,' 'steps' and 'sub-steps' using the qualitative data analysis software ATLAS.ti (version 8.4.13, 2019). See the documents below, stored on the OSF at https://osf.io/tz5da/?view_only=992aca57db814e9484a603f8e09b349b:

- `PRT–WorkshopCoding – Memo Manager.csv` contains descriptions of modelling workflow phases derived from Modelling Workflow Identification Task 1 analysis.
- `references.bib` contains all references informing Task 1 Modelling Workflow Identification.

We then synthesised a preliminary idealised scientific workflow describing ecological modelling (https://osf.io/fgd23?view_only=b60796fed71f4cc8abe9e5feac5e9465, https://osf.io/v7kjh?view_only=b60796fed71f4cc8abe9e5feac5e9465). This work informed Workshop Activity 1, whereby participants collated and categorised modelling activities from their personal modelling workflows under the main phases of this workflow. See **Appendix S2** and Section 2.3 of the manuscript for workshop details.

Task 2: Describe a General Workflow for Ecological Modelling

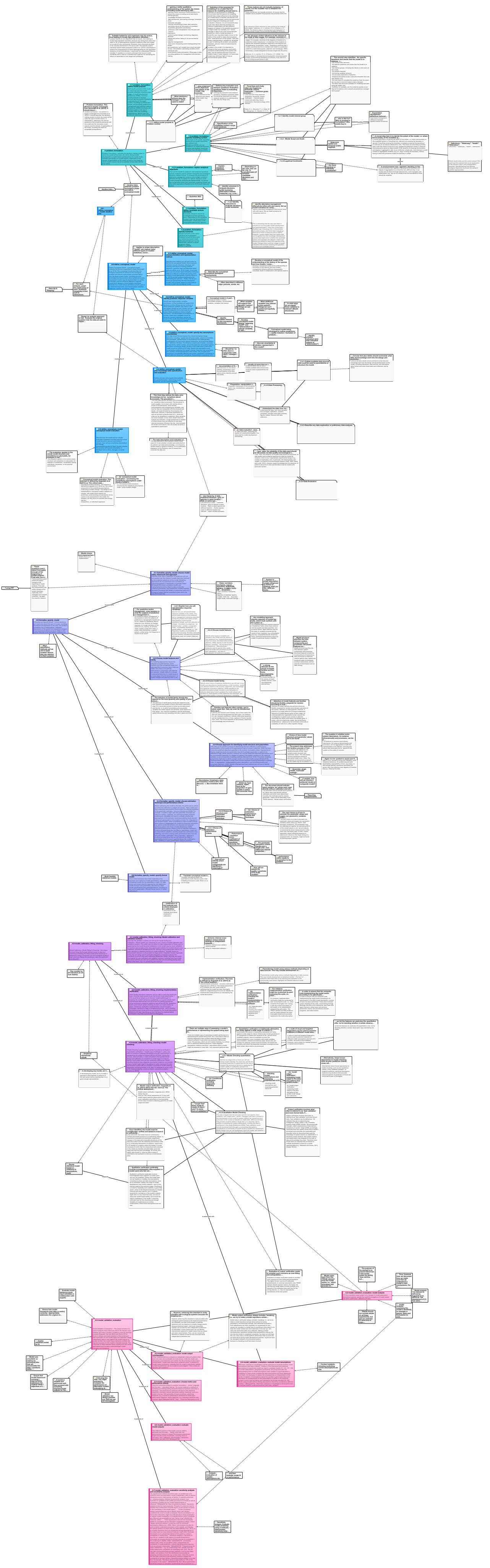
Participant responses in workshop discussions were analysed using a combination of inductive and deductive coding using the qualitative coding analysis software ATLAS.ti (ATLAS.ti Scientific Software Development GmbH, 2019). Each decision-point from the personal modelling workflows in Workshop Activity 1 were coded as belonging to one of these phases, steps and sub-steps. We revised the coding structure according to patterns and themes identified across multiple personal workflows. For example, we added the final phase "model analysis", which aimed to capture the fact that analysts usually present the results of their modelling to clients, decision-makers or other stakeholders that must interpret the evidence and are responsible for making management decisions. In this way, we identified a common or generalised workflow from different types of applied ecological modelling projects that synthesised both idealised norms and norms of practice. This final workflow underpinned our draft preregistration template used in the case study.

The documents relevant to this analysis are stored on the OSF at https://osf.io/tz5da/?view_only=992aca57db814e9484a603f8e09b349b:

- `PRT-WorkshopCoding.xml` contains the entire Atlas.TI database of coded workshop outputs in .xml format.
- `individual_sheets_FW.csv` contains decision-steps identified in individual workflow worksheets for fieldwork. Decision-steps are coded into phases and sub decision-points following Workshop Activity 1 group worksheets (see Appendix S2 for details).
- `individual_sheets_MDA.csv` contains decision-steps identified in individual workflow worksheets for modelling. Decision-steps are coded into phases and sub decision-points following workshop activity 1 group worksheets (see Appendix S2 for details).
- `PRT-WorkshopCoding - Quotation Manager.csv` Contains individual modelling workflow decision-steps coded into modelling phases and decision-points.
- `PRT-WorkshopCoding - Code Manager.csv` modelling phases and decision-steps and their descriptions, derived from Modelling Workflow Identification Task 1 and Task 2 analysis.
- **Figure S1 – Coded Workshop Outputs** (See attached file "Figure-S1-coded-workshop-outputs.pdf"). Numbered boxes correspond to phases, steps and sub-steps of the modelling process. Each different phase of the modelling process is assigned a different colour. Coloured boxes are components of the modelling process that are synthesised from a combination of literature and workshop Task 1, with accompanying text describing the phase, step or sub-step. Grey boxes correspond to verbatim text from Task 1 workflows (indicated by the symbol) or from the literature (indicated by the symbol), representing different modelling activities. The relationship of each node in the diagram is represented by arrows. Dashed arrows represent modelling activities that we have categorised as belonging to the associated modelling phase, step sub-step, or activity. Whereas solid lines represent cases where the relationship is drawn from existing taxonomies of the modelling process reported in the literature, with directionality indicating hierarchy. The ordinal numbering of the modelling component corresponds to the phase, step and sub-step in our taxonomy of the modelling process.

References

ATLAS.ti Scientific Software Development GmbH (2019) ATLAS.ti Mac (Version 8.4.13) [Computer program]. Available at: https://atlasti.com (Downloaded: 29 March 2019).



Appendix S2: User Research Workshop Materials

- Introductory Presentation 1
  https://osf.io/wezy3/?view_only=b60796fed71f4cc8abe9e5feac5e9465
- Introductory Presentation 2
  https://osf.io/egjwa/?view_only=b60796fed71f4cc8abe9e5feac5e9465

Activity	**Time**
Introduction, Background to the "Credibility Revolution", Introductions & Housekeeping	10.30 am – 11.30 am (1 hour)
Tea Break	11.30 am – 11.45 pm (15 minutes)
Breakout group – Activity 1 "Designing Preregistration Templates - Workflows and Decision Steps"	11.45 am – 1.15 pm (1.5 hours)
Lunch	1.15 pm – 2.00 pm (45 minutes)
Breakout group – Activity 2 "Challenges & Solutions to Preregistration in Ecology"	2.00 pm – 3.00 pm (1 hour)
Breakout group report back, whole workshop discussion and debrief.	3.00 pm – 4.00 pm (1 hour)

Activity 1 – Identifying Modelling Workflows

Note that this activity was run in two separate groups, with one group focusing on field work, and one focusing on modelling. Elements of the field work activity responses were collated into the analysis of the modelling activity outputs.

Materials

Workshop attendees were provided a 'Project Workflow and Decision Sequence' worksheet (https://osf.io/eza24/?view_only=b60796fed71f4cc8abe9e5feac5e9465) for individually describing decision steps and alternatives taken in a recent or memorable study. Next, each group collated their decision-steps from their individual worksheets onto group worksheets for modelling (https://osf.io/fgd23/?view_only=b60796fed71f4cc8abe9e5feac5e9465) and for field studies (https://osf.io/w68jt/?view_only=b60796fed71f4cc8abe9e5feac5e9465) with accompanying handouts describing workflow phases and steps (modelling: https://osf.io/v7kjh/?view_only=b60796fed71f4cc8abe9e5feac5e9465; field work: https://osf.io/uayt3/?view_only=b60796fed71f4cc8abe9e5feac5e9465).

Activity 2 – Identifying Challenges and Solutions to Preregistration for Modelling in Ecology

Materials

Activity 2 consisted of facilitated group discussions based on prepared questions and talking points informed by recent debates about preregistration in the literature, for example:

- What should a research workflow look like for a modeller in an applied ecological setting who wishes to preregister their work?
- What barriers are there for preregistering model-based research?
- How do we accommodate the iterative cycle of model development into the preregistration process?
- At what stage in a modelling or structured decision-making process should preregistration begin?
- Should we and how can we change the medium of the template and/or archiving platform to accommodate any roadblocks?
- How could we change the medium of both the template and the archiving platform to accommodate these issues?
- What procedural and or technical solutions could we implement to address these roadblocks to preregistration?
- What decision-tools could be used to aid in completing a preregistration?
- What sort of parsimony should we aim for in terms of the resolution of the templates and therefore their applicability across different methodologies?

Ecological Modelling Preregistration Template

[redacted]

2025-08-09

Background and Instructions

Here we present a preregistration template for ecological models in ecology, conservation and related fields. For non-trivial modelling studies, especially where model parameter and structure is in any way data-contingent, we recommend taking an *Adaptive Preregistration* approach ([redacted].

Replace author, author-affiliations and persistent ID's (e.g. ORCID iD), keywords, title and abstract metadata as relevant to your study.

All preregistration items should be completed, excluding items marked as optional or in cases where they are not applicable to your study. Additional preregistration items can be added as required at the researchers' discretion.

Table of contents

Study Information

CRediT Contribution Statement

Preregistration Item

Identify potential contributions according to the CRediT taxonomy (https://doi.org/10.1371/journal.pone.0244611.t001) and write a CRediT contribution statement.

Conflict of Interest Statement

Preregistration Item

- ☐ Explain any real or perceived conflicts of interest with this study execution. For example, any interests or activities that might be seen as influencing the research (e.g., financial interests in a test or procedure, funding by companies for research).

Data Availability Statement

Preregistration Item

Select one option from below:

- ☐ "We plan to make the data available (yes / no)," specify the planned data availability level from the following options:
  - Data access via download; usage of data for all purposes (public use file)
  - Data access via download; usage of data restricted to scientific purposes (scientific use file)
  - Data access via download; usage of data has to be agreed and defined on an individual case basis
  - Data access via secure data centre (no download, usage/analysis only in a secure data centre)
  - Data available upon email request by member of scientific community
  - Other (please specify)
- ☐ "Data will not be made available"
  - ☐ Justify reason for not making data available.

Code Availability

Preregistration Item

Select one option from below:

- ☐ "We plan to make the code available (yes / no)," specify the planned code availability level from the following options:
  - Code access via download; usage of code for all purposes (public use file)
  - Code access via download; usage of code restricted to scientific purposes (scientific use file)
  - Code access via download; usage of code has to be agreed and defined on an individual case basis
  - Code access via secure code centre (no download, usage/analysis only in a secure code centre)
  - Code available upon email request by member of scientific community
  - Other (please specify)
- ☐ "Code will not be made available"
  - ☐ Justify reason for not making code available.

Ethics

> Preregistration Item
>
> ☐ Select and respond to the relevant item below:
> - If relevant institutional ethical approval for the study has been obtained, provide the relevant identifier, and link to relevant documents.
> - If ethical approval has not yet been obtained, but is required, provide a brief overview of plans for obtaining study approval in accordance with established ethical guidelines.
> - Alternatively, if the study is exempt from ethical approval, explain exemption.

1 Problem Formulation

> Rationale & Explanation
>
> This section specifies the decision-making context in which the model will be used or the intended scope and context of conclusions. Important components include the decision maker and stakeholders (including experts) and their view on: i) the nature of the problem or decision addressed and how the scope of the modelling tool fits within the (broader) context (i.e. model purpose; ii) the spatial and temporal scales relevant to the decision context; iii) specified desired outputs; iv) role and inclusion in model development and testing; v) whether they foresee unacceptable outcomes that need to be represented in the model (i.e. as constraints), and; vi) what future scenarios does the model need to account for (noting this may be revised later). It should also provide a summary of the domain of applicability of the model, and reasonable extrapolation limits (Grimm et al., 2014).

1.1 Model Context and Purpose

> Rationale & Explanation
>
> Defining the purpose of the model is critical because the model purpose influences choices at later stages of model development (Jakeman et al., 2006). Common model purposes in ecology include: gaining a better qualitative understanding of the target system, synthesising and reviewing knowledge, and providing guidance for management and decision-making (Jakeman et al., 2006). Note that modelling objectives are distinct from the analytical objectives of the model.
> The scope of the model includes temporal and spatial resolutions, which should also be defined here (Mahmoud et al., 2009). Any external limitations on model development, analysis and flexibility should also be outlined in this section (Jakeman et al., 2006).

1.1.1 Key stakeholders and model users

> Preregistration Item
>
> *Identify relevant interest groups:*
> ☐ Who is the model for?
> ☐ Who is involved in formulating the model?
> ☐ How will key stakeholders be involved in model development?
> ☐ Describe the decision-making context in which the model will be used (if relevant).

1.1.2 Model purpose, context and problem context

Preregistration Item

Briefly outline:

- ☐ the ecological problem,
- ☐ the decision problem (if relevant), including the decision-trigger and any regulatory frameworks relevant to the problem,
- ☐ how the model will address the problem, being clear about the scope of the model i.e. is the model addressing the whole problem, or part of it? Are there any linked problems that your model should consider?
- ☐ Ensure that you specify any focal taxa and study objectives.

1.1.3 Analytical objectives

Explanation

How will the model be analysed, what analytical questions will the model be used to answer? For example, you might be using your model in a scenario analysis to determine which management decision is associated with minimum regret or the highest likelihood of improvement. Other examples from ecological decision-making include: to compare the performance of alternative management actions under budget constraint (Fraser et al., 2017), to search for robust decisions under uncertainty (McDonald-Madden et al., 2008), to choose the conservation policy that minimises uncertainty (McCarthy et al., 2011). See other examples in (Moallemi et al., 2019).

Preregistration Item

Provide detail on the analytical purpose and scope of the model:

- ☐ How will the model be analysed and what analytical questions will the model be used to answer?
- ☐ Candidate decisions should be investigated and are specified a priori. Depending on the modelling context, they may be specified by stakeholders, model users or the analyst (Moallemi et al., 2019).
  - ☐ Describe the method used to identify relevant management actions and
  - ☐ specify management actions to be considered included in the model.
  - ☐ Are there potentially unacceptable management or policy outcomes identified by stakeholders that should be captured in the model, i.e. as constraints?
- ☐ Are there scenarios that model inputs or outputs that must accommodated? Scenarios should be set a priori, (i.e. before the model is built, Moallemi et al., 2019) and may be stakeholder-defined or driven by the judgement of the modeller or other experts (Mahmoud et al., 2009).
  - ☐ If relevant, describe what processes you will use to elicit and identify relevant scenarios, e.g. literature review, structured workshops with stakeholders or decision-makers.
  - ☐ Specify scenarios under which decisions are investigated.

1.1.4 Logistical Constraints

Preregistration Item

- ☐ What degree of flexibility is required from the model? Might the model need to be quickly reconfigured to explore new scenarios or problems proposed by clients / managers / model-users?
- ☐ Are there any limitations on model development analysis and flexibility, such as time or budget constraints, for example, does a model need to be deployed rapidly?
  - ☐ When must the model be completed by, e.g. to help make a decision?

1.1.5 Model Scope, Scale and Resolution

Preregistration Item

- ☐ The choice of a model's boundaries is closely linked to the choice of how finely to aggregate the behaviour within the model (Jakeman et al., 2006) - what is the intended scale, and resolution of the model (temporal, spatial or otherwise)?
- ☐ Where is the boundary of the modelled system? Everything outside beyond the boundary and not crossing it is to be ignored within the domain of the model, and everything crossing the boundary is to be treated as external forcing (known/unknown), or else as model outputs (observed, or not, Jakeman et al., 2006).

1.1.6 Intended application of results

Explanation

Preregistration Items in this section are relevant to model transferability (Yates et al., 2018) and constraints on generality in model analysis interpretation. How far do can the results be extrapolated based on the study design (data + model + analysis)? For instance, if there are many confounding variables and not enough spatial / environmental replication, then making broader more general claims beyond the stated boundaries of the model (Section 1.1.3) may not be warranted. However, larger generalisations about results may be acceptable if the data comes from experimentally manipulated or controlled systems.

Preregistration Item

- ☐ What is the intended domain in which the model is to be applied? Are there any reasonable extrapolation limits beyond which you expect the model should not be applied (Grimm et al., 2014)?

1.2 Scenario Analysis Operationalisation

Preregistration Item (delete as necessary)

- ☐ How will you operationalise any scenarios identified in Section 1.1.3? For example, how will you operationalise any qualitative changes of interest, such as , 'deterioration' or 'improvement'?
- ☐ Describe how you will evaluate and distinguish the performance of alternative scenario outcomes
- ☐ Justify or otherwise explain how you chose these measures and determined performance criteria in relation to the analytical objectives, model purpose and modelling context, such as the risk attitudes of decision-makers and stakeholders within this system

2 Define Conceptual Model

Explanation

Conceptual models underpin the formal or quantitative model (Cartwright et al., 2016). The conceptual model describes the biological mechanisms relevant to the ecological problem and should capture basic premises about how the target system works, including any prior knowledge and assumptions about system processes. Conceptual models may be represented in a variety of formats, such as influence diagrams, linguistic model block diagram or bond graphs, and these illustrate how model drivers are linked to both outputs or observed responses, and internal (state) variables (Jakeman et al., 2006).

2.1 Choose elicitation and representation method

Preregistration Item

- ☐ Describe what method you will use to elicit or identify the conceptual model. Some common methods include interviews, drawings, and mapping techniques including influence diagrams, cognitive maps and Bayesian belief networks (Moon et al., 2019). It is difficult to decide and justify which method is most appropriate, see Moon et al. (2019) for guidance addressing this methodological question.
- ☐ Finally, how do you intend on representing the final conceptual model? This will likely depend on the method chosen to elicit the conceptual model.

2.2 Explain Critical Conceptual Design Decisions

Preregistration Item

List and explain critical conceptual design decisions (Grimm et al., 2014), including:

- ☐ spatial and temporal scales,
- ☐ selection of entities and processes,
- ☐ representation of stochasticity and heterogeneity,
- ☐ consideration of local versus global interactions, environmental drivers, etc.
- ☐ Explain and justify the influence of particular theories, concepts, or earlier models against alternative conceptual design decisions that might lead to alternative model structures.

2.3 Model assumptions and uncertainties

Preregistration Item

Specify key assumptions and uncertainties underlying the model design, describing how uncertainty and variation will be represented in the model (Moallemi et al., 2019). Sources of uncertainty may include:

- ☐ exogenous uncertainties affecting the system,
- ☐ parametric uncertainty in input data and
- ☐ structural / conceptual nonparametric uncertainty in the model.

2.4 Identify predictor and response variables

Explanation

The identification and definition of primary model input variables should be driven by scenario definitions, and by the scope of the model described in the problem formulation phase (Mahmoud et al., 2009).

Preregistration Item

Identify and define system system variables and structures, referencing scenario definitions, and the scope of the model as described within problem formulation:

- ☐ What variables would support taking this action or making this decision?
- ☐ What additional variables may interact with this system (things we can't control, but can hopefully measure)?
- ☐ What variables have not been measured, but may interact with the system (often occurs in field or observational studies)?
- ☐ What variables are index or surrogate measures of variables that we cannot or have not measured?
- ☐ In what ways do we expect these variables to interact (model structures)?
- ☐ Explain how any key concepts or terms within problem or decision-making contexts, such as regulatory terms, will be operationalised and defined in a biologically meaningful way to answer the research question appropriately?

2.5 Define prior knowledge, data specification and evaluation

ℹ Explanation

This section specifies the plan for collecting, processing and preparing data available for parameterisation, determining model structure, and for scenario analysis. It also allows the researchers to disclose any prior interaction with the data.

2.5.1 Collate available data sources that could be used to parameterise or structure the model

Preregistration Item

For pre-existing data (delete as appropriate):

- ☐ Document the identity, quantity and provenance of any data that will be used to develop, identify and test the model.
- ☐ For each dataset, is the data open or publicly available?
- ☐ How can the data be accessed? Provide a link or contact as appropriate, indicating any restrictions on the use of data.
- ☐ Date of download, access, or expected timing of future access.
- ☐ Describe the source of the data - what entity originally collected this data? (National Data Set, Private Organisational Data, Own Lab Collection, Other Lab Collection, External Contractor, Meta-Analysis, Expert Elicitation, Other).
- ☐ Codebook and meta-data. If a codebook or other meta-data is available, link to it here and / or upload the document(s).
- ☐ Prior work based on this dataset - Have you published / presented any previous work based on this dataset? Include any publications, conference presentations (papers, posters), or working papers (in-prep, unpublished, preprints) based on this dataset you have worked on.
- ☐ Unpublished Prior Research Activity - Describe any prior but unpublished research activity using these data. Be specific and transparent.
- ☐ Prior knowledge of the current dataset - Describe any prior knowledge of or interaction with the dataset before commencing this study. For example, have you read any reports or publications about this data?
- ☐ Describe how the data is arranged, in terms of replicates and covariates.

Sampling Plan (for data you will collect, delete as appropriate):

- ☐ Data collection procedures - Please describe your data collection process, including how sites and transects or any other physical unit were selected and arranged. Describe any inclusion or exclusion rules, and the study timeline.
- ☐ Sample Size - Describe the sample size of your study.
- ☐ Sample Size Rationale - Describe how you determined the appropriate sample size for your study. It could include feasibility constraints, such as time, money or personnel.
- ☐ If sample size cannot be specified, specify a stopping rule - i.e. how will you decide when to terminate your data collection?

2.5.2 Data Processing and Preparation

Preregistration Item

- ☐ Describe any data preparation and processing steps, including manipulation of environmental layers (e.g. standardisation and geographic projection) or variable construction (e.g. Principal Component Analysis).

2.5.3 Describe any data exploration or preliminary data analyses.

Explanation

In most modelling cases, it is necessary to perform preliminary analyses to understand the data and check that assumptions and requirements of the chosen modelling procedures are met. Data exploration prior to model fitting or development may include exploratory analyses to check for collinearity, spatial and temporal coverage, quality and resolution, outliers, or the need for transformations (Yates et al., 2018).

Preregistration Item

For each separate preliminary or investigatory analysis:

- ☐ State what needs to be known to proceed with further decision-making about the modelling procedure, and why the analysis is necessary.
- ☐ Explain how you will implement this analysis, as well as any techniques you will use to summarise and explore your data.
- ☐ What method will you use to represent this analysis (graphical, tabular, or otherwise, describe)
- ☐ Specify exactly which parts of the data will be used
- ☐ Describe how the results will be interpreted, listing each potential analytic decision, as well as the analysis finding that will trigger each decision, where possible.

2.5.4 Data evaluation, exclusion and missing data

Explanation

Documenting issues with reliability is important because data quality and ecological relevance might be constrained by measurement error, inappropriate experimental design, and heterogeneity and variability inherent in ecological systems (Grimm et al., 2014). Ideally, model input data should be internally consistent across temporal and spatial scales and resolutions, and appropriate to the problem at hand (Mahmoud et al., 2009).

Preregistration Item

- ☐ Describe how you will determine how reliable the data is for the given model purpose. Ideally, model input data should be internally consistent across temporal and spatial scales and resolutions, and appropriate to the problem at hand
- ☐ Document any issues with data reliability.
- ☐ How will you determine what data, if any, will be excluded from your analyses?
- ☐ How will outliers be handled? Describe rules for identifying outlier data, and for excluding a site, transect, quadrat, year or season, species, trait, etc.
- ☐ How will you identify and deal with incomplete or missing data?

2.6 Conceptual model evaluation

Preregistration Item

- ☐ Describe how your conceptual model will be critically evaluated. Evaluation includes both the completeness and suitability of the overall model structure.
- ☐ How will you critically assess any simplifying assumptions (Augusiak et al., 2014)?
- ☐ Will this process will include consultation or feedback from a client, manager, or model user.

3 Formalise and Specify Model

Explanation

In this section describe what quantitative methods you will use to build the model/s, explain how they are relevant to the client/manager/user's purpose.

