R Notebook

Anthony Davidson

February 2019

Table of Contents

Throughout my candatiture I have developed my thesis while simulationously following and attempting to combat the “Reproducibility crisis” (**peng2015?**) by writing a series of computationally reproducible pipelines (R packages and github repostories; ??) that faciliate the sound curation, analysis and communication of datasets collected for large-scale pest control programs at the national scale *[Davidson2020-Reproducibility; Invasive species database]* ??.

I use this framework to address the “unexpected” outcomes of predator removal in two key native forest types in New Zealand *[Davidson2020-Beech-forests\* 5.1; \*MPD-forests]* ??. I fit different state-space models to account for the underlying demographic processes of increased complexity. I use New Zealand PFNZ 2050 “appollo shot” to demostrate the applicability of this reproducible method to support citizen science and community driven predator control *[Davidson2020-PFNZ2050]* ??.

# 1 Overview

The overarching aim of this PhD was to examine the role of predation, competition and resource flow in regulating invasive mammal populations in New Zealand forests ecosystems. A key aspect of this research is attempting to account for unexpected outcomes of multi-species predator control. The inherant ….

I have done this by constructing and fitting ecological models that describe the population dynamics of interacting invasive species in New Zealand forests. I use the understanding derived from these models to predict the effects of management manipulations and reduce the likelihood of unanticipated outcomes. Aligning this ecological research with reproducible research practices insures the results presented within this thesis. The final result is a reproducible and repeatable by allowing open access guidelines to document the statistical code and analysis.

This research can be used support other modifications to the forecasting models presented to estimate the likely effects of species removals. In 2016 the New Zealand government funded the “Predator Free New Zealand 2050”. The disscussion combines the reproducible approach and Bayesian Modelling techniques to input additional obsersvational data to test the many othr ecosystem models that allow researchers to directly quantify the interactions among invasive species (**Peng2015?**).

## 1.1 This thesis generates:

1. A reproducible and scalable research database for invasive mammals in NZ.
2. Advanced in ecological models that incorperate both the theoretical models and observational data (**King2012?**).
3. Interactive tools to develop understanding and application of Bayesian Hierarichal Models for conservation managers and citizen researchers alike.

Submission

3.4. A candidate must sign a declaration that the thesis does notcontain any material published or written by another person except where due reference is made in the text or footnotes.

Material

produced jointly by acandidate and his/her supervisors or others can only be included in the narrative of the thesis if the candidate was explicitly involved in the original work. Any jointly-produced material in the examination submission must be accompanied by astatementclearly indicatingthe candidate’s contribution to the research.

3.5. No material or publications presented for examination for any other degree within this or any other institution will be submitted for assessment unless its incorporation in the thesis is declared in a statement. Content and structural requirements for the thesis

3.6. A doctoral thesis must make a distinctand significantcontribution to knowledge or understanding in the area of research and/or the application of knowledge to the analysis of problems in the study area; and mustafford evidence of originality.

3.7. A masters thesis shall display asound knowledge of the field of the research and include substantialcritical review of the field. 3.8. Subject to permission being obtained from publishers if necessary, the copyright of the thesis is deemed to be vested in the author.The relevant University policy for intellectual property principles as they relate to Higher Degree by Research candidates is the Intellectual Property Policy.

3.9. The followingapplies to the length of the thesis: A thesis submitted for a Doctor of Philosophy degree should not exceed 100,000 words; A thesis submitted for a Professional Doctorate degree should not exceed 60,000 words; A thesis submitted for a Masters by Research degree should not exceed 60,000 words.

3.10. Thesis requirements will in part be dictated by disciplinary requirements and the type of thesis produced.See Higher Degree by Research Thesis Submission and Examination Guidelines for the thesis requirements of submissions for examination, includingspecificguidelines and requirements for the submission ofathesis consisting of published work. Format requirements for the thesis

3.11. The University has aset ofgeneric formattingrequirements for theses[1].These requirements are set on in the Higher Degree by Research ThesisSubmission and Examination Guidelines.