3.1 Model class, modelling framework and approach

Explanation

Modelling approaches can be described as occurring on a spectrum from correlative or phenomenological to mechanistic or process-based (Yates et al., 2018); where correlative models use mathematical functions fitted to data to describe underlying processes, and mechanistic models explicitly represent processes and details of component parts of a biological system that are expected to give rise to the data (White & Marshall, 2019). A model 'class,' 'family' ' or 'type' is often used to describe a set of models each of which has a distinct but related sampling distribution (C. C. Liu & Aitkin, 2008). The model family is driven by choices about the types of variables covered and the nature of their treatment, as well as structural features of the model, such as link functions, spatial and temporal scales of processes and their interactions (Jakeman et al., 2006).

Preregistration Item

- ☐ Describe what modelling framework, approach or class of model you will use to implement your model and relate your choice to the model purpose and analytical objectives described in Section 1.1.2 and Section 1.1.3.

3.2 Choose model features and family

Explanation

All modelling approaches require the selection of model features, which conform with the conceptual model and data specified in previous steps (Jakeman et al., 2006). The choice of model are determined in conjunction with features are selected. Model features include elements such as the functional form of interactions, data structures, measures used to specify links, any bins or discretisation of continuous variables. It is usually difficult to change fundamental features of a model beyond an early stage of model development, so careful thought and planning here is useful to the modeller (Jakeman et al., 2006). However, if changes to these fundamental aspects of the model do need to change, document how and why these choices were made, including any results used to support any changes in the model.

3.2.1 Operationalising Model Variables

Preregistration Item

- ☐ For each response, predictor, and covariate, specify how these variables will be operationalised in the model. This should relate directly to the analytical and/or management objectives specified during the problem formulation phase. Operationalisations could include: the extent of a response, an extreme value, a trend, a long-term mean, a probability distribution, a spatial pattern, a time-series, qualitative change, such as a direction of change or, the frequency, location, or probability of some event occurring. Specify any treatment of model variables, including whether they are lumped / distributed, linear / non-linear, stochastic / deterministic (Jakeman et al., 2006).
- ☐ Provide a rationale for your choices, including why plausible alternatives under consideration were not chosen, and relate your justification back to the purpose, objectives, prior knowledge and or logistical constraints specified in the problem formulation phase (Jakeman et al., 2006).

3.2.2 Choose model family

Preregistration Item

- ☐ Specify which family of statistical distributions you will use in your model, and describe any transformations, or link functions.
- ☐ Include in your rational for selection, detail about which variables the model outputs are likely sensitive to, what aspects of their behaviour are important, and any associated spatial or temporal dimensions in sampling.

3.3 Describe *approach* for identifying model structure

Explanation

This section relates to the process of determining the best/most efficient/parsimonious representation of the system at the appropriate scale of concern (Jakeman et al., 2006) that best meets the analytical objectives specified in the problem formulation phase. Model structure refers to the choice of variables included in the model, and the nature of the relationship among those variables. Approaches to finding model structure and parameters may be knowledge-supported, or data-driven (Boets et al., 2015). Model selection methods can include traditional inferential approaches such as unconstrained searches of a dataset for patterns that explain variations in the response variable, or use of ensemble-modelling methods (Barnard et al., 2019). Ensemble modelling procedures might aim to derive a single model, or a multi-model average (Yates et al., 2018). Refining actions to develop a model could include iteratively dropping parameters or adding them, or aggregating / disaggregating system descriptors, such as dimensionality and processes (Jakeman et al., 2006).

Preregistration Item

- ☐ Specify what approach and methods you will use to identify model structure and parameters.
- ☐ If using a knowledge-supported approach to deriving model structure (either in whole or in part), specify model structural features, including:
  - the functional form of interactions (if any)
  - data structures,
  - measures used to specify links,
  - any bins or discretisation of continuous variables (Jakeman et al., 2006),
  - any other relevant features of the model structure.

3.4 Describe parameter estimation technique and performance criteria

Explanation

Before calibrating the model to the data, the performance criteria for judging the calibration (or model fit) are specified. These criteria and their underlying assumptions should reflect the desired properties of the parameter estimates / structure (Jakeman et al., 2006). For example, modellers might seek parameter estimates that are robust to outliers, unbiased, and yield appropriate predictive performance. Modellers will need to consider whether the assumptions of the estimation technique yielding those desired properties are suited to the problem at hand. For integrated or sub-divided models, other considerations might include choices about where to disaggregate the model for parameter estimation; e.g. spatial sectioning (streams into reaches) and temporal sectioning (piece-wise linear models) (Jakeman et al., 2006).

3.4.1 Parameter estimation technique

Preregistration Item

- ☐ Specify what technique you will use to estimate parameter values, and how you will supply non-parametric variables and/or data (e.g. distributed boundary conditions). For example, will you calibrate all variables simultaneously by optimising fit of model outputs to observations, or will you parameterise the model in a piecemeal fashion by either direct measurement, inference from secondary data, or some combination (Jakeman et al., 2006).
- ☐ Identify which variables will be parameterised directly, such as by expert elicitation or prior knowledge.
- ☐ Specify which algorithm(s) you will use for any data-driven parameter estimation, including supervised, or unsupervised machine learning, decision-tree, K-nearest neighbour or cluster algorithms (Z. Liu et al., 2018).

3.4.2 Parameter estimation / model fit performance criteria

Preregistration Item

- ☐ Specify which suite of performance criteria you will use to judge the performance of the model. Examples include correlation scores, coefficient of determination, specificity, sensitivity, AUC, etcetera (Yates et al., 2018).
- ☐ Relate any underlying assumptions of each criterion to the desired properties of the model, and justify the choice of performance metric in relation
- ☐ Explain how you will identify which model features or components are significant or meaningful.

3.5 Model assumptions and uncertainties

Preregistration Item

- ☐ Specify assumptions and key uncertainties in the formal model. Describe what gaps exist between the model conception, and the real-world problem, what biases might this introduce and how might this impact any interpretation of the model outputs, and what implications are there for evaluating model-output to inform inferences or decisions?

3.6 Specify formal model(s)

Explanation

Once critical decisions have been made about the modelling approach and method of model specification, the conceptual model is translated into the quantitative model.

Preregistration Item

- ☐ Specify all formal models
  - ☐ Note, For data-driven approaches to determining model structure and or parameterisation, it may not be possible to respond to this preregistration item. In such cases, explain why this is the case, and how you will document the model(s) used in the final analysis.
- ☐ For quantitative model selection approaches, including ensemble modelling, specify each model used in the candidate set, including any null or full/global model.

4 Model Calibration, Validation & Checking

4.1 Model calibration and validation scheme

Explanation

This section pertains to any data calibration, validation or testing schemes that will be implemented. For example, the model may be tested on data independent of those used to parameterise the model (external validation), or the model may be cross-validated on random sub-samples of the data used to parameterise the model (Barnard et al., 2019; internal cross-validation Yates et al., 2018). For some types of models, hyper-parameters are estimated from data, and may be tuned on further independent holdouts of the training data ("validation data").

Preregistration Item

- ☐ Describe any data calibration, validation and testing scheme you will implement, including any procedures for tuning or estimating model hyper-parameters (if any).

4.1.1 Describe calibration/validation data

Explanation & Rationale

The following items pertain to properties of the *datasets* used for calibration (training), validation, and testing.

Preregistration Item

If partitioning data for cross-validation or similar approach (delete as needed):

- ☐ Describe the approach specifying the number of folds that will be created, the relative size of each fold, and any stratification methods used for ensuring evenness of groups between folds and between calibration / validation data?

If using external / independent holdout data for model testing and evaluation (delete as needed):

- ☐ Which data will be used as a the testing data? What method will you be used for generating training / test data subsets?
- ☐ Describe any known differences between the training/validation and testing datasets, the relative size of each, as well as any stratification methods used for ensuring evenness of groups between data sets?
- ☐ It is preferable that any independent data used for model testing remains unknown to modellers during the process of model development, please describe the relationship modellers have to model validation data, will independent datasets be known or accessible to any modeller or analyst?

4.2 Implementation verification

Explanation & Examples

Model implementation verification is the process of ensuring that the model has been correctly implemented, and that the model performs as described by the model description (Grimm et al., 2014). This process is distinct from model checking, which assesses the model's performance in representing the system of interest (Conn et al., 2018).

- Checks for verification implementation should include i) thoroughly checking for bugs or programming errors, and ii) whether the implemented model performs as described by the model description (Grimm et al., 2014).
- Qualitative tests could include syntax checking of code, and peer-code review (Ivimey et al., 2023). Technical measures include using unit tests, or in-built checks within functions to prevent potential errors.

Preregistration Item

- ☐ What Quality Assurance measures will you take to verify the model has been correctly implemented? Specifying a priori quality assurance tests for implementation verification may help to avoid selective debugging and silent errors.

4.3 Model checking

Rationale & Explanation

"Model Checking" goes by many names ("conditional verification", "quantitative verification", "model output verification"), and refers to a series of analyses that assess a model's performance in representing the system of interest (Conn et al., 2018). Model checking aids in diagnosing assumption violations, and reveals where a model might need to be altered to better represent the data, and therefore system (Conn et al., 2018). Quantitative model checking diagnostics include goodness of fit, tests on residuals or errors, such as for heteroscedascity, cross-correlation, and autocorrelation (Jakeman et al., 2006).

4.3.1 Quantitative model checking

Preregistration Item

During this process, observed data, or data and patterns that guided model design and calibration, are compared to model output in order to identify if and where there are any systematic differences.

- ☐ Specify any diagnostics or tests you will use during model checking to assess a model's performance in representing the system of interest.
- ☐ For each test, specify the criteria that will you use to interpret the outcome of the test in assessing the model's ability to sufficiently represent the gathered data used to develop and parameterise the model.

4.3.2 Qualitative model checking

Explanation

This step is largely informal and case-specific, but requires, 'face validation' with model users / clients / managers who aren't involved in the development of the model to assess whether the interactions and outcomes of the model are feasible an defensible (Grimm et al., 2014). This process is sometimes called a "laugh test" or a "pub test" and in addition to checking the model's believability, it builds the client's confidence in the model (Jakeman et al., 2006). Face validation could include structured walk-throughs, or presenting descriptions, visualisations or summaries of model results to experts for assessment.

Preregistration Item

- ☐ Briefly explain how you will qualitatively check the model, and whether and how you will include users and clients in the process.

4.3.3 Assumption Violation Checks

Preregistration Item

The consequences of assumption violations on the interpretation of results should be assessed (Araújo et al., 2019).

- ☐ Explain how you will demonstrate robustness to model assumptions and check for violations of model assumptions.
- ☐ If you cannot perform quantitative assumption checks, describe what theoretical justifications would justify a lack of violation of or robustness to model assumptions.

- ☐ If you cannot demonstrate or theoretically justify violation or robustness to assumptions, explain why not, and specify whether you will discuss assumption violations and their consequences for interpretation of model outputs.
- ☐ If assumption violations cannot be avoided, explain how you will explore the consequences of assumption violations on the interpretation of results (To be completed in interim iterations of the preregistration, only if there are departures from assumptions as demonstrated in the planned tests above).

5 Model Validation and Evaluation

Explanation

The model validation & evaluation phase comprises a suite of analyses that collectively inform inferences about whether, and under what conditions, a model is suitable to meet its intended purpose (Augusiak et al., 2014). Errors in design and implementation of the model and their implication on the model output are assessed. Ideally independent data is used against the model outputs to assess whether the model output behaviour exhibits the required accuracy for the model's intended purpose. The outcomes of these analyses build confidence in the model applications and increase understanding of model strengths and limitations. Model evaluation including, model analysis, should complement model checking. It should evaluate model checking, and consider over-fitting and extrapolation. As the proportion of calibrated or uncertain parameters increases, so does the risk that the model seemingly works correctly, but for the wrong mechanistic reasons (Boettiger, 2022). Evaluation thus complements model checking because we can rule out the chance that the model fits the calibration data well, but has not captured the relevant ecological mechanisms of the system pertinent to the research question or the decision problem underpinning the model (Grimm et al., 2014). Evaluation of model outputs against external data in conjunction with the results from model checking provide information about the structural realism and therefore credibility of the model (Grimm & Berger, 2016).

5.1 Model output corroboration

Explanation

Ideally, model outputs or predictions are compared to independent data and patterns that were not used to develop, parameterise, or verify the model. Testing against a dataset of response and predictor variables that are spatially and/or temporally independent from the training dataset minimises the risk of artificially inflating model performance measures (Araújo et al., 2019). Although the corroboration of model outputs against an independent validation dataset is considered the 'gold standard' for showing that a model properly represents the internal organisation of the system, model validation is not always possible because empirical experiments are infeasible or model users are working on rapid-response timeframes, hence, why ecologists often model in the first place (Grimm et al., 2014). Independent predictions might instead be tested on sub-models. Alternatively, patterns in model output that are robust and seem characteristic of the system can be identified and evaluated in consultation with the literature or by experts to judge how accurate the model output is (Grimm et al., 2014).

Preregistration Item

- ☐ State whether you will corroborate the model outputs on external data, and document any independent validation data in step.
- ☐ It is preferable that any independent data used for model evaluation remains unknown to modellers during the process of model building (Dwork et al., 2015), describe the relationship modellers have to model validation data, e.g. will independent datasets be known to any modeller or analyst involved in the model building process?
- ☐ If unable to evaluate the model outputs against independent data, explain why and explain what steps you will take to interrogate the model.

5.2 Choose performance metrics and criteria

Explanation

Model performance can be quantified by a range of tests, including measures of agreement between predictions and independent observations, or estimates of accuracy, bias, calibration, discrimination refinement, resolution and skill (Araújo et al., 2019). Note that the performance metrics and criteria in this section are used for evaluating the structured and parameterised models (ideally) on independent holdout data, so this step is additional to any performance criteria used for determining model structure or parameterisation (Section 3.4.2).

Preregistration Item

- ☐ Specify what performance measures you will use to evaluate the model and briefly explain how each test relates to different desired properties of a model's performance.
- ☐ Spatial, temporal and environmental pattern of errors and variance can change the interpretation of model predictions and conservation decisions (Araújo et al., 2019), where relevant and possible, describe how you will characterise and report the spatial, temporal and environmental pattern of errors and variance.
- ☐ If comparing alternative models, specify what measures of model comparison or out-of-sample performance metrics will you use to find support for alternative models or else to optimise predictive ability. State what numerical threshold or qualities you will use for each of these metrics.

5.3 Model analysis

Rationale & Explanation

Uncertainty in models arises due to incomplete system understanding (which processes to include, or which interact), from imprecise, finite and sparse data measurements, and from uncertainty in input conditions and scenarios for model simulations or runs (Jakeman et al., 2006). Non-technical uncertainties can also be introduced throughout the modelling process, such as uncertainties arising from issues in problem-framing, indeterminicies, and modeller / client values (Jakeman et al., 2006).
The purpose of model analysis is to prevent blind trust in the model by understanding how model outputs have emerged, and to 'challenge' the model by verifying whether the model is still believable and fit for purpose if one or more parameters are changed (Grimm et al., 2014).
Model analysis should increase understanding of the model behaviour by identifying which processes and process interactions explain characteristic behaviours of the model system. Model analysis typically consists of sensitivity analyses preceded by uncertainty analyses (Saltelli et al., 2019), and a suite of other simulation or other computational experiments. The aim of such computational experiments is to increase understanding of the model behaviour by identifying which processes and process interactions explain characteristic behaviours of the model system (Grimm et al., 2014). Uncertainty analyses and sensitivity analyses augment one another to draw conclusions about model uncertainty.
Because the results from a full suite of sensitivity analysis and uncertainty analysis can be difficult to interpret due to the number and complexity of causal relations examined (Jakeman et al., 2006), it is useful for the analyst to relate the choice of analysis to the modelling context, purpose and analytical objectives defined in the problem formulation phase, in tandem with any critical uncertainties that have emerged during model development and testing prior to this point.

5.3.1 Uncertainty Analyses

Explanation

Uncertainty can arise from different modelling techniques, response data and predictor variables (Araújo et al., 2019). Uncertainty analyses characterise the uncertainty in model outputs, and identify how uncertainty in model parameters affects uncertainty in model output, but does not identify which model assumptions are driving this behaviour (Grimm et al., 2014; Saltelli et al., 2019). Uncertainty analyses can include propagating known uncertainties through the model, or by investigating the effect of different model scenarios with different parameters and modelling technique combinations (Araújo et al., 2019), for example. It could also include characterising the output distribution, such as through empirical construction using model output data points. It could also include extracting summary statistics like the mean, median and variance from this distribution, and perhaps constructing confidence intervals on the mean (Saltelli et al., 2019).

Preregistration Item

- ☐ Please describe how you will characterise model and data uncertainties, e.g. propagating known uncertainties through the model, investigating the effect of different model scenarios with different parameters and modelling technique combinations (Araújo et al., 2019), or empirically constructing model distributions from model output data points, and extracting summary statistics, including the mean, median, variance, and constructing confidence intervals (Saltelli et al., 2019).
- ☐ Relate your choice of analysis to the context and purposes of the model described in the problem formulation phase. For instance , discrepancies between model output and observed output may be important for forecasting models, where cost, benefit, an risk over a substantial period must be gauged, but much less critical for decision-making or management models where the user may be satisfied with knowing that the predicted ranking order of impacts of alternative scenarios or management options is likely to be correct, with only a rough indication of their sizes" (Jakeman et al., 2006).
- ☐ Briefly describe how you will summarise the results of these in silico experiments with graphical, tabular, or other devices, such as summary statistics.
- ☐ If the chosen modelling approach is able to explicitly articulate uncertainty due to data, measurements or baseline conditions, such as by providing estimates of uncertainty (typically in the form of probabilistic parameter covariance, Jakeman et al., 2006), specify which measure of uncertainty you will use.

5.3.2 Sensitivity analyses

Explanation

Sensitivity analysis examines how uncertainty in model outputs can be apportioned to different sources of uncertainty in model input (Saltelli et al., 2019).

Preregistration Item

- ☐ Describe the sensitivity analysis approach you will take: deterministic sensitivity, stochastic sensitivity (variability in the model), or scenario sensitivity (effect of changes based on scenarios).
- ☐ Describe any sensitivity analyses you will conduct by specifying which parameters will be held constant, which will be varied, and the range and intervals of values over which those parameters will be varied.
- ☐ State the primary objective of each sensitivity analysis, for example, to identify which input variables contribute the most to model uncertainty so that these variables can be targeted for further data collection, or alternatively to identify which variables or factors contribute little to overall model outputs, and so can be 'dropped' from future iterations of the model (Saltelli et al., 2019).

5.3.3 Model application or scenario analysis

Preregistration Item

- ☐ Specify any input conditions and relevant parameter values for initial environmental conditions and decision-variables under each scenario specified in Section 1.
- ☐ Describe any other relevant technical details of model application, such as methods for how you will implement any simulations or model projections.
- ☐ What raw and transformed model outputs will you extract from the model simulations or projections, and how will you map, plot, or otherwise display and synthesise the results of scenario and model

analyses.

- ☐ Explain how you will analyse the outputs to answer your analytical objectives. For instance, describe any trade-off or robustness analyses you will undertake to help evaluate and choose between different alternatives in consultation with experts or decision-makers.

5.3.4 Other simulation experiments / robustness analyses

Preregistration Item

- ☐ Describe any other simulation experiments, robustness analyses or other analyses you will perform on the model, including any metrics and their criteria / thresholds for interpreting the results of the analysis.

References

Araújo, M., Anderson, R., Márcia Barbosa, A., Beale, C., Dormann, C., Early, R., Garcia, R., Guisan, A., Maiorano, L., Naimi, B., O'Hara, R., Zimmermann, N., & Rahbek, C. (2019). Standards for distribution models in biodiversity assessments. *Sci Adv*, *5*(1), eaat4858. https://doi.org/10.1126/sciadv.aat4858

Augusiak, J., Van den Brink, P. J., & Grimm, V. (2014). Merging validation and evaluation of ecological models to "evaludation": A review of terminology and a practical approach. *Ecological Modelling*, *280*, 117–128. https://doi.org/10.1016/j.ecolmodel.2013.11.009

Barnard, D. M., Germino, M. J., Pilliod, D. S., Arkle, R. S., Applestein, C., Davidson, B. E., & Fisk, M. R. (2019). Can't see the random forest for the decision trees: Selecting predictive models for restoration ecology. *Restoration Ecology*. https://doi.org/10.1111/rec.12938

Boets, P., Landuyt, D., Everaert, G., Broekx, S., & Goethals, P. L. M. (2015). *Evaluation and comparison of data-driven and knowledge-supported bayesian belief networks to assess the habitat suitability for alien macroinvertebrates*. *74*, 92–103. https://doi.org/10.1016/j.envsoft.2015.09.005

Boettiger, C. (2022). The forecast trap. *Ecol Lett*. https://doi.org/10.1111/ele.14024

Cartwright, S. J., Bowgen, K. M., Collop, C., Hyder, K., Nabe-Nielsen, J., Stafford, R., Stillman, R. A., Thorpe, R. B., & Sibly, R. M. (2016). Communicating complex ecological models to non-scientist end users. *Ecological Modelling*, *338*, 51–59. https://doi.org/10.1016/j.ecolmodel.2016.07.012

Conn, P. B., Johnson, D. S., Williams, P. J., Melin, S. R., & Hooten, M. B. (2018). A guide to bayesian model checking for ecologists. *Ecological Monographs*, *9*, 341–317. https://doi.org/10.1002/ecm.1314

Dwork, C., Feldman, V., Hardt, M., Pitassi, T., Reingold, O., & Roth, A. (2015). The reusable holdout: Preserving validity in adaptive data analysis. *Science*, *349*(6248), 636–638. https://doi.org/10.1126/science.aaa9375

Fraser, H., Rumpff, L., Yen, J. D. L., Robinson, D., & Wintle, B. A. (2017). Integrated models to support multiobjective ecological restoration decisions. *Conservation Biology*, *31*(6), 1418–1427. https://doi.org/10.1111/cobi.12939

[redacted] & [redacted] *EcoConsPreReg: A guide to adaptive preregistration for model-based research in ecology and conservation*. https://doi.org/10.5281/ZENODO.10807029

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

Grimm, V., & Berger, U. (2016). Structural realism, emergence, and predictions in next-generation ecological modelling: Synthesis from a special issue. *Ecological Modelling*, *326*, 177–187. https://doi.org/10.1016/j.ecolmodel.2016.01.001

Ivimey, E. R., Pick, J. L., Bairos, K. R., Culina, A., Gould, E., Grainger, M., Marshall, B. M., Moreau, D., Paquet, M., Royauté, R., Sánchez, A., Silva, I., & Windecker, S. M. (2023). Implementing code review in the scientific workflow: Insights from ecology and evolutionary biology. *Journal of Evolutionary Biology*. https://doi.org/https://doi.org/10.1111/jeb.14230

Jakeman, A. J., Letcher, R. A., & Norton, J. P. (2006). Ten iterative steps in development and evaluation of environmental models. *Environmental Modelling & Software*, *21*(5), 602–614. https://doi.org/10.1016/j.envsoft.2006.01.004

Liu, C. C., & Aitkin, M. (2008). Bayes factors: Prior sensitivity and model generalizability. *Journal of Mathematical Psychology*, *52*(6), 362–375. https://doi.org/10.1016/j.jmp.2008.03.002

Liu, Z., Peng, C., Work, T., Candau, J.-N., DesRochers, A., & Kneeshaw, D. (2018). Application of machine-learning methods in forest ecology: Recent progress and future challenges. *Environmental Reviews*, *26*(4), 339–350. https://doi.org/https://doi.org/10.1139/er-2018-0034

Mahmoud, M., Liu, Y., Hartmann, H., Stewart, S., Wagener, T., Semmens, D., Stewart, R., Gupta, H., Dominguez, D., Dominguez, F., Hulse, D., Letcher, R., Rashleigh, B., Smith, C., Street, R., Ticehurst, J., Twery, M., Delden, H. van, Waldick, R., … Winter, L. (2009). A formal framework for scenario development in support of environmental decision-making. *Environmental Modelling & Software*, *24*(7), 798–808. https://doi.org/10.1016/j.envsoft.2008.11.010

McCarthy, M. A., Thompson, C. J., Moore, A. L., & Possingham, H. P. (2011). Designing nature reserves in the face of uncertainty. *Ecology Letters*, *14*(5), 470–475. https://doi.org/10.1111/j.1461-0248.2011.01608.x

McDonald-Madden, E., Baxter, P. W. J., & Possingham, H. P. (2008). Making robust decisions for conservation with restricted money and knowledge. *Journal of Applied Ecology*, *45*(6), 1630–1638. https://doi.org/10.1111/j.1365-2664.2008.01553.x

Moallemi, E. A., Elsawah, S., & Ryan, M. J. (2019). Strengthening "good" modelling practices in robust decision support: A reporting guideline for combining multiple model-based methods. *Mathematics and Computers in Simulation*. https://doi.org/10.1016/j.matcom.2019.05.002

Moon, K., Guerrero, A. M., Adams, Vanessa. M., Biggs, D., Blackman, D. A., Craven, L., Dickinson, H., & Ross, H. (2019). Mental models for conservation research and practice. *Conservation Letters*, *12*(3), e12642. https://doi.org/10.1111/conl.12642

Saltelli, A., Aleksankina, K., Becker, W., Fennell, P., Ferretti, F., Holst, N., Li, S., & Wu, Q. (2019). Why so many published sensitivity analyses are false: A systematic review of sensitivity analysis practices. *Environmental Modelling & Software*, *114*, 29–39. https://doi.org/10.1016/j.envsoft.2019.01.012

White, C. R., & Marshall, D. J. (2019). Should we care if models are phenomenological or mechanistic. *Trends in Ecology & Evolution*, *34*(4), 276–278. https://doi.org/10.1016/j.tree.2019.01.006

Yates, K., Bouchet, P., Caley, M., Mengersen, K., Randin, C., Parnell, S., Fielding, A., Bamford, A., Ban, S., Barbosa, A., Dormann, C., Elith, J., Embling, C., Ervin, G., Fisher, R., Gould, S., Graf, R., Gregr, E., Halpin, P., … Sequeira, A. (2018). Outstanding challenges in the transferability of ecological models. *Trends Ecol. Evol. (Amst.)*, *33*(10), 790–802. https://doi.org/10.1016/j.tree.2018.08.001

Preregistration for Vegetation Responses to Environmental Flows

[redacted]

2024-03-25

Table of contents

1 Problem Formulation

Rationale & Explanation

This section specifies the decision-making context in which the model will be used or the intended scope and context of conclusions. Important components include the decision maker and stakeholders (including experts) and their view on: i) the nature of the problem or decision addressed and how the scope of the modelling tool fits within the (broader) context (i.e. model purpose; ii) the spatial and temporal scales relevant to the decision context; iii) specified desired outputs; iv) role and inclusion in model development and testing; v) whether they foresee unacceptable outcomes that need to be represented in the model (i.e. as constraints), and; vi) what future scenarios does the model need to account for (noting this may be revised later). It should also provide a summary of the domain of applicability of the model, and reasonable extrapolation limits (Grimm et al., 2014).

1.1 Model Context and Purpose

Rationale & Explanation

Defining the purpose of the model is critical because the model purpose influences choices at later stages of model development (Jakeman et al., 2006). Common model purposes in ecology include: gaining a better qualitative understanding of the target system, synthesising and reviewing knowledge, and providing guidance for management and decision-making (Jakeman et al., 2006). Note that modelling objectives are distinct from the analytical objectives of the model.
The scope of the model includes temporal and spatial resolutions, which should also be defined here (Mahmoud et al., 2009). Any external limitations on model development, analysis and flexibility should also be outlined in this section (Jakeman et al., 2006).

1.1.1 Key stakeholders and model users

Preregistration Item

Identify relevant interest groups:
- ☑ Who is the model for?
- ☑ Who is involved in formulating the model?
- ☑ How will key stakeholders be involved in model development?
- ☑ Describe the decision-making context in which the model will be used (if relevant).

This preregistration document relates to the data analysis of a study on vegetation responses to environmental flows. The study forms one component of a larger set of studies within Stage 6 of the Victorian Environmental Flows Monitoring and Assessment Program (VEFMAP) managed by DELWP and delivered through the Arthur Rylah Institute (ARI). The study is funded by the state of Victoria, with a hierarchy of clients ranging from the Premier, to the Water Minister, Dep. Secretary, and program managers within the DELWP Water and Catchments division. The project is also co-funded through the Murray Darling Basin Plan, with corresponding requirements to contribute findings to Basin Plan reporting outputs. Other key stakeholders for this work are the Victorian Environmental Water Holder (VEWH) who manage the environmental water entitlements throughout the state, and the Victorian Catchment Management Authorities (CMAs) that manage the delivery of environmental water along individual waterways. Findings from this research can then be used to directly influence: management decisions through the CMAs; water allocation through the VEWH; and water investment through DELWP, the State Government and the Murray Darling Basin Authority. Other stakeholders, including the public, water authorities and institutions such as Parks Victoria (among many others) have a minor active role within the study.

The model formulation is developed by the researchers at ARI, with guidance from the program managers at DELWP, external researchers and an Independent Review Panel of external researchers for VEFMAP. The general model structure was proposed by the ARI researchers, which was initially interrogated, modified and validated with the program manager and external input. A more formal development of the model formulation was conducted in a collaboration between ARI researchers and external researchers with specific statistical expertise.

The model outputs, including recommendations, will be shared with decision makers (funders and managers) to guide their respective decisions. Decisions will require multiple inputs, such as the outcomes of this study and others, regulatory frameworks, funding availability, physical and practical constraints, and others. This study does not contribute to those decisions outside of providing the study outputs.

The three primary users of the model outputs are: 1) funders of environmental water (i.e. state and federal governments) to understand the magnitude of the outcomes of their investment, and the limitations or barriers to benefits; 2) managers of environmental water (i.e. CMAs) to guide and improve management decisions for environmental benefits; and 3) other researchers working in the field of waterway flow management and environmental flows locally and internationally.

1.1.2 Model purpose, context and problem context

Preregistration Item

Briefly outline:

- ☑ the ecological problem,
- ☑ the decision problem (if relevant), including the decision-trigger and any regulatory frameworks relevant to the problem,
- ☑ how the model will address the problem, being clear about the scope of the model i.e. is the model addressing the whole problem, or part of it? Are there any linked problems that your model should consider?
- ☑ Ensure that you specify any focal taxa and study objectives.

The ecological problem is related to river regulation. Natural flow regimes are major drivers of ecological processes within waterways. Natural flow regimes are altered by water storage and extraction, i.e. regulation. Environmental flows aim to provide flow components in a waterway to replace natural components that have been removed from the flow regime, specifically to provide ecological benefit to multiple taxa and physical processes. The decision problem relates to maximising the effectiveness of the environmental water deliveries to achieve their objectives. Decision triggers can occur for the funding of environmental water, i.e. whether the outcomes are worth the investment or if the investment needs to be altered. Decision triggers can also occur for the management of environmental water, i.e. whether the outcomes suggest changes to the current management process. For this study, the evidence is provided for others to address those decisions. There are various regulatory frameworks that are relevant to the funding and delivery of environmental water, but they are not within the scope of this assessment as they are in the hands of the decision makers.

The model itself within this study aims to provide clear evidence of the influence of environmental flows, and other factors (particularly exotic vegetation and livestock grazing), on native vegetation (cover and diversity) within regulated river channels. Outputs will aim to give practical evidence and the implications of responses to directly improve and guide the management of environmental water delivery for native vegetation benefit. Given that the model will incorporate data from across a number of waterways across a large part of Victoria, the model applications will speak to each of those waterways, as well as potential extrapolation to waterways not surveyed. The models will not aim to provide explicit predictions within the waterways surveyed, or to specific un-surveyed waterways, but extension of these models for prediction will be highlighted for future investigation beyond this study.

1.1.3 Analytical objectives

Explanation

How will the model be analysed, what analytical questions will the model be used to answer? For example, you might be using your model in a scenario analysis to determine which management decision is associated with minimum regret or the highest likelihood of improvement. Other examples from ecological decision-making include: to compare the performance of alternative management actions under budget constraint (Fraser et al., 2017), to search for robust decisions under uncertainty (McDonald-Madden et al., 2008), to choose the conservation policy that minimises uncertainty (McCarthy et al., 2011). See other examples in Moallemi et al. [(2019)].

Preregistration Item

Provide detail on the analytical purpose and scope of the model:

- ☑ How will the model be analysed and what analytical questions will the model be used to answer?
- ☐ Candidate decisions should be investigated and are specified a priori. Depending on the modelling context, they may be specified by stakeholders, model users or the analyst (Moallemi et al., 2019).
  - ☐ Describe the method used to identify relevant management actions and
  - ☐ specify management actions to be considered included in the model.
  - ☐ Are there potentially unacceptable management or policy outcomes identified by stakeholders that should be captured in the model, i.e. as constraints?
- ☑ Are there scenarios that model inputs or outputs that must accommodated? Scenarios should be set a priori, (i.e. before the model is built, Moallemi et al., 2019) and may be stakeholder-defined or

driven by the judgement of the modeller or other experts (Mahmoud et al., 2009).

- ☐ If relevant, describe what processes you will use to elicit and identify relevant scenarios, e.g. literature review, structured workshops with stakeholders or decision-makers.
- ☐ Specify scenarios under which decisions are investigated.

The model objectives for this study are to provide evidence for vegetation responses to environmental flow delivery over short (<1yr) to medium (<5yr) term with consideration of additional significant factors, such as exotic vegetation cover and livestock grazing. These objectives can be captured in the following questions: what is the short-term vegetation response to a single environmental flow event? what is the vegetation response to repeated flow events that have been delivered in recent years? how much is the native vegetation response to environmental flows limited by the abundance of exotic vegetation or livestock grazing?

The models will aim to quantify vegetation responses (indicated by the level or change in plant cover or diversity) to environmental flows and to quantify the effect of exotic vegetation cover or livestock grazing presence on these responses. Vegetation cover data that will be the response variables to models (and exotic vegetation covariates) are collected on the individual species level but will likely be grouped into relevant response classes. These data also enable evaluation of vegetation diversity within relevant response classes. The specific groupings are yet to be determined. More information on these groupings is provided below in Section 2.4.

The models themselves will be analysed by assessing the model fit and parameters indicating suitability of model structure given the data. The specific tests to be used will depend on the model structure used. More information is provided in Section 3.2 and Section 2.3. Future study will investigate the predictive capacity of the models within and between different waterways to enable transferability of the data, but this is beyond the scope of the current piece of work and is not described further here.

Future research will involve testing scenarios of different flow regimes to what we have observed. This includes estimating the expected data we would have collected if the flows that were delivered were not delivered - thus effectively modelling a control or counterfactual dataset to the observed data. It also includes testing various future scenarios with different hypothetical future regimes in the medium (<5yr) or long (>5yr) term. However, again this is likely to be beyond the scope of the current study which will focus solely on the collected data.

1.1.4 Logistical Constraints

Preregistration Item

- ☑ What degree of flexibility is required from the model? Might the model need to be quickly reconfigured to explore new scenarios or problems proposed by clients / managers / model-users?
- ☑ Are there any limitations on model development analysis and flexibility, such as time or budget constraints, for example, does a model need to be deployed rapidly?
  - ☑ When must the model be completed by, e.g. to help make a decision?

The models will require a certain amount of flexibility to be able to change the time frames over which data are compared and the set of predictor variables that will change depending on the time frames selected. Funding constraints and client objectives may also require changes to the modelling process or purpose at any point. Currently there are no significant limitations on model development - sufficient time and funding resources - but this may change.

1.1.5 Model Scope, Scale and Resolution

Preregistration Item

- ☐ The choice of a model's boundaries is closely linked to the choice of how finely to aggregate the behaviour within the model (Jakeman et al., 2006)
  - what is the intended scale, and resolution of the model (temporal, spatial or otherwise)?
- ☐ Where is the boundary of the modelled system? Everything outside beyond the boundary and not crossing it is to be ignored within the domain of the model, and everything crossing the boundary is to be treated as external forcing (known/unknown), or else as model outputs (observed, or not, Jakeman et al., 2006).

Each of the models described in this preregistration is aggregated differently depending on the response variable and depending on the particular model specification.

Vegetation Cover Models

- Full Model: Models vegetation cover for any given transect
- Simplified Model 1: "flow regime model"
- Simplified Model 2: "flow events model"

Species Richness Models

1.1.6 Intended application of results

> **Explanation**
>
> Preregistration Items in this section are relevant to model transferability (Yates et al., 2018) and constraints on generality in model analysis interpretation. How far do can the results be extrapolated based on the study design (data + model + analysis)? For instance, if there are many confounding variables and not enough spatial / environmental replication, then making broader more general claims beyond the stated boundaries of the model (Section 1.1.3) may not be warranted. However, larger generalisations about results may be acceptable if the data comes from experimentally manipulated or controlled systems.

> **Preregistration Item**
>
> ☑ What is the intended domain in which the model is to be applied? Are there any reasonable extrapolation limits beyond which you expect the model should not be applied (Grimm et al., 2014)?

The models developed in this study will be applied to only the river systems for which data was collected and used to fit these models. However, improved understanding about the ecological responses to environmental flows for the modelled systems may inform future conceptual models and statistical models based thereon.

Given that the primary goal of this study is to improve understanding about the effects of environmental flows on vegetation, these models resulting from this preregistration should not be directly used to inform predictions on which to base decisions about the management of environmental flows. The models developed in this study will instead be appropriately refined for use within a predictive modelling context.

2 Define Conceptual Model

> **Explanation**
>
> Conceptual models underpin the formal or quantitative model (Cartwright et al., 2016). The conceptual model describes the biological mechanisms relevant to the ecological problem and should capture basic premises about how the target system works, including any prior knowledge and assumptions about system processes. Conceptual models may be represented in a variety of formats, such as influence diagrams, linguistic model block diagram or bond graphs, and these illustrate how model drivers are linked to both outputs or observed responses, and internal (state) variables (Jakeman et al., 2006).

2.1 Choose elicitation and representation method

> **Preregistration Item**
>
> ☑ Describe what method you will use to elicit or identify the conceptual model. Some common methods include interviews, drawings, and mapping techniques including influence diagrams, cognitive maps and Bayesian belief networks (Moon et al., 2019). It is difficult to decide and justify which method is most appropriate, see Moon et al. (2019) for guidance addressing this methodological question.
>
> ☑ Finally, how do you intend on representing the final conceptual model? This will likely depend on the method chosen to elicit the conceptual model.

We have developed a series of relevant conceptual models for this work over the past four years that build on previous published work by other researchers, as well as our own research, observations and many discussions. These conceptual models are described in text within program reports and are summarised in diagrams. The models are biological/ecological only and do not include links to decision makers or values as these constraints are not within our capability to influence. Therefore, the models primarily summarise the dominant drivers of vegetation attributes (cover, diversity, distribution, composition) within waterways. Not all important drivers are included because it is impossible to account for everything within our study. Our final model for this particular study will be described in text and summarised within one or more non-quantitative diagrams.

2.2 Explain Critical Conceptual Design Decisions

Preregistration Item

List and explain critical conceptual design decisions (Grimm et al., 2014), including:

- ☑ spatial and temporal scales,
- ☑ selection of entities and processes,
- ☑ representation of stochasticity and heterogeneity,
- ☑ consideration of local versus global interactions, environmental drivers, etc.
- ☑ Explain and justify the influence of particular theories, concepts, or earlier models against alternative conceptual design decisions that might lead to alternative model structures.

This current study will focus on two different time frames: short term (months) in relation to before and after event surveys; as well as long-term (years) in relation to patterns resulting from previous years of particular flow regimes. The models therefore need to describe the short term responses of plants to events as well as the cumulative responses of plants to multiple types of events within years repeated over several years, i.e. regimes.

The entities will be the vegetation response variables (cover, diversity, composition and perhaps distribution) as well as various flow variables and other factors such as bank elevation, site, river system, rainfall and livestock grazing. The flow variables will depend on the time frame. Short term responses will include the presence and magnitude of a flow event and potentially the time of year and duration of the event. Long term responses will include the presence of individual flow events over the previous years (number of years TBC) and categorisation of the types of events that occurred, e.g. was a winter/spring flow natural or e-flow? Was it small or large? Categories TBC.

The primary process included is the inundation of plants by river flow. Rainfall to provide water will be included as a covariate rather than a process. Livestock grazing will also be evaluated as a damaging process through various mechanisms (plant consumption, trampling, soil compaction, nutrient addition, pugging, etc.) but these mechanisms are difficult to untangle so will most likely be combined if we can't do so adequately.

There is a large amount of heterogeneity in vegetation patterns within and between sites as well as stochasticity in the responses to events (and regimes). While both will play a large role in the modelling process of this study, neither is substantially addressed in the conceptual models for responses. One key component that has been considered though is the issue of habitat unsuitability for some plants in some areas, for example, there are many locations on a bank (such as very steep banks) where it is very difficult for some or any plants to occur, so these sites will have no or low cover or diversity regardless of flows. Because of this, we may consider models that evaluate change in cover only where plants occur at a sample in at least one survey - i.e. excluding sample locations with zeros through the whole dataset. This is defensible because locations where occupancy is impossible are not effective for evaluating flow responses, however, the frequency and distribution of these 'zero' samples would need to be described as well.

There are clear local and global interactions within our study. Local interactions are dominated by the species abundance and composition of plants within a sample area (plant interactions). Rainfall may also interact with flow events and would influence sites or groups of sites separately. Global interactions include the effect of season (time of year) on the responses, which is largely influenced by day length and temperatures.

2.3 Model assumptions and uncertainties

Preregistration Item

Specify key assumptions and uncertainties underlying the model design, describing how uncertainty and variation will be represented in the model (Moallemi et al., 2019). Sources of uncertainty may include:

- ☑ exogenous uncertainties affecting the system,
- ☑ parametric uncertainty in input data and
- ☑ structural / conceptual nonparametric uncertainty in the model.

Exogenous uncertainty

By 'exogenous uncertainties' I'm assuming this refers to uncertainties in any of the possible predictor variables for the model that are not direct treatment variables, even those that may not be used. There are a large number of potential exogenous uncertainties affecting the system because the surveys are done in the natural environment. These include:

- The spatial variability of rainfall. For rainfall to be accessible to a plant it needs to fall on or very near the plant in most cases. However, rainfall is spatially patchy and a rain event will affect different sites in an area differently. So our rain gauge data will only be an approximation of the actual rainfall at a specific site.
- Climate. Aside from rainfall, there are many climate elements that will influence plants, such as temperature, humidity and solar radiation (light). We can provide estimated values for some of these at sites but this will not account for the spatial and temporal variation in these factors that may influence our results.
- Soil properties. We do not have accurate (or any in some cases) information on soil type at each site. Nor do we have an expectation of how soils might influence plant attributes and responses to flows. We may be able to obtain basic information from spatial mapping of soil types, but this would be at a very coarse scale and would also not account for shallow surface soil deposits or variation through the soil profile. We have some data on soil types at some sites from soil cores but not all.
- Funghi. Funghi may be considered exogenous or endogenous, but it may have a role in determining plant patterns and responses to flows. We have no data on this, nor any specific expectations on how funghi may influence results.

Parametric uncertainty

I think this is referring to the other predictor variables in the model that are directly associated with treatments. These are primarily associated with flow but also grazing and potentially exotic plant cover.

- Flow elevation zones (if used). There is uncertainty in our GPS measurements of elevation of sub-transects and quadrats. There is also uncertainty in the flow elevation through time. Also, the flow elevation peak may only last for a brief period so the duration affected by the peak height may be very short. This means that the flow elevation zones that combine all of this information are uncertain. Even if we do not use zones, each of the input data listed here will be included and will be uncertain as indicated.
- Flow magnitude and duration. Flow magnitude (discharge) is recorded in most regulated streams at set gauges. The data from these gauges is usually calibrated by the data manager at some point after the raw data are entered. We have little idea of how much uncertainty there is in the uncorrected or corrected data that we obtain from online resources. Also, it is assumed that the magnitude (and timing and duration of flows) at the gauge is representative of the sites near that gauge, but we are unsure how much that may vary spatially.
- Flow velocity. Flow velocity can have a significant impact on plants if it is fast enough to damage or remove plants, or if it alters the soil. However the calculation of velocity at any given point on a river is difficult and varies at all points throughout the river cross section depending on the flow magnitude, flow height, channel form and obstacles. We do not have the capacity to calculate or even broadly approximate velocities at different flows so this variable cannot be included. However, we have seen very little field evidence of plant damage or removal from regulated flows on waterways and velocities during most regulated flows are typically not high on the bank where the plants are.
- Livestock grazing is specified as present or absent only at each site, not the intensity or duration of grazing. This adds a lot of uncertainty to this variable because the intensity (density of animals) and duration are important factors for the impact. However, this information is difficult to obtain from landholders or estimate from the site. We do have the animal ID though (cattle or sheep). Also, grazing impacts are spatially patchy depending on the ease of access - livestock don't impact steep slopes or inaccessible ledges or obstacles much compared to flatter or gentler slopes. So within a grazed site, the grazing impact may vary between transects.

- We will consider models for native plant cover that include exotic plant cover as a predictor. The categorisation of native and exotic is fairly well recognised in Victoria and should not pose an issue, but the categorisation of the plant group is variable (see Section 2.5.2). Also, the estimate of cover is uncertain.

Structural non-parametric uncertainty

Model uncertainty is a given. The models we will use will not fit the data perfectly but we will be able to quantify this uncertainty in our model estimates. Uncertainty will be captured for our vegetation input data based on the number of samples we have within our hierarchical data structure. Uncertainty will not be accounted for in our uncertain variables such as rainfall or flow where we have no way of estimating this uncertainty. Unknown uncertainty will be somewhat captured in random effects but this will be an approximation only. More information about the model form and potential uncertainty is provided in Section 3 but is not contained within our conceptual model.

2.4 Identify predictor and response variables

> **ℹ Explanation**
>
> The identification and definition of primary model input variables should be driven by scenario definitions, and by the scope of the model described in the problem formulation phase (Mahmoud et al., 2009).

> **Preregistration Item**
>
> *Identify and define system variables and structures, referencing scenario definitions, and the scope of the model as described within problem formulation:*
>
> - ☑ What variables would support taking this action or making this decision?
> - ☑ What additional variables may interact with this system (things we can't control, but can hopefully measure)?
> - ☑ What variables have not been measured, but may interact with the system (often occurs in field or observational studies)?
> - ☑ What variables are index or surrogate measures of variables that we cannot or have not measured?
> - ☑ In what ways do we expect these variables to interact (model structures)?
> - ☑ Explain how any key concepts or terms within problem or decision-making contexts, such as regulatory terms, will be operationalised and defined in a biologically meaningful way to answer the research question appropriately?

Response Variables

The response variables will be associated with plant cover and diversity, which are the most commonly specified variables of interest within stated management objectives. These variables are also widely used within ecological studies and are directly transferable/translatable to many other studies. The specific variables used are likely to be:

- Plant cover (by species or species group) at a particular time
- Change in plant cover (by species or group) over a particular period of time relating to the occurrence of one or more flow events
- Plant diversity (in relevant species response groups) at a particular time
- Change in plant diversity (in relevant species response groups) over a particular period of time relating to the occurrence of one or more flow events

Predictor Variables

Based on our problem formulation, the expected predictor variables to use within our models are:

- Bank elevation (by zone or elevation measure GPS)
- Flow elevation, duration, time.
- Livestock grazing (likely presence or absence (but can provide approximate intensity category), potentially by grazer type sheep/cattle)
- Rainfall (potentially, but need to be careful about what period of rainfall data is relevant to the responses)
- Site variables (random effects for hierarchical structure: sub-transects/quadrats $<$ transects $<$ site $<$ system/basin)
- Exotic vegetation cover or diversity (for models with native vegetation as the response)

- Time/season (the period over which the event has occurred, for regimes this will be standardised to a set number of years so will not be included as time but would be categorised into regime type TBC)

Each of these variables is important for influencing decisions about the use of environmental water to achieve benefits for (native) riparian vegetation. Additional variables that may be useful but are difficult to obtain at the appropriate spatial resolution (such as soils and funghi) are unlikely to be added. We expect there to be interactions between flow and bank elevation, hence the need to incorporate the elevation in our models. There may be interactions between bank elevation and livestock impacts due to the softer ground at the bank margin being more susceptible to trampling. Exotic vegetation will also interact with bank elevation and flows. Rainfall in summer is likely to be more influential than rainfall in winter.

2.5 Define prior knowledge, data specification and evaluation

ℹ Explanation

This section specifies the plan for collecting, processing and preparing data available for parameterisation, determining model structure, and for scenario analysis. It also allows the researchers to disclose any prior interaction with the data.