3.12. It is expected that the supervisory panel will provide editorialadvice to the candidate.

3.13. Candidates may use a professional editor in preparingtheir thesis for submission but must strictly follow the guidelines set out in the Higher Degree by Research ThesisSubmission and Examination

Guidelines.

[1]For research theses incorporatingcreative production these formattingrequirements relate to the exegeticalcomponent of the thesis

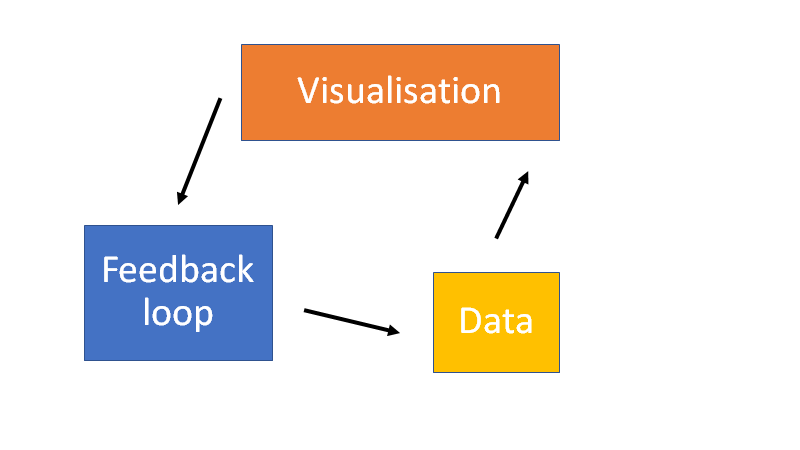
# 2 Introduction

Invasive species are regarded to have large impacts native species globally (**Doherty2016b?**). Meta-analysis highlight the unexpected impacts of not monitoring and account for dynamic changes in both species interactions and environmental changes (**Tompkins2006?**; **Caut2009?**). Under these conditions, repeatability and computational reproducibility are vital aspects to the development of this research (**Buerkle2019?**).

This is an attempt to link several projects I have been working on under the umbrella of my tidypipes approach.

NOTE: This is my first attempt to apply my tidyPipes workflow (Figure above) to my project management and integrate this into my normal workflow.

Generally, this work is focused on writing the scripts to automate the intergration between emails, PhD timeline and other projects. To do this we need to import datasets and modify the structure of these imputs to match the information needed to contruct a timeline of tasks and objectives.



## 2.1 Code, data and manuscript isn’t enough

A recent publication suggests that the accessibility and application of the code that is keep along side standard scientific publications is infact often incomplete or inefficient to achieve its goal of reproducibility.

They achieved their reproducible methods by using RMarkdown and the following yaml content:

---  
title: Low availability of code in ecology&#58; a call for urgent action  
author: Antica Culina, Ilona van den Berg, Simon Evans, Alfredo Sanchez-Tojar  
date: 'Last updated: 01 February, 2022'  
output:  
 html\_document:  
 code\_download: yes  
 code\_folding: hide  
 toc: yes  
 toc\_float: yes  
 pdf\_document:  
 toc: yes  
 word\_document:  
 toc: yes  
bibliography: Culina\_et\_al.bib  
subtitle: Supporting Information  
---

and the following R package information:

devtools::session\_info()