2.5.1 Collate available data sources that could be used to parameterise or structure the model

Preregistration Item

For pre-existing data (delete as appropriate):

- ☑ Document the identity, quantity and provenance of any data that will be used to develop, identify and test the model.
- ☑ For each dataset, are the data open or publicly available?
- ☑ How can the data be accessed? Provide a link or contact as appropriate, indicating any restrictions on the use of data.
- ☐ Date of download, access, or expected timing offuture access.
- ☑ Describe the source of the data - what entity originally collected these data? (National Data Set, Private Organisational Data, Own Lab Collection, Other Lab Collection, External Contractor, Meta-Analysis, Expert Elicitation, Other).
- ☐ Codebook and meta-data. If a codebook or other meta-data are available, link to it here and / or upload the document(s).
- ☑ Prior work based on this dataset - Have you published / presented any previous work based on this dataset? Include any publications, conference presentations (papers, posters), or working papers (in-prep, unpublished, preprints) based on this dataset you have worked on.
- ☑ Unpublished Prior Research Activity - Describe any prior but unpublished research activity using these data. Be specific and transparent.
- ☑ Prior knowledge of the current dataset - Describe any prior knowledge of or interaction with the dataset before commencing this study. For example, have you read any reports or publications about these data?
- ☑ Describe how the data are arranged, in terms of replicates and covariates.

For this section, I have used the relevant points above from both of the options above given that we have collected the data ourselves.

- The data were collected over a four year period (2016-2020) as part of the Victorian Environmental Flows Monitoring and Assessment Program. The program outline and the methods are outlined in reports found at this website: https://www.ari.vic.gov.au/research/rivers-and-estuaries/assessing-benefits-of-water-for-the-environment
- Data were collected from 44 sites across Victoria with sampling year and number of surveys varying between sites, from 3 times in one 'water year' to 11 times in four 'water years'. Each site had 5-10 permanent transects established and each of those transects had a series of sub-transects at increasing bank elevations at which data were collected each survey. More detail on how the data were collected is available in the manual.
- The data are not yet publicly available but will be available from an online database at a later date as a requirement of publicly funded data. There will be standard conditions of use for the data respecting IP and research contributions.

- A metadata file has not yet been produced for all of the relevant data files, but the column names have been intuitively labelled. A file will be produced and stored with the data or in an accessible repository. *Task/issue to create metadata file*
- There have been many unpublished client reports produced based on small subsets of these data but with very limited data analysis. There have also been many oral presentations and various summary documents (flyers etc.) shared with relevant stakeholders. A summary including some preliminary analysis of these data is provided in this published client report, available at the above website (Tonkin et al., 2020). There are other published papers in development based on this program but not on this dataset.
- Existing knowledge of the data comes from collecting it ourselves, conducting thorough data checking and cleaning, producing output summaries (species lists, sample sizes, etc.) and conducting a preliminary analysis on a small subset of the data for the report highlighted above and a separate study on one time period from one river: Sutton, N., Houghton, J., Vietz, G., Jones, C., Mole, B., Morris, K., Gower, T. 2020. Influence of Intervalley Transfers (IVT) on the Riverbanks and Bank Vegetation of the Goulburn and Campaspe Rivers. Report by Streamology and Arthur Rylah Institute for the Department of Environment, Land, Water and Planning. June, 2020.

2.5.2 Data Processing and Preparation

Preregistration Item

☑ Describe any data preparation and processing steps, including manipulation of environmental layers (e.g. standardisation and geographic projection) or variable construction (e.g. Principal Component Analysis).

This analysis is relatively complex, due to the large dataset and the integration of different data sources (vegetation data, flow, transect elevation, site attributes). Each of these data sources requires some careful data processing and preparation for use within the models. Here we outline the major actions - listed within issue #23 in our repository.

GPS data correlation The specific relevance of this issue for flow elevation is detailed below, but here we summarise the general issues with GPS correlation. High-accuracy GPS points are required for three parts of this study: 1) vegetation sampling locations, 2) flow elevation (data loggers), 3) bank profiles and soil moisture loggers. In this study, there were various sources of these GPS points, with each source not corresponding exactly to the others. A calibration is then used to align all points at a site for each source. The data sources include:

- Points for flow loggers on Campaspe: Unimelb
- Points for veg survey locations on Campaspe: ARI using UPG equipment
- Points for all other veg survey locations: ARI using ARI equipment (equiv to UPG)
- Zero grade heights for permanent water loggers: Ventia

In all cases, we will calibrate to the levels of the vegetation survey locations, so that the flow data and vegetation data align.

Flow data correlation with site elevation Flow data in this study was processed externally by a hydrology consultant with data and site information provided by the ARI team. For each site surveyed, flow data were compiled and an output of flow level (elevation in AHD) was produced at regular intervals (interval time depending on data inputs). While all vegetation survey sites were attempted to have matching flow data, some data were impossible to acquire due to the lack of critical flow information at the specific site and we are unable to evaluate full flow hydrology levels for these sites. The full hydrology data processing procedure is outlined in a separate document provided by the consultant and is not replicated here. It is critically important for our study that the flow elevations correlate with measured elevations of each vegetation sub-transect. The sub-transect elevations were obtained by ARI staff at the exact locations of the sub-transects, using GPS devices and processing software that was tested and proven sufficiently accurate for the study. While the relative elevation among sub-transect points appears to be accurate, the comparative elevation to the flow levels was in some cases misaligned. This is however a relatively easy fix, once the difference between the transect and flow data elevations is determined, that value is simply added or subtracted from either dataset for calibration. The differences were possible due to water level elevations recorded at each site that could compare flow levels with sub-transect levels directly. A second issue is that the river channels typically decline in elevation from upstream to downstream within a site, resulting in a decline in elevation across the transect locations within a site. This is particularly true for sites further upstream, whereas lowland reaches can be very flat. This means that a single flow level for each site will be inaccurately indicating flow heights at some transect locations. Again, this can be addressed through data calibration for each site so that the flow level speaks to each transect location within a site. This transect-specific calibration is required for all transects within the study with flow data available. This is then a three step process: 1) determine any discrepancy between sub-transect elevations and flow elevations for each site, 2) determine the within-site variation in elevation across transects, 3) provide a calibration value for all transects at all sites with flow data to enable rapid calibration.

Determination of flow events/timelines for evaluation Once accurate and fully aligned data are produced via the processes indicated above, it is important to carefully consider the flow variables to be used within models. Our conceptual understanding of vegetation responses to flows suggests that the depth, duration and timing (season) are important attributes of flow that dictate responses. These variables can be extracted from the flow data that we have for each site.

Important considerations for flow depth The most important aspect of flow depth is where or not a plant is fully submerged or not by the flow. Our VEFMAP Stage 6 experimental research has shown us that a plant that is only partly submerged is likely to be much less affected than a plant fully submerged. This is the case for most non-aquatic species. Therefore, when we consider the impacts of flow level, the flow elevation is very important, for example, the threshold of plant responses my not occur at the maximum elevation of a flow event but instead at around 40cm below the peak level where most plants will have been fully submerged. However, this impact may vary between species based on their height, growth form, and physiological responses to inundation.

Important considerations for flow duration Flow duration is important because vegetation responses vary dramatically with duration. Our VEFMAP Stage 6 experimental research has shown us that most plants are tolerant of short periods of inundation (and there may be some positive effects at very short durations depending on climate attributes and soil). But as submergence duration increases, most non-aquatic plants will decline in health until they eventually reach a survival threshold. The specific duration of a non-lethal flow may not be critical in evaluating the effect of a flow, but that is critical information for a lethal flow.

Important considerations for flow timing Flow timing is important because vegetation responses vary with flow timing. Our VEFMAP Stage 6 experimental research has shown us that many plants are more negatively affected (more rapidly) by submergence in warmer seasons. Separating winter/spring flow events from summer/autumn events appears to be the most important distinction from this point of view and we don't expect vegetation responses to be equivalent in these periods.

Important considerations for flow regimes When considering flow regimes, we firstly will need to determine a time range that is relevant, e.g. the last 3 years, 5 years,...prior to surveys. We will also probably need to categorise or otherwise quantify waterways with particular regimes to indicate the set of regime components that will influence vegetation patterns. For example, a 'full' regime might have baseflows and/or low-flows as well as at least one spring fresh and summer/autumn fresh. A 'summer' regime might have all but the spring fresh/high-flow, while a 'spring' regime might have all but the summer/autumn fresh. There may also be a category for 'variable' regime within recent years. For categorical regimes, the regime also needs to be consistent across the relevant set of years. Depending on the variability between sites and between years, we might need to consider multiple options for evaluating the regime. Each of the sites needs to be assessed to list the regime within recent years to determine the best options - which has not been done yet.

Determining the variables to use Initially, we will use the most simple flow variables that directly relate to management of e-flows. The most commonly delivered e-flows are baseflows and spring freshes. These events are usually consistently delivered each year at the same flow magnitude. This allows us to evaluate the cumulative effect of those flow levels on vegetation, as well as the effects of a single flow event. This can be done by determining the mean baseflow level and mean peak fresh level from our flow data and using those levels as indicators to compare vegetation trends above and below those elevations. This will be our first point of evaluation for flow. As we progress the analysis and determine how closely or loosely vegetation patterns align with these levels, we will explore alternative approaches and variables for use in the evaluation, such as the number of days inundated by a certain depth over a given period. Importantly, the usefulness of these evaluations depends a lot on the vegetation species or groupings that are evaluated in the models against these flow metrics.

Vegetation categorisation into functional/response groups Different species respond to the environment differently. The same is true for flow responses. This is why all of our vegetation surveys are initially done at the individual species level. However, individual species distributions and abundances are spatially and temporally patchy, so there can be great value in aggregating data from similar species to produce larger data pools. This is the norm for many ecological studies. The critical step is how species are aggregated. There is a balance between too refined groups (with little benefit from increased data pools as the pools increase by small amounts) and too broad groups (there is a large amount of species variation within a group meaning that responses are unclear). One well-established grouping system for wetland plants is by Brock and Casanova (1997) that groups plants by their affinity for different water regimes - called wetland plant functional groups (WPFG). In many cases, the species and the responses are directly equivalent to riparian systems, however there are some notable variations. There are also many different WPFGs, which are impractical to evaluate separately in many cases. In this study, we do not have a pre-determined approach for how vegetation groups will be used. We have compiled all of the relevant WPFG assignments for each species recorded and have determined our own classes in which these groups are nested (with variation required in few cases where the riparian response is expected or observed to be different from the wetland responses). We will initially conduct our analyses using the three hierarchical levels of data aggregation: species < WPFG < broad classes. Based on these initial explorations, we will progress the final evaluation approach. This exploratory approach is important for a study like this where there is not a clear precedent in the published literature for the most appropriate vegetation groupings.

Determination of grazing covariate The most simple form of grazing covariate is a binary score of present or absent. However, as described in Section 2.3, there is a lot of variation in the effect of grazing relating to the intensity (density of animals) and the timing (season of grazing). While we do not have this information, we can roughly categorise the grazing intensity based on site observations, which may be informative. Additionally, we have data on grazer animal (sheep or cattle) which may be important for some or all questions. We need to consider these options and develop candidate variables to test in our models.

Determination of exotic vegetation covariate The exotic vegetation covariate is relatively straightforward in one sense because it will be simple cover and/or diversity estimates that we have collected. However, there is a likely interaction between the impact of exotic species and the terrestrial/riparian grouping, for example, it is possible that terrestrial exotics are less of a problem than riparian ones even with the same amount of cover because they occupy the same habitat and seasonal niches as native riparian species. So this comes back to the vegetation groupings described above which need to be resolved to determine the variables used.

2.5.3 Describe any data exploration or preliminary data analyses

> **Explanation**
>
> In most modelling cases, it is necessary to perform preliminary analyses to understand the data and check that assumptions and requirements of the chosen modelling procedures are met. Data exploration prior to model fitting or development may include exploratory analyses to check for collinearity, spatial and temporal coverage, quality and resolution, outliers, or the need for transformations (Yates et al., 2018).

> **Preregistration Item**
>
> *For each separate preliminary or investigatory analysis:*
>
> - ☑ State what needs to be known to proceed with further decision-making about the modelling procedure, and why the analysis is necessary.
> - ☑ Explain how you will implement this analysis, as well as any techniques you will use to summarise and explore your data.
> - ☑ What method will you use to represent this analysis (graphical, tabular, or otherwise, describe)
> - ☑ Specify exactly which parts of the data will be used
> - ☑ Describe how the results will be interpreted, listing each potential analytic decision, as well as the analysis finding that will trigger each decision, where possible.

Given the complex nature of this analysis and the uncertainties in the data structure due to the data processing that needs to occur, there are many preliminary checks that may need to be conducted. For example, depending on the vegetation groupings that we decide to use, and how we account for 'zero samples' (see Section 2.2) there may be very different amounts of skew or zero-inflation in the response or predictor variables. Initially, we will need to do simple tests of data distributions within hierarchies to check for data spread and prevalence of zeros or outliers. We will also need to assess the collinearity or interactions between candidate predictor variables to ensure relevant interactions are captured and to reduce model overfitting. Our data are spatially and temporally variable with a relatively large spatial coverage but the data density is not equal across space and time, i.e. some sites have been surveyed more than others and in different years. Random effects for site should account for the spatial bias but we will need to carefully consider the interpretation of the outputs and explore options for accounting for temporal variability.

We will conduct a pilot analysis of a subset of the full data-set in order to develop an initial set of candidate models. This pilot data will consist of both hydrological and vegetation data for a single System, *Campaspe*. Although the last three years of data are missing from the current Campaspe dataset, it still contains observations from multiple years which will facilitate setting up a multi-year model. The pilot analysis aims to address uncertainties in how best to specify the candidate models, including how some variables should be operationalised, and which variables should be included given issues like multicolinearity. The analysis aims to identify any major issues that we might need to address within our modelling of the full dataset.

Some questions we will investigate are:

- How do we classify best flow regime based on inundation data?
- What is the relevant regime time-frame to consider for quantifying the regime? 3 / 5 years, something else?
- Are there any outliers?
- What distributional assumptions do we need to make in our models?
- Will there be issues with zero-inflated count data?
- Do the available data support the desired model structures, or do these require simplification?

To resolve these uncertainties we will conduct some exploratory data analyses. Specifically, we will visually assess distributions of the response variables, test correlations between all pairs of numerical (non-categorical) predictor variables, and calculate counts of non-zero observations within each category of any categorical variables included in the analysis (accounted for nested structures and interactions).

We will also fit two initial models to the pilot dataset, one for each key response variable (richness, cover):

```r
library(glmmTMB)

# autoregressive model for cover
cover_ar_model <- glmmTMB(
  # specify an autoregressive model structure to model change in plant cover
  plant_hits ~ log_plant_hits_tm1 +
    # assess group- and origin-specific impacts of broad flow "regimes"
    wpfg * origin * (days_above_baseflow + days_above_springfresh) +
    # fixed effects for functional group, zone (bank elev.), period (before/after),
    # and origin (with interactions)
    wpfg * zone * period * origin +
    # fixed effect for grazing impacts (binary variable)
    grazing +
    (1 | site / transect) +    # random effects for transects nested within sites
    (1 | metres) +             # random effect for location of site up the streambank
    (1 | survey_year) +        # random effect for survey year
    family = poisson,          # assume count distribution of plant_hits
    ziformula = ~ wpfg,        # allow zero-inflation, with functional group-specific parameters
    data = veg_cover_ar
)

# model for species richness
richness_model <- glmmTMB(
  species_richness ~
    # assess group- and origin-specific impacts of broad flow "regimes"
    wpfg * origin * (days_above_baseflow + days_above_springfresh) +
    # fixed effects for functional group, zone (bank elev.), period (before/after),
    # and origin (with interactions)
    wpfg * zone * period * origin +
    # fixed effect for grazing impacts (binary variable)
    grazing +
    (1 | site / transect) +    # random effects for transects nested within sites
    (1 | metres) +             # random effect for location of site up the streambank
    (1 | survey_year) +        # random effect for survey year
    offset(npoint),            # offset to account for number of points measured at each transect
    family = poisson,          # assume count distribution of plant_hits
    ziformula = ~ wpfg,        # allow zero-inflation, with functional group-specific parameters
    data = veg_richness
)
```

All models fitted to the pilot data set will undergo the same model checks as the candidate models fitted to the full dataset. The primary focus of model checks for the pilot analysis is to identify appropriate model structures, which has two main steps. First, assessing whether model converged and generated reliable parameter estimates, which in the case of the proposed `glmmTMB` models is assessed internally and printed on model return. Second, assessing whether the model structure (particularly the error distribution) is appropriate for the data, which is supported by posterior predictive checks. The details of these steps are provided in Section 4.3.1.

It is important to note that models fitted to the pilot dataset may not translate directly to the final analysis due to differences in the data structure. These differences will change the distribution of observations within categories, and will introduce a new random effect for water body (or system), which is not required for the single-water body pilot analysis. Due to these changes, the final analysis still requires model checking and may require changes to the model structure.

2.5.4 Data evaluation, exclusion and missing data

Explanation

Documenting issues with reliability is important because data quality and ecological relevance might be constrained by measurement error, inappropriate experimental design, and heterogeneity and variability inherent in ecological systems (Grimm et al., 2014). Ideally, model input data should be internally consistent across temporal and spatial scales and resolutions, and appropriate to the problem at hand (Mahmoud et al., 2009).

Preregistration Item

- ☑ Describe how you will determine how reliable the data are for the given model purpose. Ideally, model input data should be internally consistent across temporal and spatial scales and resolutions, and appropriate to the problem at hand
- ☑ Document any issues with data reliability.
- ☑ How will you determine what data, if any, will be excluded from your analyses?
- ☑ How will outliers be handled? Describe rules for identifying outlier data, and for excluding a site, transect, quadrat, year or season, species, trait, etc.
- ☑ How will you identify and deal with incomplete or missing data?

Data reliability At this stage I am unsure about how we will quantitatively evaluate how reliable the data are for the questions being addressed. Qualitatively, one of the most important considerations will be to assess how logical the outcomes are in relation to our observations and conceptual understanding. We have a very good mental model of these systems due to the large amounts of fieldwork in various areas over the years collecting the dataset used in this study as well as relevant data for other studies. As for data quality, there are likely to be minor errors in such a large dataset that we cannot detect with our quality checking procedures, however, we are very confident that the data have been consistently collected (almost every survey was led by the same individual with the other surveyors being from a consistent set of experienced individuals), in the same places (all transects were permanently marked and relocated at each survey), and the data checking process has been extremely thorough (documented in git). All this considered, we believe that this is one of the most reliable vegetation datasets available anywhere in Victoria for its size and complexity.

Data exclusions Data will only be excluded if we believe it is incorrect, and it can't be corrected, it is irrelevant to a particular model/assessment, or it precludes model fitting. For example, if evaluating survey intervals relating to spring fresh delivery, we can only include years and sites where such a delivery actually occurred. Incorrect data are easily detected where values lie outside possible or plausible ranges, but in other cases they can be very difficult to identify - we are confident that the former have been well accounted for in the datasets but the latter is only partially accounted for and there may be some minor errors that we cannot isolate. All species will be included, but unknown species that have no possible grouping identifier, e.g. native/exotic or life form, may not be possible to include. The vast majority of these occurrences are for seedlings that are too small to be identified. In most cases seedlings had minimal impact on plant cover, so this would have few implications for evaluation of cover, but this may have a greater impact on species richness. Decisions will need to be made for certain unknown species categories, particularly those that are more common in the dataset. At this stage only species that can reliably assigned to a relevant group for a given model will be included. Furthermore, the pilot analysis illustrated extreme zero-inflation for several groups, which prevented model fitting. Consequently these groups were excluded from the pilot study modelling and they may or may not be included in the full dataset analysis, depending on whether the models can be successfully fit.

Plant functional groups "Atl_native", "Ate_native", "Tda_unknown" are removed for all models because these wpfg are thought to be incorrectly specified (typos).

Missing data There are a small number of cases where we have incomplete or missing data due to various circumstances in the surveys. Currently these are indicated as NA in the dataset and form a small proportion of the overall dataset.

2.6 Conceptual model evaluation

Preregistration Item

- ☑ Describe how your conceptual model will be critically evaluated. Evaluation includes both the completeness and suitability of the overall model structure.
- ☑ How will you critically assess any simplifying assumptions (Augusiak et al., 2014)?
- ☐ Will this process include consultation or feedback from a client, manager, or model user?

Model analysis results will be used to update the existing conceptual model (described in Section 2). Specifically, model results will be used to inform understanding about the short-term (flow models) and long-term (regime models) impacts of environmental flows combined with grazing and exotic vegetation on native vegetation (cover and density) within regulated river channels monitored in this dataset.

3 Formalise and Specify Model

Explanation

In this section describe what quantitative methods you will use to build the model/s, explain how they are relevant to the client/manager/user's purpose.

3.1 Model class, modelling framework and approach

Explanation

Modelling approaches can be described as occurring on a spectrum from correlative or phenomenological to mechanistic or process-based (Yates et al., 2018); where correlative models use mathematical functions fitted to data to describe underlying processes, and mechanistic models explicitly represent processes and details of component parts of a biological system that are expected to give rise to the data (White & Marshall, 2019). A model 'class,' 'family' or 'type' is often used to describe a set of models each of which has a distinct but related sampling distribution (C. C. Liu & Aitkin, 2008). The model family is driven by choices about the types of variables covered and the nature of their treatment, as well as structural features of the model, such as link functions, spatial and temporal scales of processes and their interactions (Jakeman et al., 2006).

Preregistration Item

- ☑ Describe what modelling framework, approach or class of model you will use to implement your model and relate your choice to the model purpose and analytical objectives described in Section 1.1.2 and Section 1.1.3.

We will use a correlative model approach for this analysis, where we will attempt to detect and describe patterns in recorded vegetation data in relation to a series of covariates. Specifically, the analysis aims to provide clear evidence of the influence of environmental flows, and other factors (particularly exotic vegetation and livestock grazing), on native vegetation (cover and diversity) within regulated river channels (Section 1.1.3). We will initially use generalised linear mixed effects models (GLMM) with zero-inflated Poisson family distribution for cover data and species richness (counts). The exact model family will be determined based on an iterative process of model fitting and model checking (with posterior predictive checks) to ensure that the fitted model family is appropriate for the data being modelled (see Section 3.2, below). We believe that GLMMs are a robust approach for our objectives and data structure, with a key feature of the random effects allowing for the hierarchical sampling design and repeated measures.

Although alternative distributions may be suitable for cover data (e.g., binomial, beta), these distributions introduce link functions that complicate the specification of an autoregressive model structure. For this reason, a Poisson distribution was used to model cover data, noting that the Poisson approximates the binomial in the limit of a large number of trials. The Poisson distribution, specified with an offset, still models proportional cover but does not constrain values to sit below an upper bound (values must still be non-negative).

3.2 Choose model features and family

Explanation

All modelling approaches require the selection of model features, which conform with the conceptual model and data specified in previous steps (Jakeman et al., 2006). The choice of model are determined in conjunction with features are selected. Model features include elements such as the functional form of interactions, data structures, measures used to specify links, any bins or discretisation of continuous variables. It is usually difficult to change fundamental features of a model beyond an early stage of model development, so careful thought and planning here is useful to the modeller (Jakeman et al., 2006). However, if changes to these fundamental aspects of the model do need to change, document how and why these choices were made, including any results used to support any changes in the model.

3.2.1 Operationalising Model Variables

Preregistration Item

- ☑ For each response, predictor, and covariate, specify how these variables will be operationalised in the model. This should relate directly to the analytical and/or management objectives specified during the problem formulation phase. Operationalisations could include: the extent of a response, an extreme value, a trend, a long-term mean, a probability distribution, a spatial pattern, a time-series, qualitative change, such as a direction of change or, the frequency, location, or probability of some event occurring. Specify any treatment of model variables, including whether they are lumped / distributed, linear / non-linear, stochastic / deterministic (Jakeman et al., 2006).
- ☑ Provide a rationale for your choices, including why plausible alternatives under consideration were not chosen, and relate your justification back to the purpose, objectives, prior knowledge and or logistical constraints specified in the problem formulation phase (Jakeman et al., 2006).

The two primary response variables are plant cover and species richness (see Section 2.4), which will be operationalised in slightly different ways depending on the time frame:

1. short term-in response to a single event (i.e. before and after); and
2. medium term (2-10 years) in relation to typical flow regimes (flow elevation/duration/timing) over that period.

Both will be using the extent of a response of particular groups of plants at particular bank elevations in relation to flow events (factor condition of before or after an specific event) or flow values (e.g. days of flow to a specific elevation within a year/season), as well as the additional covariates of exotic plant cover and livestock grazing. This will enable an evaluation of variation in plant responses at different elevations in relation to the primary management action (flow) as well as identify the relative impacts of flow and the covariates of exotic plants (interacting with flow) and livestock grazing.

3.2.2 Choose model family

Preregistration Item

- ☑ Specify which family of statistical distributions you will use in your model, and describe any transformations, or link functions.
- ☑ Include in your rational for selection, detail about which variables the model outputs are likely sensitive to, what aspects of their behaviour are important, and any associated spatial or temporal

dimensions in sampling.

Posterior predictive checks from the pilot analysis indicated high levels of zero-inflation with some over-dispersion. The final models in the pilot analysis allowed for zero-inflation parameters to differ among plant functional groups (`ziformula = ~ wpfg`) to account for different proportions of zeros among functional groupings, but did not account for over-dispersion. Over-dispersion is commonly accounted for by using negative binomial models, which we could potentially fit using groups-specific dispersion parameters using the `glmmTMB` argument `dispformula = ~ wpfg`. However, all attempts to fit negative binomial models resulted in non-convergence in our pilot analysis. Despite some degree of over-dispersion, given that the zero-inflated Poisson models converged and reliably captured the proportion of zeros in the pilot dataset, we propose using this approach for all models fitted to the full dataset, including any simplified models.

The two key vegetation response variables have Poisson and binomial families (distributions) as indicated in Section 3.3, but will both be modelled as zero-inflated Poisson distributions for the reasons outlined in Section 3.1. All models will use a log link function. These distributions have been selected based on the type of data and the expected data distributions, particularly the fact that observations of richness and cover are both recorded as counts but include many zero values. There may be unexpected issues with these approaches due to actual data distributions not matching our expectations, such as an unaccounted for over-dispersion of the data.

Should either over-dispersion or zero-inflation be identified in any fitted model, we will try alternative distributions.

Should over dispersion in any model be identified, we will re-fit the models using a negative binomial distribution. However, the models re-fitted using the negative binomial distribution will only be accepted over the Poisson models if they *both* improve over-dispersion *and* do not decline in model performance, particularly model fit, zero-inflation, and posterior predictive checks.

While we are unable to *a priori* precisely weight these criteria in determining the final distribution, the final decision will be guided by the model's overall ability to capture key associations reliably. This is especially likely to occur when there is no 'perfect model' and there is no dominant alternative choice of model distribution.

3.3 Describe *approach* for identifying model structure

Explanation

This section relates to the process of determining the best/most efficient/parsimonious representation of the system at the appropriate scale of concern (Jakeman et al., 2006) that best meets the analytical objectives specified in the problem formulation phase. Model structure refers to the choice of variables included in the model, and the nature of the relationship among those variables. Approaches to finding model structure and parameters may be knowledge-supported, or data-driven (Boets et al., 2015). Model selection methods can include traditional inferential approaches such as unconstrained searches of a dataset for patterns that explain variations in the response variable, or use of ensemble-modelling methods (Barnard et al., 2019). Ensemble modelling procedures might aim to derive a single model, or a multi-model average (Yates et al., 2018). Refining actions to develop a model could include iteratively dropping parameters or adding them, or aggregating / disaggregating system descriptors, such as dimensionality and processes (Jakeman et al., 2006).

Preregistration Item

- ☑ Specify what approach and methods you will use to identify model structure and parameters.
- ☑ If using a knowledge-supported approach to deriving model structure (either in whole or in part), specify model structural features, including:
  - the functional form of interactions (if any)
  - data structures,
  - measures used to specify links,
  - any bins or discretisation of continuous variables (Jakeman et al., 2006),
  - any other relevant features of the model structure.

Structure estimation: model structure specification is knowledge-driven rather than data-driven, with further refinement and simplification guided by the results of the pilot analysis (Section 2.5.3).

Interactions are expected within the proposed models, such as those between elevation and flow, as well as flow, livestock grazing and weeds. For some model options, flow data may not be required, and elevation may be used as a surrogate for flow (e.g. x elevation represents x flow). In this case, a three-way interaction between elevation, grazing and weeds may be used to evaluate vegetation responses. Event-based (short term) models evaluating changes before and after an event would require a fourth interacting term for period (before or after event). Isolating the effects of elevation, flow or period could then be done using post-hoc tests. Data structures are broadly defined in Section 2.2 and Section 2.4 for the different variables proposed for the study. Plant richness data are counts, cover data are hits (successes) from points (trials, where n=40 for all sub-transects) (modelled as *Poisson*, see Section 3.1), flow data may take a range of forms indicated earlier, elevation is provided in *mAHD* but is likely to be input into the model as an ordinal categorical factor with bins based on known elevation of flow events (e.g. freshes), grazing at this stage may be a binomial indicator of presence or absence, or alternatively as a categorical factor indicating the presence/absence of cattle or sheep.

The model structure will be refined through the pilot analysis (described in Section 2.5.3.), one aspect of which will be to assess model convergence with different levels of interactions (model convergence checks are described in Section 4.3.1). If models do not converge with the above-specified interactions, higher order interactions will be progressively removed from the model (i.e., three way interactions will be removed, then two-way interactions).

The following terms will be included as random effects: transects nested within sites, point location (metres up the streambank), and survey year.

3.4 Describe parameter estimation technique and performance criteria

> **Explanation**
>
> Before calibrating the model to the data, the performance criteria for judging the calibration (or model fit) are specified. These criteria and their underlying assumptions should reflect the desired properties of the parameter estimates / structure (Jakeman et al., 2006). For example, modellers might seek parameter estimates that are robust to outliers, unbiased, and yield appropriate predictive performance. Modellers will need to consider whether the assumptions of the estimation technique yielding those desired properties are suited to the problem at hand. For integrated or sub-divided models, other considerations might include choices about where to disaggregate the model for parameter estimation; e.g. spatial sectioning (streams into reaches) and temporal sectioning (piece-wise linear models) (Jakeman et al., 2006).

3.4.1 Parameter estimation technique

> **Preregistration Item**
>
> - ☑ Specify what technique you will use to estimate parameter values, and how you will supply non-parametric variables and/or data (e.g. distributed boundary conditions). For example, will you calibrate all variables simultaneously by optimising fit of model outputs to observations, or will you parameterise the model in a piecemeal fashion by either direct measurement, inference from secondary data, or some combination (Jakeman et al., 2006).
> - ☑ Identify which variables will be parameterised directly, such as by expert elicitation or prior knowledge.
> - ☑ Specify which algorithm(s) you will use for any data-driven parameter estimation, including supervised, or unsupervised machine learning, decision-tree, K-nearest neighbour or cluster algorithms (Z. Liu et al., 2018).

Initially, our model structure will be based on our conceptual model of the system/response that has also guided our data collection. In this case, we do not have a very large set of potential environmental variables to select from that may or may not be influential. Instead, we have selected variables and monitoring approaches to collect data that reflect our expectations of correlation/causation. Given this, there is unlikely to be a large need for a model selection process where variables are added or removed sequentially to evaluate relative model performance. However, there will be critical evaluation of all model components to determine their value to the model, such as the hierarchical nature of the data and the need (or not) to include all levels of the hierarchy in the final models. The model selection process is an important part of the study in itself and will be described in the resulting paper, particularly where it relates to relative effects of different variables and relative importance of spatial and temporal scales. Additionally, there will be some investigation of the effectiveness/appropriateness of different forms of certain variables, such as: bank elevation (continuous or categorical), vegetation groupings, and flow variables (Section 2.4). The primary focus of the model is to evaluate the relative impacts of the key variables in line with the conceptual model and hypotheses, so model performance will align with that objective, including model fit parameters of residual plots, model uncertainty, unexplained variation, degrees of freedom, and fit statistics. Evaluating the model performance for predictive capacity within and beyond the dataset may not be required in the current study as it is likely beyond the scope, but this would involve testing predictive capacity within and between waterways using e.g. a cross-validation approach.

Parameterisation: Parameter estimation will be data-driven, and implemented with the `glmmTMB` R package (Brooks et al., 2017), which uses Template Model Builder software to fit flexible GLMM-type models, amongst other model types. `glmmTMB` uses maximum likelihood to estimate parameters for both parametric and non-parametric factor data (nominal or ordinal) variables (e.g. categorised elevation and/or flow data, as well as binary or categorical grazing variables). We will optimise the model by comparing fit of model outputs to observations (residuals versus fitted values).

3.4.2 Parameter estimation / model fit performance criteria

> Preregistration Item
>
> - ☑ Specify which suite of performance criteria you will use to judge the performance of the model. Examples include correlation scores, coefficient of determination, specificity, sensitivity, AUC, etcetera (Yates et al., 2018).
> - ☑ Relate any underlying assumptions of each criterion to the desired properties of the model, and justify the choice of performance metric in relation
> - ☑ Explain how you will identify which model features or components are significant or meaningful.

R^2, a measure of agreement between fitted and observed values, is the primary performance criterion we will use to evaluate the performance of each model, coupled with posterior predictive checks to assist the suitability of normal structures (Section 4) . When comparing among alternative models, models with a higher R^2 will be preferred over those with lower R^2. As an approximate guide for judging model fit, we will use the following thresholds R^2 in the absence of cross-validation:

- < 0.25 is poor,
- 0.25 - 0.5 is moderate,
- 0.5 - 0.75 is good,
- 0.75 - 1.0 is excellent, but probably indicates overfitting.

3.5 Model assumptions and uncertainties

> Preregistration Item
>
> - ☑ Specify assumptions and key uncertainties in the formal model. Describe what gaps exist between the model conception, and the real-world problem, what biases might this introduce and how might this impact any interpretation of the model outputs, and what implications are there for evaluating model-output to inform inferences or decisions?

Poisson and negative binomial models all assume linearity in model parameters, independence between individual observations, as well as the multiplicative effects of independent variables.

Zero-inflation

Based on the results of the pilot study (Section 2.5.3), all Poisson models allow for zero-inflation parameters to differ among plant functional groups (using `glmmTMB::` argument `ziformula=~ wpfg`).

Over-dispersion

Should over-dispersion (where the conditional variance of the outcome variable is greater than the conditional mean) be detected in models specified with a Poisson distribution, negative-binomial models will be fitted because they allow for over-dispersion by estimating the mean and variance independently (Kruppa & Hothorn, 2021) and assume that extra-Poisson variance is a quadratic function of the mean (Lindén & Mäntyniemi, 2011). Negative binomial models may allow for group-specific dispersion parameters using the `glmmTMB::` argument `dispformula = ~ wpfg`.

Issues of over-dispersion are not expected to be fully resolved by the model specifications outlined in Section 3, however proposed model specifications were a compromise between ideal specification and the limitations of our data.

Other Model Assumptions & Sources of Uncertainty

For each specified model we may fit, further assumptions are described below in Section 3.

A formal sensitivity analysis has not yet been completed and so we are unaware of what variables the model outputs are most sensitive to. The many potential forms of flow and/or elevation data is a key area of possible variation in the variable behaviour, which is currently unknown. There are also many spatial and temporal dimensions associated with the data, such as the hierarchical scales of sub-transect, transect, site, reach, waterway, basin, and State. There will be spatial autocorrelation within the data at each of these levels, such as northern waterways or river basins being more similar to southern counterparts due to climate and geomorphology differences. The extent of the variation among the different spatial scale is currently unknown and is an important aspect of the study to evaluate. Temporal patterns are also important due to the longitudinal nature of the data collection and different seasons of survey (i.e. different proximity to different flow/climate periods). Careful consideration of these aspects will be important in the model design, as per Section 2.4.

3.6 Specify formal model(s)

Explanation

Once critical decisions have been made about the modelling approach and method of model specification, the conceptual model is translated into the quantitative model.

Preregistration Item

- [x] Specify all formal models
  - [ ] Note, for data-driven approaches to determining model structure and/or parameterisation, it may not be possible to respond to this preregistration item. In such cases, explain why this is the case, and how you will document the model(s) used in the final analysis.
- [x] For quantitative model selection approaches, including ensemble modelling, specify each model used in the candidate set, including any null or full/global model.

Vegetation Cover Models

Based on the pilot analysis (Section 2.5.3) we have derived three additional model structures to be fitted on the full dataset that are simplified versions of the full models specified in Section 2.5.3. While these simplified models are not ideal, the full models with three-way interactions were too complex given the data, and failed to converge and generate reliable parameter estimates on the pilot dataset. Consequently, the full models will be fitted again to the full dataset.

Full Model, full dataset, GitHub Issue 64

We will attempt to fit a single full model using all data, and adding a random effect for `waterbody` to account for variation between river systems.

Should the full model fit to all data be computationally feasible and converge, we will add the three-way interaction `zone * period * wpfg` back into the full model specification (discarded in the pilot analysis). If this converges it will be used as the basis for final outputs, if not, the simplified models listed below will be used.

Should the full model fitted to the full dataset not be computationally feasible or converge, we will follow a similar strategy for the pilot-study modelling, working from the full model towards simplified models, as there is potential for the full dataset to support a model of intermediate complexity (more complex than the simplified models, but not as complex as the full model). We will fit the following simplified models in the case that the full model fitted to the full dataset does not converge:

Simplified model 1: "flow regime" model, GitHub Issue 65

The aim of this simplified model is to examine how past flow conditions influence vegetation cover while capturing the average effects of flows at different levels over multiple years. The model does not consider zone or period and identifies functional group-specific impacts of days above baseflow or spring fresh levels.

This model includes two flow predictors (`days_above_baseflow_std`, `days_above_spring_fresh_std`), and the full suite of random effects.

```
cover_ar_TMBmod_1 <- glmmTMB::glmmTMB(
   hits ~ log_hits_tm1 +
     days_above_baseflow_std * wpfg * origin +
     days_above_springfresh_std * wpfg * origin +
    # days_above_baseflow_std^2 + days_above_springfresh_std^2 +
     #   zone * period +
   #  zone + period +
   #  grazing + wpfg  +
     (1 | site / transect) +
     #(1 | site / period) +
     (1 | metres) +
     (1 | survey_year),
   # offset(npoint),
   family = poisson,
   ziformula=~ wpfg,
   #dispformula =~ wpfg ,
   data = veg_cover_ar_sum |>
     filter(!wpfg_ori %in% c("Atl_native",
                             "Ate_native",
                             "Tda_unknown"))
 )
```

Simplified model 2a: "flow events" Model, version 1, GitHub Issue 66

The aim of the 'flow events' models (2a,2b) is to examine how vegetation cover changes in specific zones before and after key flow events (spring and summer freshes). Model 2a seeks to examine the effects on vegetation before and after specific flow events assuming that different plant functional groups (`wpfg`s) have different cover levels but similar responses to flows.

In the pilot analysis, model results indicated that vegetation cover differed in its responses to spring and summer freshes in each zone (below baseflow, baseflow to spring fresh, above spring fresh). Below baseflow level, vegetation increased following the spring fresh and remained high following the summer event (Figure 5, Appendix S7). In the baseflow-to-spring fresh zone, vegetation increased following the spring fresh but returned to pre-spring levels following the summer fresh (Figure 5, Appendix S7). In the above-spring fresh zone, vegetation level declined following both the spring and the summer fresh (Figure 5, Appendix S7).

Consequently, this model includes several categorical predictors as independent fixed effects (`origin`, `wpfg`, `grazing`, and a zone-by-period interaction) and the full suite of random effects, but no `days_above_` predictors.)

This first version of the flow event model allows functional groups to have different levels of cover but assumes that changes following flow events are similar in all groups within a zone. Although this is not an ideal model structure, it is likely that functional groups are restricted to particular zones, in which case the zone-by-period interaction may capture some of the variation attributable to functional groupings.

```
cover_ar_TMBmod_2 <- glmmTMB::glmmTMB(
  hits ~ log_hits_tm1 +
    # days_above_baseflow_std*wpfg*origin +
    # days_above_springfresh_std*wpfg*origin +
    # days_above_baseflow_std^2 +
    # days_above_springfresh_std_sq +
    zone * period +
    origin + wpfg +
    grazing +
    (1 | site / transect) +
    #(1 | site / period) +
    (1 | metres) +
    (1 | survey_year),
  # offset(npoint),
  family = poisson,
  ziformula = ~ wpfg,
  # dispformula =~ wpfg ,
```

```
  data = veg_cover_ar_sum |>
    filter(!wpfg_ori %in% c("Atl_native", "Ate_native", "Tda_unknown"))
)
```

Simplified model 2b: "flow events" model, version 2, GitHub Issue 67

The aim is to examine how vegetation cover of each functional grouping changes before and after key flow events (spring and summer freshes). This second version (Simplified model 2b) allows functional groups to have different responses to flow events (spring and summer freshes) but assumes that vegetation in all zones changes similarly following each flow event. As for simplified model 2a, this model structure is not ideal, but provides a method to distinguish `wpfg`-specific responses to flow events.

This model includes several categorical predictors as independent fixed effects (`origin`, `zone`, `grazing`), a `wpfg`-by-`period` interaction, as well as the full suite of random effects, and no `days_above_` predictors.)

```
cover_ar_TMBmod_3 <- glmmTMB::glmmTMB(
  hits ~ log_hits_tm1 +
    # days_above_baseflow_std*wpfg*origin +
    # days_above_springfresh_std*wpfg*origin +
    # days_above_baseflow_std^2 +
    # days_above_springfresh_std_sq +
    zone + wpfg * period +
    origin +
    grazing +
    (1 | site / transect) +
    # (1 | site / period) +
    (1 | metres) +
    (1 | survey_year),
  # offset(npoint),
  family = poisson,
  ziformula = ~ wpfg,
  # dispformula =~ wpfg ,
  data = veg_cover_ar_sum |>
    filter(!wpfg_ori %in% c("Atl_native", "Ate_native", "Tda_unknown"))
)
```

Species Richness Models

Full Model, Full Dataset, GitHub Issue 68

Simplified Model 1: "Flow Regime" model, GitHub Issue 69

```
richness_ar_TMBmod_1 <- glmmTMB::glmmTMB(
  richness ~
    days_above_baseflow_std * wpfg * origin +
    days_above_springfresh_std * wpfg * origin +
    # days_above_baseflow_std^2 +
    # days_above_springfresh_std^2 +
    # zone * period +
    #zone *period + zone*wpfg + wpfg*period +
    # grazing + origin +
    (1 | site / transect) +
    #(1 | site / period) +
    (1 | metres) +
    (1 | survey_year),
  # offset(npoint),
  family = poisson,
  #family = nbinom2,
  #ziformula=~ wpfg,
  # dispformula =~ wpfg ,
  data = veg_richness |>
    filter(!wpfg_ori %in% c("Atl_native",
                            "Ate_native",
                            "Tda_unknown"))
)
```

Simplified Model 2: "Flow Events" model, GitHub Issue 70

Rather than splitting the flow events model into two different versions, as we did for the cover models, we have combined these into a single model for richness.

```
richness_ar_TMBmod_2 <- glmmTMB::glmmTMB(
  richness ~
    #days_above_baseflow_std +
    # days_above_springfresh_std +
    # days_above_baseflow_std^2 +
    # days_above_springfresh_std^2 +
    # zone * period +
    zone * period + zone * wpfg + wpfg * period +
    grazing + origin +
    (1 | site / transect) +
    # (1 | site / period) +
    (1 | metres) +
    (1 | survey_year),
  # offset(npoint),
  family = poisson,
  #family = nbinom2,
  #ziformula=~ wpfg,
  # dispformula =~ wpfg ,
  data = veg_richness |>
    filter(!wpfg_ori %in% c("Atl_native",
                            "Ate_native",
                            "Tda_unknown")) |>
    filter(!wpfg %in% c("Sk", "Se"))
)
```

4 Model Calibration, Validation & Checking

4.1 Model calibration and validation scheme

Explanation

This section pertains to any data calibration, validation or testing schemes that will be implemented. For example, the model may be tested on data independent of those used to parameterise the model (external validation), or the model may be cross-validated on random sub-samples of the data used to parameterise the model (Barnard et al., 2019; internal cross-validation Yates et al., 2018). For some types of models, hyper-parameters are estimated from data, and may be tuned on further independent holdouts of the training data ("validation data").

Preregistration Item

- [ ] Describe any data calibration, validation and testing scheme you will implement, including any procedures for tuning or estimating model hyper-parameters (if any).

4.1.1 Describe calibration/validation data

Explanation & Rationale

The following items pertain to properties of the *datasets* used for calibration (training), validation, and testing.

🔥 Preregistration Item

If using external / independent holdout data for model testing and evaluation (delete as needed):

- ☑ Which data will be used as the testing data? What method will you be used for generating training / test data subsets?
- ☑ Describe any known differences between the training/validation and testing datasets, the relative size of each, as well as any stratification methods used for ensuring evenness of groups between data sets.
- ☑ It is preferable that any independent data used for model testing remains unknown to modellers during the process of model development. Describe the relationship modellers have with model validation data. Will independent datasets be known or accessible to any modeller or analyst?

Due to the complexity of the analysis, and the focus on inference, no data partitioning / testing on external data will be used. However, we will use a subset of the data for exploratory pilot analysis that informs the final model specifications (Campaspe catchment only, Section 2.5.3).

4.2 Implementation verification

ℹ Explanation & Examples

Model implementation verification is the process of ensuring that the model has been correctly implemented, and that the model performs as described by the model description (Grimm et al., 2014). This process is distinct from model checking, which assesses the model's performance in representing the system of interest (Conn et al., 2018).

- Checks for verification implementation should include i) thoroughly checking for bugs or programming errors, and ii) whether the implemented model performs as described by the model description (Grimm et al., 2014).
- Qualitative tests could include syntax checking of code, and peer-code review (Ivimey et al., 2023). Technical measures include using unit tests, or in-built checks within functions to prevent potential errors.

🔥 Preregistration Item

- ☑ What Quality Assurance measures will you take to verify the model has been correctly implemented? Specifying a priori quality assurance tests for implementation verification may help to avoid selective debugging and silent errors.

Implementation verification will be assessed using a number of techniques, but will broadly follow the approach proposed by Ivimey-Cook *et al.* (2023):

1. Code will be reviewed periodically using the GitHub flow model where code is submitted for independent review by collaborators before being merged into the working copy of the code repository. Code will be assessed either by attempting to reproduce the code in the pull-request, or by visual inspection.
2. Defensive programming techniques, in-line error checking, functionalisation, modularisation and documentation of analysis code will be used as preventative measures to catch bugs and ensure proper code implementation (Ivimey et al., 2023),
3. Finally, the final analysis and results will be subjected to a more substantial peer-code review from collaborators further removed from code writing and analysis implementation (likely CJ or EG), following the 4R's (Ivimey et al., 2023).

4.3 Model checking

ℹ Rationale & Explanation

"Model Checking" goes by many names ("conditional verification", "quantitative verification", "model output verification"), and refers to a series of analyses that assess a model's performance in representing the system of interest (Conn et al., 2018). Model checking aids in diagnosing assumption violations, and reveals where a model might need to be altered to better represent the data, and therefore system (Conn et al., 2018). Quantitative model checking diagnostics include goodness of fit, tests on residuals or errors, such as for heteroscedascity, cross-correlation, and autocorrelation (Jakeman et al., 2006).

4.3.1 Quantitative model checking

Preregistration Item

During this process, observed data, or data and patterns that guided model design and calibration, are compared to model output in order to identify if and where there are any systematic differences.

- ☑ Specify any diagnostics or tests you will use during model checking to assess a model's performance in representing the system of interest.
- ☑ For each test, specify the criteria that will you use to interpret the outcome of the test in assessing the model's ability to sufficiently represent the gathered data used to develop and parameterise the model.

Posterior predictive checks will be conducted to assess the degree of zero-inflation and over-dispersion. Posterior checks compare the distribution of observed data against a distribution simulated from the fitted model and are assessed quantitatively or graphically (Conn et al., 2018). We will plot both distributions and visually compare the distributions: close alignment of the two distributions indicates that the specified model structure is appropriate, whereas deviations between the observed distribution and the model-generated distribution indicate potential assumption violations and mismatches between the data and model.

For graphical posterior checks, there is no threshold but any disagreement, particularly in key aspects of the model (zero inflation, and counts in the range of the majority of the data) would suggest an alternative distribution is worth considering. Often no distribution will be perfect, so it's a choice between two imperfect options. (in which case, we will take into consideration the balance between model balance simplicity, R^2, and the "key aspects" of model such as zero inflation and the bulk of the data).

Should over-dispersion and zero-inflation be present in the fitted models, alternative model families / structures (Section 3.2) may be investigated and compared to the primary models outlined in Section 2.5.3.

4.3.2 Qualitative model checking

Explanation

This step is largely informal and case-specific, but requires, 'face validation' with model users / clients / managers who aren't involved in the development of the model to assess whether the interactions and outcomes of the model are feasible an defensible (Grimm et al., 2014). This process is sometimes called a "laugh test" or a "pub test" and in addition to checking the model's believability, it builds the client's confidence in the model (Jakeman et al., 2006). Face validation could include structured walk-throughs, or presenting descriptions, visualisations or summaries of model results to experts for assessment.

Preregistration Item

- ☑ Briefly explain how you will qualitatively check the model, and whether and how you will include users and clients in the process.

In terms of qualitative assessment, model results will be checked by field experts Chris Jones & Lyndsey Vivian, who will assess model-estimated associations for their plausibility given their expert knowledge of the underlying target system.

4.3.3 Assumption Violation Checks

Preregistration Item

The consequences of assumption violations on the interpretation of results should be assessed (Araújo et al., 2019).

- ☐ Explain how you will demonstrate robustness to model assumptions and check for violations of model assumptions.
- ☐ If you cannot perform quantitative assumption checks, describe what theoretical justifications would justify a lack of violation of or robustness to model assumptions.
- ☐ If you cannot demonstrate or theoretically justify violation or robustness to assumptions, explain why not, and specify whether you will discuss assumption violations and their consequences for interpretation of model outputs.
- ☐ If assumption violations cannot be avoided, explain how you will explore the consequences of assumption violations on the interpretation of results. *This step is to be completed in interim iterations of the preregistration, only if there are departures from assumptions as demonstrated in the planned tests above.*

5 Model Validation and Evaluation

Explanation

The model validation & evaluation phase comprises a suite of analyses that collectively inform inferences about whether, and under what conditions, a model is suitable to meet its intended purpose (Augusiak et al., 2014). Errors in design and implementation of the model and their implication on the model output are assessed. Ideally independent data are used against the model outputs to assess whether the model output behaviour exhibits the required accuracy for the model's intended purpose. The outcomes of these analyses build confidence in the model applications and increase understanding of model strengths and limitations. Model evaluation, including model analysis, should complement model checking. It should evaluate model checking, and consider over-fitting and extrapolation. As the proportion of calibrated or uncertain parameters increases, so does the risk that the model seemingly works correctly, but for the wrong mechanistic reasons (Boettiger, 2022). Evaluation thus complements model checking because we can rule out the chance that the model fits the calibration data well, but has not captured the relevant ecological mechanisms of the system pertinent to the research question or the decision problem underpinning the model (Grimm et al., 2014). Evaluation of model outputs against external data in conjunction with the results from model checking provide information about the structural realism and therefore credibility of the model (Grimm & Berger, 2016).

5.1 Model output corroboration

Explanation

Ideally, model outputs or predictions are compared to independent data and patterns that were not used to develop, parameterise, or verify the model. Testing against a dataset of response and predictor variables that are spatially and/or temporally independent from the training dataset minimises the risk of artificially inflating model performance measures (Araújo et al., 2019). Although the corroboration of model outputs against an independent validation dataset is considered the 'gold standard' for showing that a model properly represents the internal organisation of the system, model validation is not always possible because empirical experiments are infeasible or model users are working on rapid-response time-frames, hence, why ecologists often model in the first place (Grimm et al., 2014). Independent predictions might instead be tested on sub-models. Alternatively, patterns in model output that are robust and seem characteristic of the system can be identified and evaluated in consultation with the literature or by experts to judge how accurate the model output is (Grimm et al., 2014).

Preregistration Item

- ☑ State whether you will corroborate the model outputs on external data, and document any independent validation data in this step.
- ☑ It is preferable that any independent data used for model evaluation remains unknown to modellers during the process of model building (Dwork et al., 2015), describe the relationship modellers have to model validation data, e.g. will independent datasets be known to any modeller or analyst involved in the model building process?
- ☑ If unable to evaluate the model outputs against independent data, explain why and explain what steps you will take to interrogate the model.

Model evaluation and validation on external data will not be undertaken in this study due to the complexity of the models (hierarchical structure combined with interactions) and data (potential for over-dispersion and zero-inflation) and also due to the focus of the analysis on inference.

For the same reasons, we will not partition the data for model validation on independent data using cross-validation or other data-partitioning approaches. Instead, in-sample model fit assessment will be conducted (Section 3.4.2).

Models may be validated on newly collected independent data in the future, but external validation is outside the scope of this study.

We have partitioned the data for pilot analysis, see Section 2.5.3 for details of this subsetting. Pilot analysis was performed blind to the full dataset, and was performed only on the pilot data subset.

5.2 Choose performance metrics and criteria

Explanation

Model performance can be quantified by a range of tests, including measures of agreement between predictions and independent observations, or estimates of accuracy, bias, calibration, discrimination refinement, resolution and skill (Araújo et al., 2019). Note that the performance metrics and criteria in this section are used for evaluating the structured and parameterised models (ideally) on independent holdout data, so this step is additional to any performance criteria used for determining model structure or parameterisation (Section 3.4.2).

Preregistration Item

- ☐ Specify what performance measures you will use to evaluate the model and briefly explain how each test relates to different desired properties of a model's performance.
- ☐ Spatial, temporal and environmental pattern of errors and variance can change the interpretation of model predictions and conservation decisions (Araújo et al., 2019). Where relevant and possible, describe how you will characterise and report the spatial, temporal and environmental pattern of errors and variance.
- ☐ If comparing alternative models, specify what measures of model comparison or out-of-sample performance metrics will you use to find support for alternative models or else to optimise predictive ability. State what numerical threshold or qualities you will use for each of these metrics.

5.3 Model analysis

Rationale & Explanation

Uncertainty in models arises due to incomplete system understanding (which processes to include, or which interact), from imprecise, finite and sparse data measurements, and from uncertainty in input conditions and scenarios for model simulations or runs (Jakeman et al., 2006). Non-technical uncertainties can also be introduced throughout the modelling process, such as uncertainties arising from issues in problem-framing, indeterminicies, and modeller / client values (Jakeman et al., 2006).
The purpose of model analysis is to prevent blind trust in the model by understanding how model outputs

have emerged, and to 'challenge' the model by verifying whether the model is still believable and fit for purpose if one or more parameters are changed (Grimm et al., 2014).
Model analysis should increase understanding of the model behaviour by identifying which processes and process interactions explain characteristic behaviours of the model system. Model analysis typically consists of sensitivity analyses preceded by uncertainty analyses (Saltelli et al., 2019), and a suite of other simulation or other computational experiments. The aim of such computational experiments is to increase understanding of the model behaviour by identifying which processes and process interactions explain characteristic behaviours of the model system (Grimm et al., 2014). Uncertainty analyses and sensitivity analyses augment one another to draw conclusions about model uncertainty.
Because the results from a full suite of sensitivity analysis and uncertainty analysis can be difficult to interpret due to the number and complexity of causal relations examined (Jakeman et al., 2006), it is useful for the analyst to relate the choice of analysis to the modelling context, purpose and analytical objectives defined in the problem formulation phase, in tandem with any critical uncertainties that have emerged during model development and testing prior to this point.

5.3.1 Uncertainty Analyses

Explanation

Uncertainty can arise from different modelling techniques, response data and predictor variables (Araújo et al., 2019). Uncertainty analyses characterise the uncertainty in model outputs, and identify how uncertainty in model parameters affects uncertainty in model output, but does not identify which model assumptions are driving this behaviour (Grimm et al., 2014; Saltelli et al., 2019). Uncertainty analyses can include propagating known uncertainties through the model, or by investigating the effect of different model scenarios with different parameters and modelling technique (Araújo et al., 2019), for example. It could also include characterising the output distribution, such as through empirical construction using model output data points. It could also include extracting summary statistics like the mean, median and variance from this distribution, and perhaps constructing confidence intervals on the mean

Preregistration Item

- ☐ Please describe how you will characterise model and data uncertainties, e.g. propagating known uncertainties through the model, investigating the effect of different model scenarios with different parameters and modelling technique combinations (Araújo et al., 2019), or empirically constructing model distributions from model output data points, and extracting summary statistics, including the mean, median, variance, and constructing confidence intervals (Saltelli et al., 2019).
- ☐ Relate your choice of analysis to the context and purposes of the model described in the problem formulation phase. For instance, discrepancies between model output and observed output may be important for forecasting models, where cost, benefit, an risk over a substantial period must be gauged, but much less critical for decision-making or management models where the user may be satisfied with knowing that the predicted ranking order of impacts of alternative scenarios or management options is likely to be correct, with only a rough indication of their sizes" (Jakeman et al., 2006).
- ☐ Briefly describe how you will summarise the results of these in silico experiments with graphical, tabular, or other devices, such as summary statistics.
- ☐ If the chosen modelling approach is able to explicitly articulate uncertainty due to data, measurements or baseline conditions, such as by providing estimates of uncertainty (typically in the form of probabilistic parameter covariance, Jakeman et al., 2006), specify which measure of uncertainty you will use.

Uncertainty analyses will not be conducted.

5.3.2 Sensitivity analyses

Explanation

Sensitivity analysis examines how uncertainty in model outputs can be apportioned to different sources of uncertainty in model input (Saltelli et al., 2019).

Preregistration Item

- ☐ Describe the sensitivity analysis approach you will take: deterministic sensitivity, stochastic sensitivity (variability in the model), or scenario sensitivity (effect of changes based on scenarios).
- ☐ Describe any sensitivity analyses you will conduct by specifying which parameters will be held constant, which will be varied, and the range and intervals of values over which those parameters will be varied.
- ☐ State the primary objective of each sensitivity analysis, for example, to identify which input variables contribute the most to model uncertainty so that these variables can be targeted for further data collection, or alternatively to identify which variables or factors contribute little to overall model outputs, and so can be 'dropped' from future iterations of the model (Saltelli et al., 2019).

No sensitivity analyses will be performed on any models developed in this study.

5.3.3 Model application or scenario analysis

Preregistration Item

- ☐ Specify any input conditions and relevant parameter values for initial environmental conditions and decision-variables under each scenario specified in Section 1.
- ☐ Describe any other relevant technical details of model application, such as methods for how you will implement any simulations or model projections.
- ☑ What raw and transformed model outputs will you extract from the model simulations or projections, and how will you map, plot, or otherwise display and synthesise the results of scenario and model analyses.
- ☐ Explain how you will analyse the outputs to answer your analytical objectives. For instance, describe any trade-off or robustness analyses you will undertake to help evaluate and choose between different alternatives in consultation with experts or decision-makers.

Model analysis

Model features will be identified using visual plots of estimates, i.e. forest plots of most coefficients, and marginal effects plots (plots of predicted values of y against values of x) under different combinations of of the key flow, zone, and period predictor variables.

Forest Plots: For each fitted model we will extract both fixed-effect and random-effect coefficient estimates and their 95% Confidence Intervals using `parameters::model_parameters()` (Lüdecke et al., 2020) or similar. Estimates will then be visualised as forest plots in R.

Predicted Values Plots: We will use the `effects::effect_plot()` function (Fox & Hong, 2009), or similar, to create marginal effects plots for each fitted model.

Scenario analysis will not be used in this study, but will be used in future predictive modelling studies based on the models developed in this study.

5.3.4 Other simulation experiments / robustness analyses

Preregistration Item

- ☐ Describe any other simulation experiments, robustness analyses or other analyses you will perform on the model, including any metrics and their criteria / thresholds for interpreting the results of the analysis.

No further simulation experiments, robustness or other analyses will be performed on the model other than analyses described above in this document.

References

Araújo, M., Anderson, R., Márcia Barbosa, A., Beale, C., Dormann, C., Early, R., Garcia, R., Guisan, A., Maiorano, L., Naimi, B., O'Hara, R., Zimmermann, N., & Rahbek, C. (2019). Standards for distribution models in biodiversity assessments. *Sci Adv*, *5*(1), eaat4858. https://doi.org/10.1126/sciadv.aat4858

Augusiak, J., Van den Brink, P. J., & Grimm, V. (2014). Merging validation and evaluation of ecological models to "evaludation": A review of terminology and a practical approach. *Ecological Modelling*, *280*, 117–128. https://doi.org/10.1016/j.ecolmodel.2013.11.009

Barnard, D. M., Germino, M. J., Pilliod, D. S., Arkle, R. S., Applestein, C., Davidson, B. E., & Fisk, M. R. (2019). Can't see the random forest for the decision trees: Selecting predictive models for restoration ecology. *Restoration Ecology*. https://doi.org/10.1111/rec.12938

Boets, P., Landuyt, D., Everaert, G., Broekx, S., & Goethals, P. L. M. (2015). *Evaluation and comparison of data-driven and knowledge-supported bayesian belief networks to assess the habitat suitability for alien macroinvertebrates*. *74*, 92–103. https://doi.org/10.1016/j.envsoft.2015.09.005

Boettiger, C. (2022). The forecast trap. *Ecol Lett*. https://doi.org/10.1111/ele.14024

Brock, M. A., & Casanova, M. T. (1997). *Frontiers in ecology; building the links* (N. Klomp & I. Lunt, Eds.; pp. 181–192). Elsevier Science.

Brooks, M. E., Kristensen, K., Benthem, J. ,. van, Koen, Magnusson, A., Berg, Casper,W., Nielsen, A., Skaug, Hans,J., Mächler, M., & Bolker, Benjamin,M. (2017). glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. *The R Journal*, *9*(2), 378. https://doi.org/10.32614/RJ-2017-066

Cartwright, S. J., Bowgen, K. M., Collop, C., Hyder, K., Nabe-Nielsen, J., Stafford, R., Stillman, R. A., Thorpe, R. B., & Sibly, R. M. (2016). Communicating complex ecological models to non-scientist end users. *Ecological Modelling*, *338*, 51–59. https://doi.org/10.1016/j.ecolmodel.2016.07.012

Conn, P. B., Johnson, D. S., Williams, P. J., Melin, S. R., & Hooten, M. B. (2018). A guide to bayesian model checking for ecologists. *Ecological Monographs*, *9*, 341–317. https://doi.org/10.1002/ecm.1314

Dwork, C., Feldman, V., Hardt, M., Pitassi, T., Reingold, O., & Roth, A. (2015). The reusable holdout: Preserving validity in adaptive data analysis. *Science*, *349*(6248), 636–638. https://doi.org/10.1126/science.aaa9375

Fox, J., & Hong, J. (2009). Effect displays in R for multinomial and proportional-odds logit models: Extensions to the effects package. *Journal of Statistical Software*, *32*(1), 1–24. https://doi.org/10.18637/jss.v032.i01

Fraser, H., Rumpff, L., Yen, J. D. L., Robinson, D., & Wintle, B. A. (2017). Integrated models to support multiobjective ecological restoration decisions. *Conservation Biology*, *31*(6), 1418–1427. https://doi.org/10.1111/cobi.12939

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

Grimm, V., & Berger, U. (2016). Structural realism, emergence, and predictions in next-generation ecological modelling: Synthesis from a special issue. *Ecological Modelling*, *326*, 177–187. https://doi.org/10.1016/j.ecolmodel.2016.01.001

Ivimey, E. R., Pick, J. L., Bairos, K. R., Culina, A., Gould, E., Grainger, M., Marshall, B. M., Moreau, D., Paquet, M., Royauté, R., Sánchez, A., Silva, I., & Windecker, S. M. (2023). Implementing code review in the scientific workflow: Insights from ecology and evolutionary biology. *Journal of Evolutionary Biology*. https://doi.org/https://doi.org/10.1111/jeb.14230

Jakeman, A. J., Letcher, R. A., & Norton, J. P. (2006). Ten iterative steps in development and evaluation of environmental models. *Environmental Modelling & Software*, *21*(5), 602–614. https://doi.org/10.1016/j.envsoft.2006.01.004

Kruppa, J., & Hothorn, L. (2021). A comparison study on modeling of clustered and overdispersed count data for multiple comparisons. *Journal of Applied Statistics*, *48*(16), 3220–3232. https://doi.org/10.1080/02664763.2020.1788518

Lindén, A., & Mäntyniemi, S. (2011). Using the negative binomial distribution to model overdispersion in ecological count data. *Ecology*, *92*(7), 1414–1421. https://doi.org/10.1890/10-1831.1

Liu, C. C., & Aitkin, M. (2008). Bayes factors: Prior sensitivity and model generalizability. *Journal of Mathematical Psychology*, *52*(6), 362–375. https://doi.org/10.1016/j.jmp.2008.03.002

Liu, Z., Peng, C., Work, T., Candau, J.-N., DesRochers, A., & Kneeshaw, D. (2018). Application of machine-learning methods in forest ecology: Recent progress and future challenges. *Environmental Reviews*, *26*(4), 339–350. https://doi.org/https://doi.org/10.1139/er-2018-0034

Lüdecke, D., Ben-Shachar, M., Patil, I., & Makowski, D. (2020). Extracting, computing and exploring the parameters of statistical models using r. *Journal of Open Source Software*, *5*(53), 2445. https://doi.org/10.21105/joss.02445

Mahmoud, M., Liu, Y., Hartmann, H., Stewart, S., Wagener, T., Semmens, D., Stewart, R., Gupta, H., Dominguez, D., Dominguez, F., Hulse, D., Letcher, R., Rashleigh, B., Smith, C., Street, R., Ticehurst, J., Twery, M., Delden, H. van, Waldick, R., … Winter, L. (2009). A formal framework for scenario development in support of environmental decision-making. *Environmental Modelling & Software*, *24*(7), 798–808. https://doi.org/10.1016/j.envsoft.2008.11.010

McCarthy, M. A., Thompson, C. J., Moore, A. L., & Possingham, H. P. (2011). Designing nature reserves in the face of uncertainty. *Ecology Letters*, *14*(5), 470–475. https://doi.org/10.1111/j.1461-0248.2011.01608.x

McDonald-Madden, E., Baxter, P. W. J., & Possingham, H. P. (2008). Making robust decisions for conservation with restricted money and knowledge. *Journal of Applied Ecology*, *45*(6), 1630–1638. https://doi.org/10.1111/j.1365-2664.2008.01553.x

Moallemi, E. A., Elsawah, S., & Ryan, M. J. (2019). Strengthening "good" modelling practices in robust decision support: A reporting guideline for combining multiple model-based methods. *Mathematics and Computers in Simulation*. https://doi.org/10.1016/j.matcom.2019.05.002

Moon, K., Guerrero, A. M., Adams, Vanessa. M., Biggs, D., Blackman, D. A., Craven, L., Dickinson, H., & Ross, H. (2019). Mental models for conservation research and practice. *Conservation Letters*, *12*(3), e12642. https://doi.org/10.1111/conl.12642

Saltelli, A., Aleksankina, K., Becker, W., Fennell, P., Ferretti, F., Holst, N., Li, S., & Wu, Q. (2019). Why so many published sensitivity analyses are false: A systematic review of sensitivity analysis practices. *Environmental Modelling & Software*, *114*, 29–39. https://doi.org/10.1016/j.envsoft.2019.01.012

Tonkin, Z., Jones, C., Clunie, P., Vivian, L., Amtstaetter, F., Jones, M., Koster, W., Mole, B., O'Connor, J., Brooks, J., Caffrey, L., & Lyon, J. (2020). *Victorian Environmental Flows Monitoring and Assessment Program*.

White, C. R., & Marshall, D. J. (2019). Should we care if models are phenomenological or mechanistic. *Trends in Ecology & Evolution*, *34*(4), 276–278. https://doi.org/10.1016/j.tree.2019.01.006

Yates, K., Bouchet, P., Caley, M., Mengersen, K., Randin, C., Parnell, S., Fielding, A., Bamford, A., Ban, S., Barbosa, A., Dormann, C., Elith, J., Embling, C., Ervin, G., Fisher, R., Gould, S., Graf, R., Gregr, E., Halpin, P., … Sequeira, A. (2018). Outstanding challenges in the transferability of ecological models. *Trends Ecol. Evol. (Amst.)*, *33*(10), 790–802. https://doi.org/10.1016/j.tree.2018.08.001

Appendix S7: Case Study Pilot Analysis Summary Report

Pilot analysis of VEFMAP riparian vegetation surveys

2024-01-24

Overview

A pilot study was conducted to identify model structures that were estimable (i.e., generated reliable parameter estimates) and captured key features of the data (e.g., many zeros). The details are outlined in the pre-registration document. In short, the process began with a preferred model structure and refined this by removing interactions until a suitable model was identified.

The preferred model structure was (in high-level, R notation):

```
Veg. response ~ wpfg * origin * flow predictors +
     wpfg * origin * zone * period +
     grazing +
     (1 | site / transect) +
     (1 | transect / metres) +
     (1 | survey_year) +
     (1 | species) +
     offset(npoints)
```

The model for vegetation cover included an additional term for the previous cover in the previous survey, which made this an autoregressive model (i.e., modelling the change in cover rather than absolute cover).

The flow predictors were simple metrics intended to capture broad features of the flow regime at a site. The two metrics were the number of days in which discharge was above baseflow levels, and the number of days in which discharge was above spring fresh levels. Values were calculated for each water year, and a 3-year rolling mean was used to represent broad changes in the multi-year flow regime (cf. variation within or among individual years). Quadratic transforms of these metrics were also included in the model.

Models for cover and species richness were fitted with zero-inflated Poisson distributions, with negative binomial distributions as an alternative to address over-dispersion. For cover, this differs from the (generally preferred) use of a bounded distribution (e.g., binomial or beta). However, parameters of bounded distributions are challenging to estimate when many values are near the bounds (lower or upper), and typically introduce link functions that do not allow straightforward specifications of autoregressive models. For this reason, the Poisson distribution was used in models of cover, noting that this approximates the binomial distribution in the limit of large numbers of trials.

Initial models aimed to include data at the level of species. This was infeasible for two reasons. First, data resolved to the level of species required an extremely large data set (due to the many species recorded), which made model estimation computationally challenging. Second, the large data set (with all species) comprised mostly zero observations, resulting in extreme zero-inflation that affected model convergence. The final models were fitted to cover summed over all species within a plant functional grouping, noting that these summed cover estimates sometimes (but rarely) exceeded 100% due to species overlap.

Key decisions and final model structures

1. High-level interactions are too challenging to estimate. Instead proposed three models for vegetation cover:
    a. A flow regime model that examines flow predictors only (with random effects).
    b. A flow event model (version 1) that examines zone-by-period interactions but assumes that all functional groups have similar patterns (within a zone and period).
    c. A flow event model (version 2) that examines functional group-by-period interactions but assumes that all zones respond similarly.
2. Species richness model is still being finalised, but should at least be able to encompass versions 1 and 2 of the cover model in a single analysis.
3. These models ask high-level questions about the effects of flows (and flow events) on vegetation cover and species richness and provide generalisable insight into how flows could be delivered to influence vegetation outcomes.
4. Quadratic flow effects too complex to fit: removed.
5. Metres within transect too complex to include: replaced with a random effect for metres, which assumes that the distance up the bank might influence vegetation outcomes but in similar ways among sites.
6. Zero-inflation specified as a function of plant functional group to allow for different proportions of zeros among groupings.
7. Over-dispersion present but not resolved by negative binomial model (even with wpfg-specific dispersion parameters): zero-inflated Poisson model retained, acknowledging lingering issues of over-dispersion.
8. Offset removed because all observations were based on the same number of points.

Decision points for full analysis

Several decisions will need to be made prior to the full analysis (which should begin ASAP).

1. Assess whether the full zone-by-period-by-wpfg interaction is possible with the full data set. If not, use simplified models outlined above.
2. Decide whether to fit a single model to all waterbodies or separate models for each (decision based on model convergence and posterior checks – a single model is preferable if feasible).
3. Decide on how to visualise model outputs.

Vegetation cover: regime model

This model includes two flow predictors (days above baseflow, days above spring fresh) and the full suite of random effects. The aim is to examine how past flow conditions influence vegetation cover. The model does not consider zone or period.

The model identifies functional group-specific impacts of days above baseflow or spring fresh levels. Interpretations here are based on the predicted associations (Figures 2 and 3) because interpreting forest plots of coefficients (Figure 1) is challenging in models with many interactions. Positive effects of baseflows are clearest for native ATl and ATe and exotic Tdr and Tda, all of which have higher cover under baseflows. Positive effects of spring freshes are clearest for native Sk, Tda, and Ate and exotic ATe. Negative effects of spring freshes are clearest for Tdr.

Model structure

```
# lets attempt to fit the hydrology model first -
cover_ar_TMBmod_1 <- glmmTMB::glmmTMB(
  hits ~ log_hits_tm1 +
    days_above_baseflow_std*wpfg*origin + days_above_springfresh_std*wpfg*origin +
   # days_above_baseflow_std^2 + days_above_springfresh_std^2 +
    #   zone * period +
  #  zone + period +
  #  grazing + wpfg  +
    (1 | site / transect) +
    #(1 | site / period) +
    (1 | metres) +
    (1 | survey_year),
  # offset(npoint),
  family = poisson,
  ziformula=~ wpfg,
  #dispformula =~ wpfg ,
  data = veg_cover_ar_sum |> filter(!wpfg_ori %in% c("Atl_native", "Ate_native", "Tda_unknown"))
)
```

Conditional model:

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-6.338926	3.945021	-1.61	0.1081	
log_hits_tm1	0.466865	0.004717	98.98	< 2e-16	***
days_above_baseflow_std	3.208391	3.369341	0.95	0.3410	
wpfgATe	6.158428	3.941779	1.56	0.1182	
wpfgATl	3.547942	0.684107	5.19	2.15e-07	***
wpfgSe	-21.385108	25.954264	-0.82	0.4100	
wpfgSk	1.100790	1.345171	0.82	0.4132	
wpfgTda	7.307450	3.941643	1.85	0.0638	.
wpfgTdr	7.736796	3.941596	1.96	0.0497	*
originnative	2.784402	3.994830	0.70	0.4858	
days_above_springfresh_std	-0.123400	0.390857	-0.32	0.7522	
days_above_baseflow_std:wpfgATe	-2.987176	3.367936	-0.89	0.3751	
days_above_baseflow_std:wpfgATl	0.155333	0.657344	0.24	0.8132	
days_above_baseflow_std:wpfgSe	14.468021	26.401090	0.55	0.5837	
days_above_baseflow_std:wpfgSk	0.182377	1.265110	0.14	0.8854	
days_above_baseflow_std:wpfgTda	-2.975603	3.367609	-0.88	0.3769	
days_above_baseflow_std:wpfgTdr	-2.989015	3.367573	-0.89	0.3748	
days_above_baseflow_std:originnative	-2.304494	3.430569	-0.67	0.5017	
wpfgATe:originnative	-1.514880	3.995046	-0.38	0.7045	
wpfgATl:originnative	NA	NA	NA	NA	
wpfgSe:originnative	NA	NA	NA	NA	
wpfgSk:originnative	NA	NA	NA	NA	
wpfgTda:originnative	-2.579748	3.994900	-0.65	0.5184	
wpfgTdr:originnative	-3.710273	3.995021	-0.93	0.3530	
wpfgATe:days_above_springfresh_std	1.010669	0.382893	2.64	0.0083	**
wpfgATl:days_above_springfresh_std	-1.755976	0.272420	-6.45	1.15e-10	***
wpfgSe:days_above_springfresh_std	2.328011	3.084072	0.75	0.4503	
wpfgSk:days_above_springfresh_std	-0.323750	0.546548	-0.59	0.5536	
wpfgTda:days_above_springfresh_std	0.255564	0.381392	0.67	0.5028	
wpfgTdr:days_above_springfresh_std	0.271463	0.381315	0.71	0.4765	
originnative:days_above_springfresh_std	2.036161	0.468210	4.35	1.37e-05	***
days_above_baseflow_std:wpfgATe:originnative	2.376384	3.430972	0.69	0.4885	
days_above_baseflow_std:wpfgATl:originnative	NA	NA	NA	NA	
days_above_baseflow_std:wpfgSe:originnative	NA	NA	NA	NA	
days_above_baseflow_std:wpfgSk:originnative	NA	NA	NA	NA	
days_above_baseflow_std:wpfgTda:originnative	2.166421	3.430644	0.63	0.5277	
days_above_baseflow_std:wpfgTdr:originnative	2.291317	3.430663	0.67	0.5042	
wpfgATe:originnative:days_above_springfresh_std	-2.761693	0.469796	-5.88	4.14e-09	***
wpfgATl:originnative:days_above_springfresh_std	NA	NA	NA	NA	
wpfgSe:originnative:days_above_springfresh_std	NA	NA	NA	NA	
wpfgSk:originnative:days_above_springfresh_std	NA	NA	NA	NA	
wpfgTda:originnative:days_above_springfresh_std	-1.953152	0.468932	-4.17	3.11e-05	***
wpfgTdr:originnative:days_above_springfresh_std	-2.770192	0.472856	-5.86	4.67e-09	***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

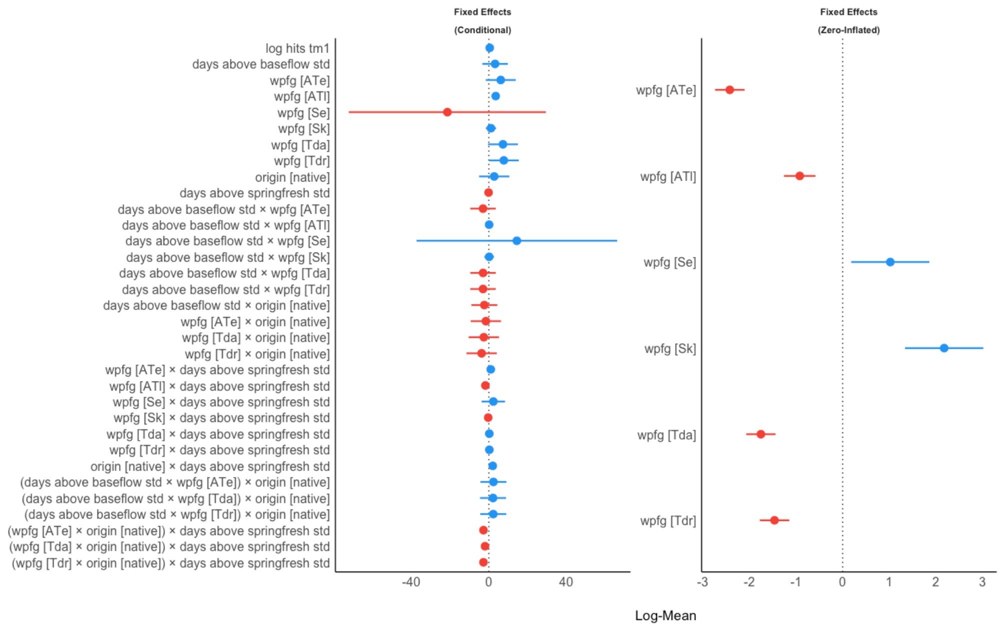


Figure 1.

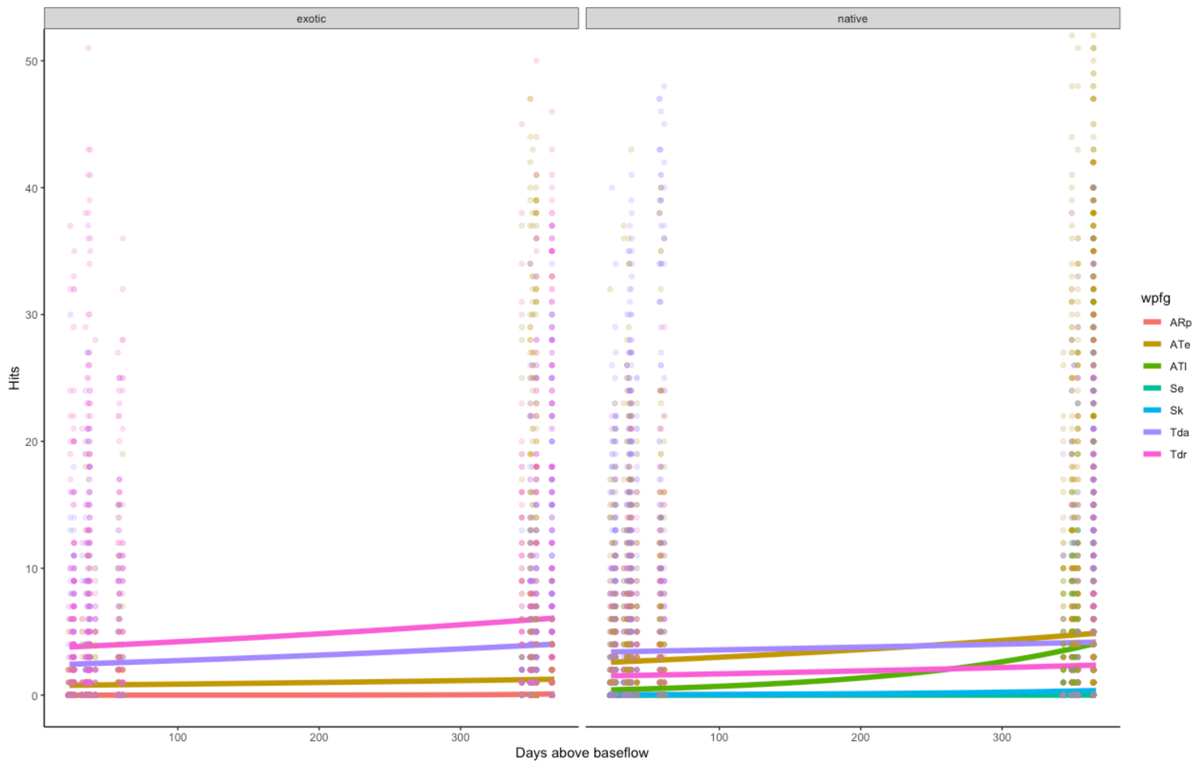


Figure 2. Model predicted plant cover for the three way interaction between ‘Days above baseflow’, ‘functional group’ and ‘origin’. Points are raw data.

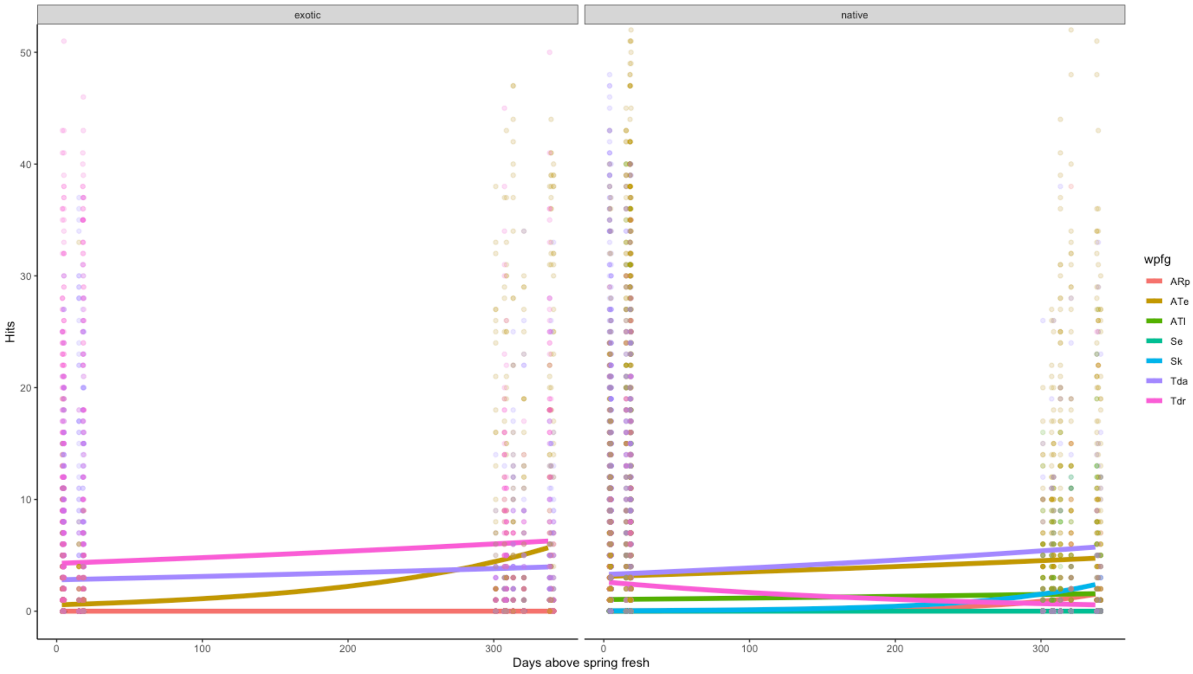


Figure 3. Model predicted plant cover for the three way interaction between 'Days above springfresh', 'functional group' and 'origin'. Points are raw data.

Vegetation cover: flow event model (version 1)

This model includes several categorical predictors (origin, wpfg, grazing, and a zone-by-period interaction) and the full suite of random effects. The aim is to examine how vegetation cover changes in specific zones before and after key flow events (spring and summer freshes). This first version allows functional groups to have different levels of cover *but assumes that changes following flow events are similar in all groups within a zone*. Although this is not an ideal model structure, it is likely that functional groups are restricted to particular zones, in which case the zone-by-period interaction may capture some of the variation attributable to functional groupings.

The model indicated that vegetation cover differed in its responses to spring and summer freshes in each zone (below baseflow, baseflow to spring fresh, above spring fresh). Below baseflow level, vegetation increased following the spring fresh and remained high following the summer event (Figure 5). In the baseflow-to-spring fresh zone, vegetation increased following the spring fresh but returned to pre-spring levels following the summer fresh (Figure 5). In the above-spring fresh zone, vegetation level declined following both the spring and the summer fresh (Figure 5).

Model structure

```
cover_ar_TMBmod_2 <- glmmTMB::glmmTMB(
  hits ~ log_hits_tm1 +
   # days_above_baseflow_std + days_above_springfresh_std +
   # days_above_baseflow_std^2 + #days_above_springfresh_std_sq +
    zone*period +
     origin + wpfg +
    grazing +
    (1 | site / transect) +
    #(1 | site / period) +
    (1 | metres) +
    (1 | survey_year),
  # offset(npoint),
  family = poisson,
  ziformula=~ wpfg,
  #dispformula =~ wpfg ,
  data = veg_cover_ar_sum |> filter(!wpfg_ori %in% c("Atl_native", "Ate_native", "Tda_unknown"))
)
```

```
Random effects:

Conditional model:
 Groups       Name        Variance  Std.Dev.
 transect:site (Intercept) 0.1495148 0.38667
 site         (Intercept) 0.1083039 0.32910
 metres       (Intercept) 0.1042668 0.32290
 survey_year  (Intercept) 0.0006314 0.02513
Number of obs: 36817, groups:  transect:site, 50; site, 5; metres, 22; survey_year, 3

Conditional model:
                                               Estimate Std. Error z value Pr(>|z|)
(Intercept)                                    0.507407   0.202834    2.50   0.0124 *
log_hits_tm1                                   0.503130   0.004658  108.00  < 2e-16 ***
zonebaseflow_to_springfresh                    0.135690   0.019849    6.84 8.14e-12 ***
zonebelow_baseflow                             0.027219   0.062187    0.44   0.6616
periodafter_summer                            -0.095429   0.019586   -4.87 1.10e-06 ***
periodbefore_spring                            0.275100   0.019116   14.39  < 2e-16 ***
originnative                                   0.022715   0.011601    1.96   0.0502 .
wpfgATe                                        0.568841   0.074505    7.63 2.26e-14 ***
wpfgATl                                        0.392510   0.077440    5.07 4.01e-07 ***
wpfgSe                                        -0.532349   0.235144   -2.26   0.0236 *
wpfgSk                                         0.785098   0.158281    4.96 7.04e-07 ***
wpfgTda                                        0.580179   0.075459    7.69 1.49e-14 ***
wpfgTdr                                        0.740681   0.076091    9.73  < 2e-16 ***
grazingY                                      -0.414008   0.207655   -1.99   0.0462 *
zonebaseflow_to_springfresh:periodafter_summer -0.049931   0.022829   -2.19   0.0287 *
zonebelow_baseflow:periodafter_summer          0.072623   0.060393    1.20   0.2292
zonebaseflow_to_springfresh:periodbefore_spring -0.474324  0.023176  -20.47  < 2e-16 ***
zonebelow_baseflow:periodbefore_spring        -0.844370   0.080058  -10.55  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

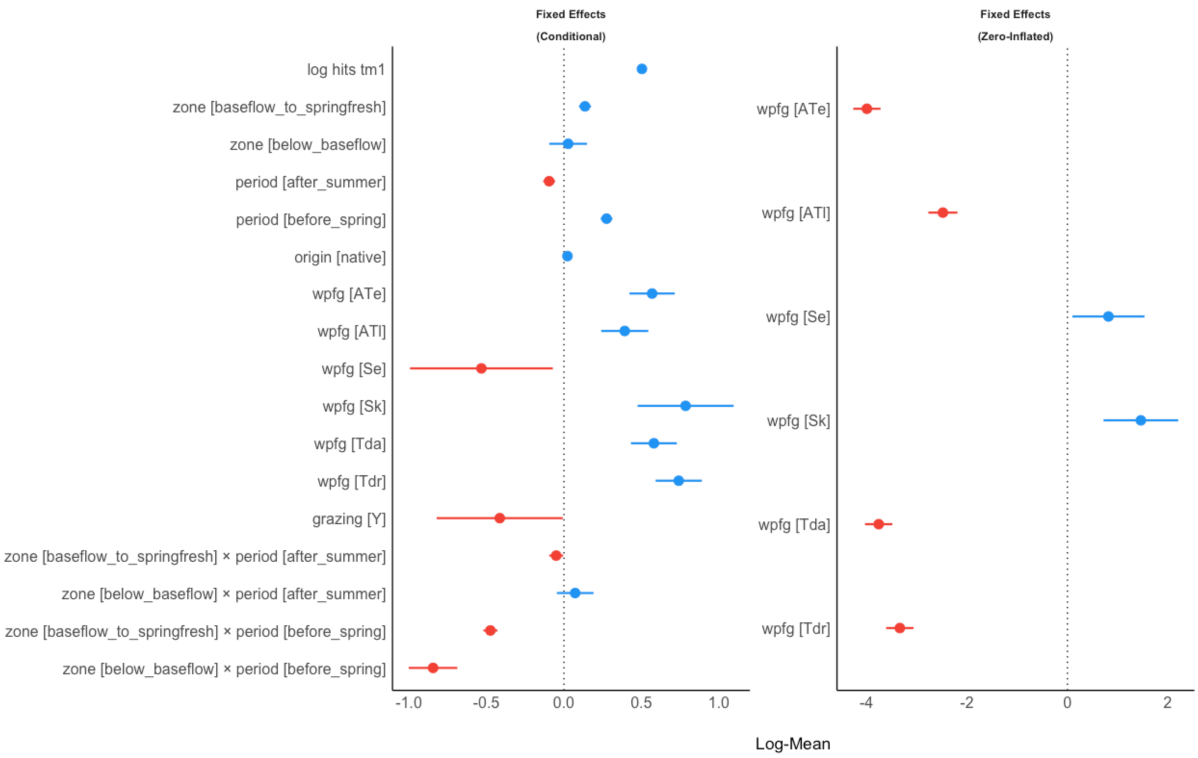


Figure 4.

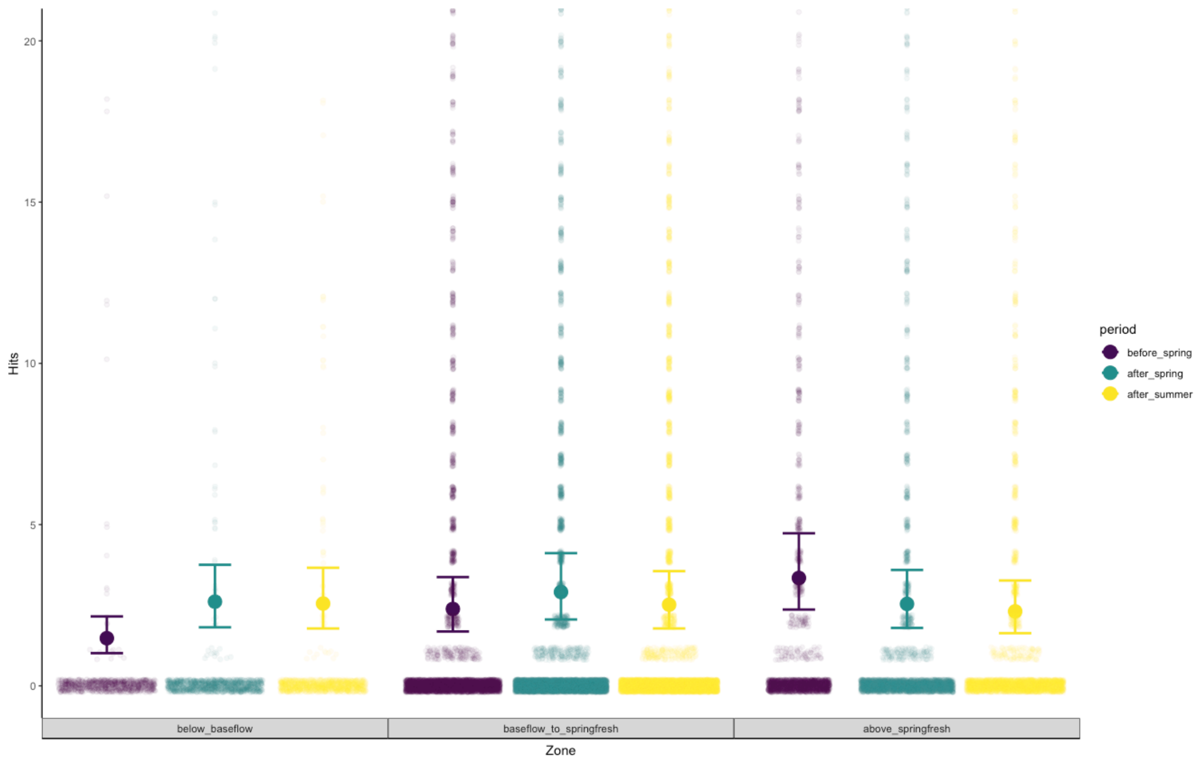


Figure 5. Model predicted plant cover for the two way interaction between ‘period’ and ‘zone’. Points are raw data with the spread across x-axis representing data density.

Vegetation cover: flow event model (version 2)

This model includes several categorical predictors (origin, zone, grazing, and a wpfg-by-period interaction) and the full suite of random effects. The aim is to examine how vegetation cover of each functional grouping changes before and after key flow events (spring and summer freshes). This second version allows functional groups to have different responses to flow events (spring and summer freshes) *but assumes that vegetation in all zones changes similarly following each flow event*. As for version 1, this is not an ideal model structure but provides an avenue to tease apart wpfg-specific responses to flow events.

The model indicated that cover of each functional group differed in its response to spring and summer freshes (Figure 7). Key patterns were as follows:

- ARp: increased after the spring fresh and again after the summer fresh.
- ATe: increased after the spring fresh and returned to pre-spring levels after the summer fresh.
- ATl: as for ATe.
- Se: increased after the spring fresh and largely maintained this level after the summer fresh.
- Sk: no change following the spring fresh but large increases following the summer fresh.
- Tda: no change following either flow event.
- Tdr: reduced following spring fresh then recovered slightly following the summer fresh, but not to pre-spring levels.

Model structure

```
cover_ar_TMBmod_3 <- glmmTMB::glmmTMB(
  hits ~ log_hits_tm1 +
    # days_above_baseflow_std + days_above_springfresh_std +
    # days_above_baseflow_std^2 + #days_above_springfresh_std_sq +
    zone + wpfg * period +
     origin +
    grazing +
    (1 | site / transect) +
    #(1 | site / period) +
    (1 | metres) +
    (1 | survey_year),
  # offset(npoint),
  family = poisson,
  ziformula=~ wpfg,
  #dispformula =~ wpfg ,
  data = veg_cover_ar_sum |> filter(!wpfg_ori %in% c("Atl_native", "Ate_native", "Tda_unknown"))
)
```

```
Conditional model:
 Groups       Name        Variance Std.Dev.
 transect:site (Intercept) 0.143451 0.37875
 site         (Intercept) 0.103689 0.32201
 metres       (Intercept) 0.099915 0.31609
 survey_year  (Intercept) 0.001128 0.03359
Number of obs: 36817, groups:  transect:site, 50; site, 5; metres, 22; survey_year, 3

Conditional model:
                             Estimate Std. Error z value Pr(>|z|)
(Intercept)                  0.490933   0.220532    2.23 0.026006 *
log_hits_tm1                 0.526108   0.004773  110.22  < 2e-16 ***
zonebaseflow_to_springfresh -0.020485   0.014807   -1.38 0.166522
zonebelow_baseflow          -0.135442   0.051703   -2.62 0.008803 **
wpfgATe                      0.762082   0.120587    6.32 2.62e-10 ***
wpfgATl                      0.826599   0.124319    6.65 2.95e-11 ***
wpfgSe                      -0.219954   0.361766   -0.61 0.543187
wpfgSk                      -0.409777   0.497890   -0.82 0.410493
wpfgTda                      0.640675   0.121310    5.28 1.28e-07 ***
wpfgTdr                      0.361223   0.122625    2.95 0.003222 **
periodafter_summer           0.259131   0.152882    1.69 0.090080 .
periodbefore_spring         -0.769916   0.262410   -2.93 0.003346 **
originnative                 0.022738   0.011705    1.94 0.052058 .
grazingY                    -0.406325   0.203425   -2.00 0.045780 *
wpfgATe:periodafter_summer  -0.570432   0.153550   -3.71 0.000203 ***
wpfgATl:periodafter_summer  -0.864219   0.159835   -5.41 6.41e-08 ***
wpfgSe:periodafter_summer   -0.445005   0.462139   -0.96 0.335586
wpfgSk:periodafter_summer    1.502212   0.526754    2.85 0.004347 **
wpfgTda:periodafter_summer  -0.310229   0.153846   -2.02 0.043749 *
wpfgTdr:periodafter_summer   0.052045   0.155180    0.34 0.737333
wpfgATe:periodbefore_spring  0.414490   0.262945    1.58 0.114948
wpfgATl:periodbefore_spring -0.060499   0.268355   -0.23 0.821633
wpfgSe:periodbefore_spring  -0.968411   0.865331   -1.12 0.263088
wpfgSk:periodbefore_spring   0.760442   0.736038    1.03 0.301531
wpfgTda:periodbefore_spring  0.611485   0.263133    2.32 0.020133 *
wpfgTdr:periodbefore_spring  1.539382   0.263473    5.84 5.14e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

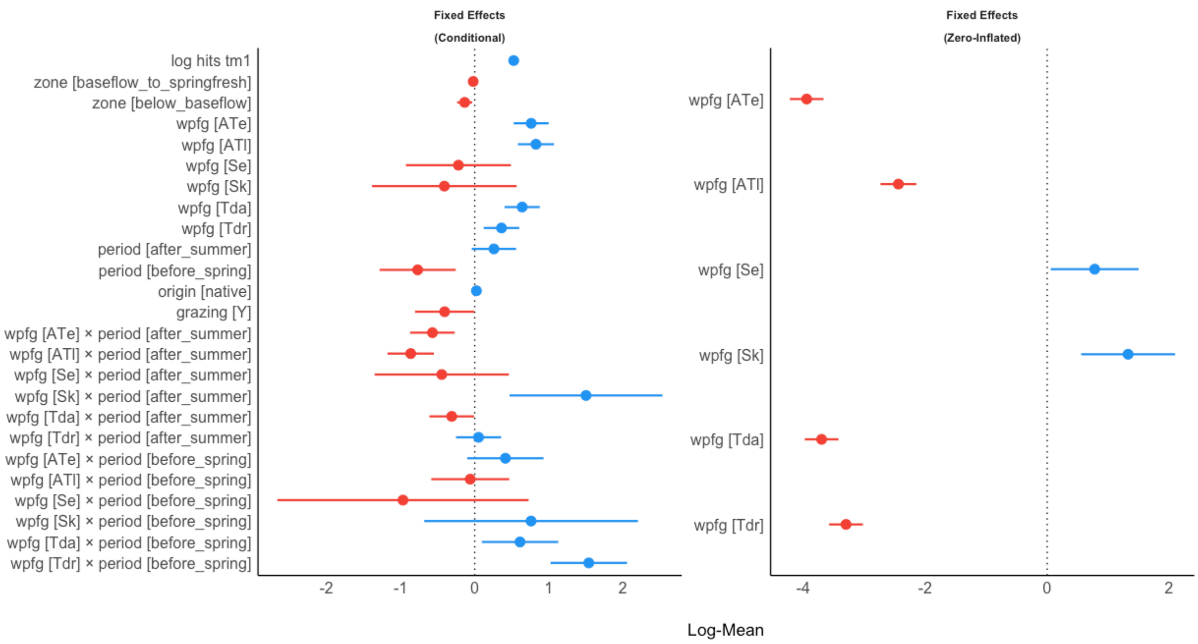


Figure 6.

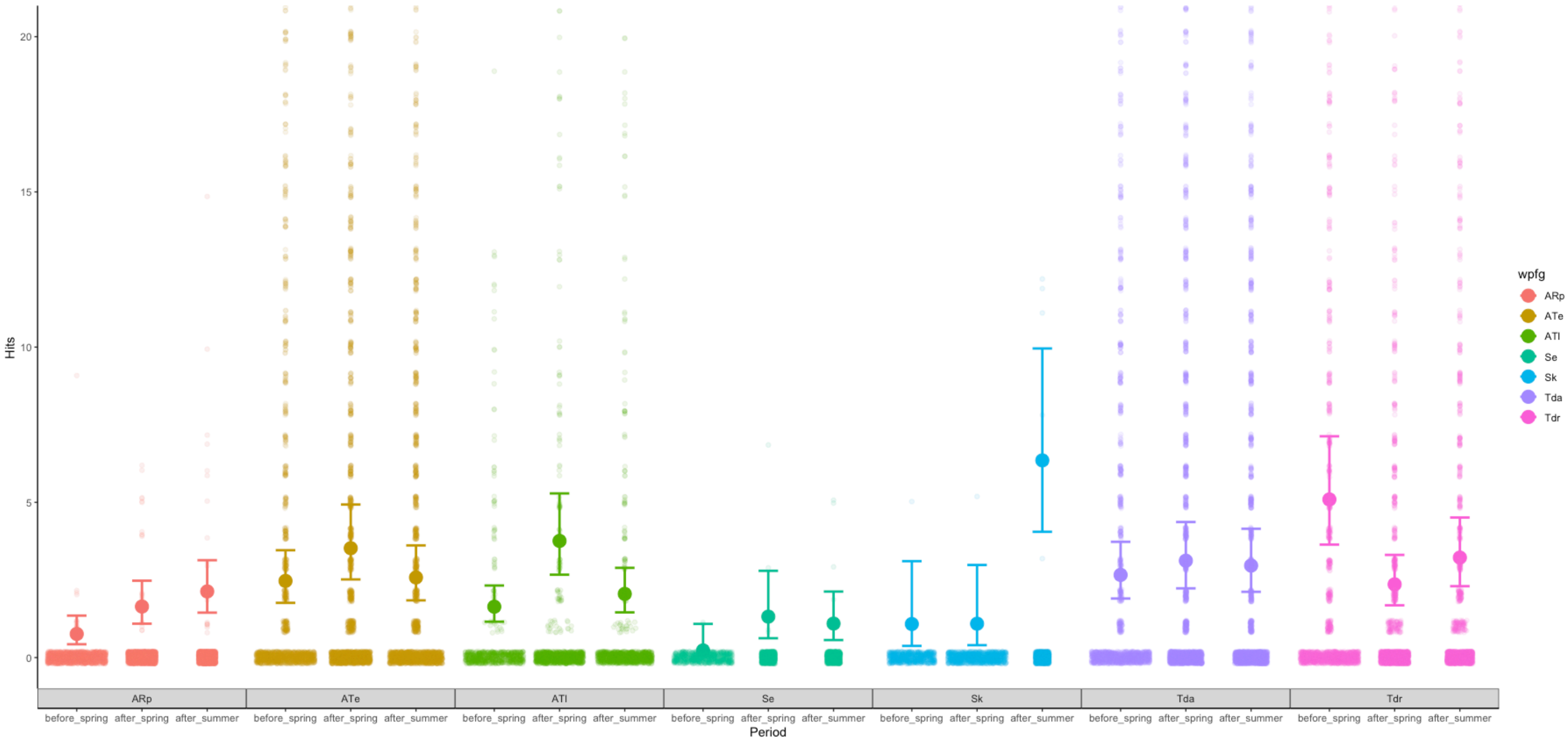


Figure 7. Model predicted plant cover for the two way interaction between 'period' and 'functional group'. Points are raw data with the spread across x-axis representing data density.

Species richness model

Species richness models are not fully determined yet. The below structure captures similar patterns to version 1 and version 2 of the cover model. The model includes two-way interactions between zone and period, zone and plant functional group, and period and plant functional group. Together, these terms aim to identify responses to flow events within a given zone or functional group, while accounting for differences in species richness of each functional group among zones.

The models illustrate a general pattern of relatively low temporal variation in species richness once accounting for spatial patterns due to zone and functional grouping. Figure 9 illustrates that total species richness does not change substantially within each zone following the spring or summer fresh. Similarly, Figure 11 shows that total species richness of each plant functional group does not change substantially following spring or summer freshes. By contrast, Figure 10 illustrates that the species richness of each plant functional group differs among zones. As expected, aquatic groupings are more species-richness below the baseflow level, whereas terrestrial groupings are more species richness above the spring fresh level.

Model structure

Coauthor-3 comment: coauthor-4 to update, the dispformula is probably ignored for a Poisson model, use ziformula instead. Can flow metrics be included?

```
richness_ar_TMBmod_1 <- glmmTMB::glmmTMB(
  richness ~
    #days_above_baseflow_std + days_above_springfresh_std +
    # days_above_baseflow_std^2 + days_above_springfresh_std^2 +
    #   zone * period +
     zone *period + zone*wpfg + wpfg*period +
      grazing + origin +
    (1 | site / transect) +
    #(1 | site / period) +
    (1 | metres) +
    (1 | survey_year),
  # offset(npoint),
  family = poisson,
  #ziformula=~ wpfg,
  dispformula =~ wpfg ,
  data = veg_richness |> filter(!wpfg_ori %in% c("Atl_native", "Ate_native", "Tda_unknown"))
)
```

Best fitting so far – the three way interaction had convergence issues

```
Conditional model:
                            Estimate Std. Error z value Pr(>|z|)    
(Intercept)                 0.490933   0.220532    2.23 0.026006 *  
log_hits_tm1                0.526108   0.004773  110.22  < 2e-16 ***
zonebaseflow_to_springfresh -0.020485  0.014807   -1.38 0.166522    
zonebelow_baseflow          -0.135442  0.051703   -2.62 0.008803 ** 
wpfgATe                     0.762082   0.120587    6.32 2.62e-10 ***
wpfgATl                     0.826599   0.124319    6.65 2.95e-11 ***
wpfgSe                     -0.219954   0.361766   -0.61 0.543187    
wpfgSk                     -0.409777   0.497890   -0.82 0.410493    
wpfgTda                     0.640675   0.121310    5.28 1.28e-07 ***
wpfgTdr                     0.361223   0.122625    2.95 0.003222 ** 
periodafter_summer          0.259131   0.152882    1.69 0.090080 .  
periodbefore_spring        -0.769916   0.262410   -2.93 0.003346 ** 
originnative                0.022738   0.011705    1.94 0.052058 .  
grazingY                   -0.406325   0.203425   -2.00 0.045780 *  
wpfgATe:periodafter_summer -0.570432   0.153550   -3.71 0.000203 ***
wpfgATl:periodafter_summer -0.864219   0.159835   -5.41 6.41e-08 ***
wpfgSe:periodafter_summer  -0.445005   0.462139   -0.96 0.335586    
wpfgSk:periodafter_summer   1.502212   0.526754    2.85 0.004347 ** 
wpfgTda:periodafter_summer -0.310229   0.153846   -2.02 0.043749 *  
wpfgTdr:periodafter_summer  0.052045   0.155180    0.34 0.737333    
wpfgATe:periodbefore_spring  0.414490   0.262945    1.58 0.114948    
wpfgATl:periodbefore_spring -0.060499   0.268355   -0.23 0.821633    
wpfgSe:periodbefore_spring -0.968411   0.865331   -1.12 0.263088    
wpfgSk:periodbefore_spring  0.760442   0.736038    1.03 0.301531    
wpfgTda:periodbefore_spring  0.611485   0.263133    2.32 0.020133 *  
wpfgTdr:periodbefore_spring  1.539382   0.263473    5.84 5.14e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

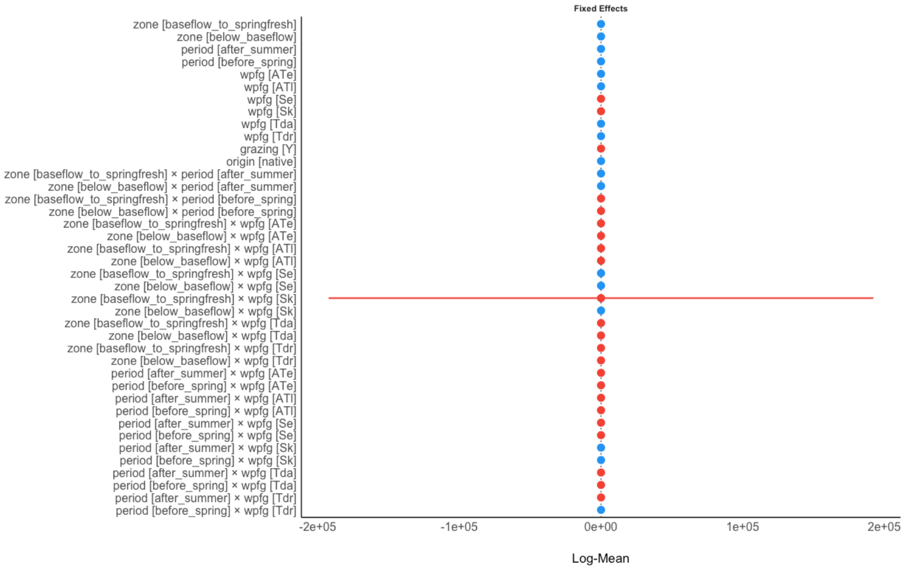


Figure 8. Something funny here. Coauthor-3: suggest issue with the Sk functional grouping in that zone. Assess following model tweaks.

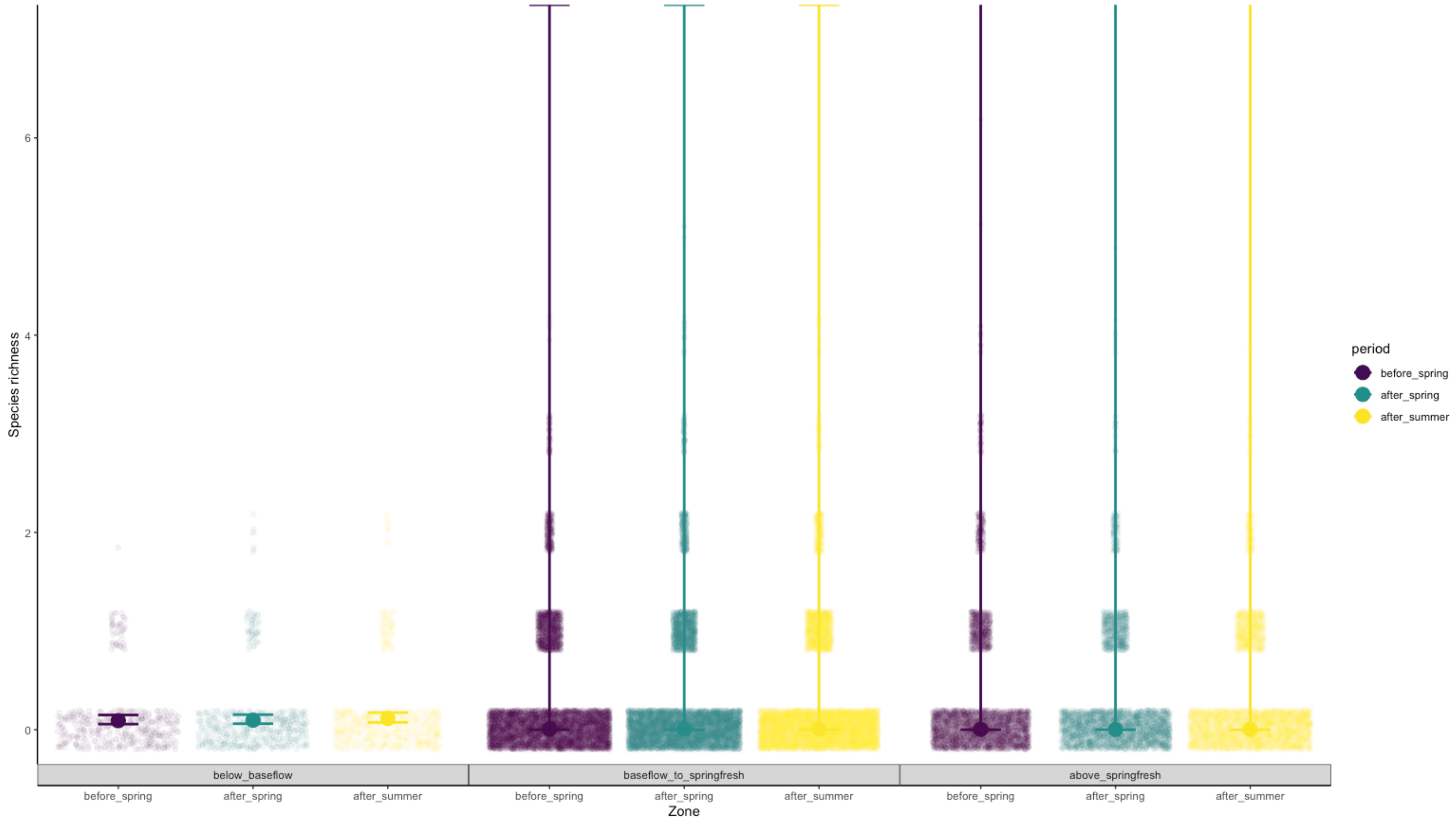


Figure 9. Model predicted species richness for the two way interaction between ‘period’ and ‘zone’. Points are raw data with the spread across x-axis representing data density.

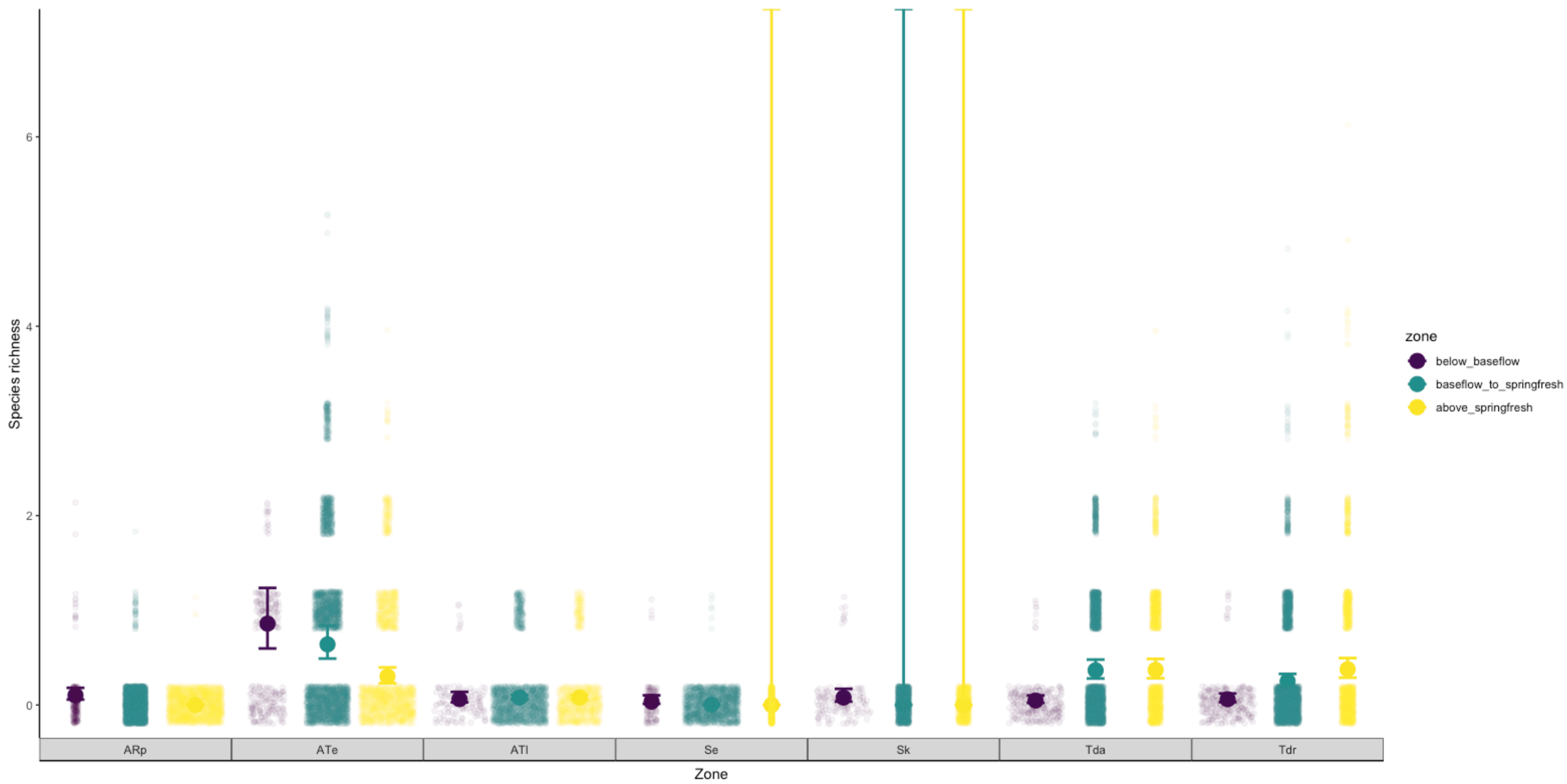


Figure 10. Model predicted species richness for the two way interaction between 'zone' and 'functional group'. Points are raw data with the spread across x-axis representing data density.

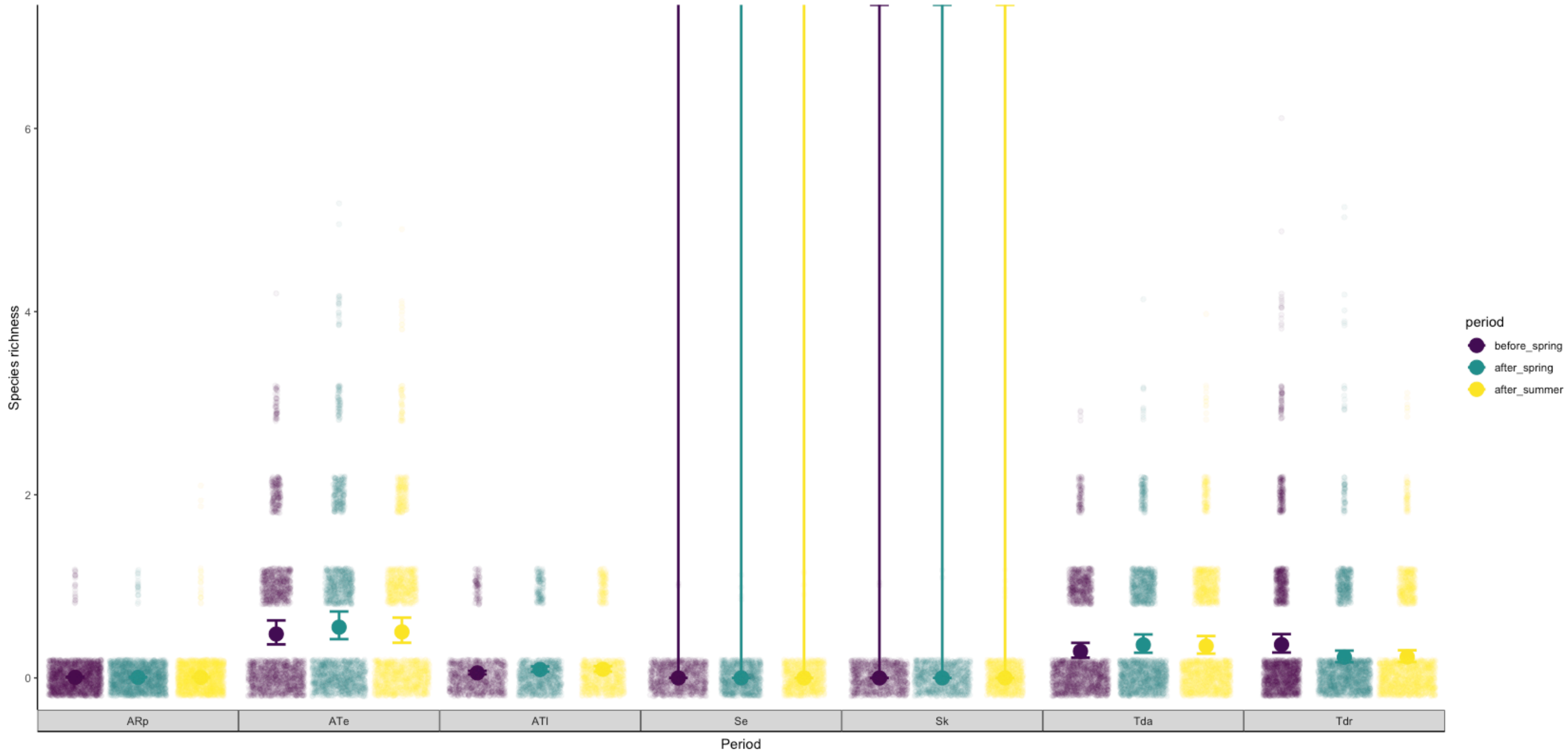


Figure 11. Model predicted species richness for the two way interaction between ‘period’ and ‘functional group’. Points are raw data with the spread across x-axis representing data density.