## - Session info ---------------------------------------------------------------  
## setting value  
## version R version 4.1.2 (2021-11-01)  
## os Windows 10 x64 (build 19042)  
## system x86\_64, mingw32  
## ui RTerm  
## language (EN)  
## collate English\_United States.1252  
## ctype English\_United States.1252  
## tz Australia/Sydney  
## date 2022-02-01  
## pandoc 2.17.0.1 @ C:/PROGRA~1/Pandoc/ (via rmarkdown)  
##   
## - Packages -------------------------------------------------------------------  
## package \* version date (UTC) lib source  
## bookdown 0.24 2021-09-02 [1] CRAN (R 4.1.2)  
## brio 1.1.3 2021-11-30 [1] CRAN (R 4.1.2)  
## cachem 1.0.6 2021-08-19 [1] CRAN (R 4.1.2)  
## callr 3.7.0 2021-04-20 [1] CRAN (R 4.1.2)  
## cli 3.1.0 2021-10-27 [1] CRAN (R 4.1.2)  
## crayon 1.4.2 2021-10-29 [1] CRAN (R 4.1.2)  
## desc 1.4.0 2021-09-28 [1] CRAN (R 4.1.2)  
## devtools 2.4.3 2021-11-30 [1] CRAN (R 4.1.2)  
## digest 0.6.29 2021-12-01 [1] CRAN (R 4.1.2)  
## ellipsis 0.3.2 2021-04-29 [1] CRAN (R 4.1.2)  
## evaluate 0.14 2019-05-28 [1] CRAN (R 4.1.2)  
## fastmap 1.1.0 2021-01-25 [1] CRAN (R 4.1.2)  
## fs 1.5.2 2021-12-08 [1] CRAN (R 4.1.2)  
## glue 1.6.0 2021-12-17 [1] CRAN (R 4.1.2)  
## highr 0.9 2021-04-16 [1] CRAN (R 4.1.2)  
## htmltools 0.5.2 2021-08-25 [1] CRAN (R 4.1.2)  
## knitr 1.37 2021-12-16 [1] CRAN (R 4.1.2)  
## lifecycle 1.0.1 2021-09-24 [1] CRAN (R 4.1.2)  
## magrittr 2.0.1 2020-11-17 [1] CRAN (R 4.1.2)  
## memoise 2.0.1 2021-11-26 [1] CRAN (R 4.1.2)  
## pkgbuild 1.3.1 2021-12-20 [1] CRAN (R 4.1.2)  
## pkgload 1.2.4 2021-11-30 [1] CRAN (R 4.1.2)  
## png 0.1-7 2013-12-03 [1] CRAN (R 4.1.1)  
## prettyunits 1.1.1 2020-01-24 [1] CRAN (R 4.1.2)  
## processx 3.5.2 2021-04-30 [1] CRAN (R 4.1.2)  
## ps 1.6.0 2021-02-28 [1] CRAN (R 4.1.2)  
## purrr 0.3.4 2020-04-17 [1] CRAN (R 4.1.2)  
## R6 2.5.1 2021-08-19 [1] CRAN (R 4.1.2)  
## remotes 2.4.2 2021-11-30 [1] CRAN (R 4.1.2)  
## rlang 0.4.12 2021-10-18 [1] CRAN (R 4.1.2)  
## rmarkdown 2.11 2021-09-14 [1] CRAN (R 4.1.2)  
## rprojroot 2.0.2 2020-11-15 [1] CRAN (R 4.1.2)  
## rstudioapi 0.13 2020-11-12 [1] CRAN (R 4.1.2)  
## sessioninfo 1.2.2 2021-12-06 [1] CRAN (R 4.1.2)  
## stringi 1.7.6 2021-11-29 [1] CRAN (R 4.1.2)  
## stringr 1.4.0 2019-02-10 [1] CRAN (R 4.1.2)  
## testthat 3.1.2 2022-01-20 [1] CRAN (R 4.1.2)  
## usethis 2.1.5 2021-12-09 [1] CRAN (R 4.1.2)  
## withr 2.4.3 2021-11-30 [1] CRAN (R 4.1.2)  
## xfun 0.29 2021-12-14 [1] CRAN (R 4.1.2)  
## yaml 2.2.1 2020-02-01 [1] CRAN (R 4.1.1)  
##   
## [1] C:/Users/Home1\_local/Documents/R/win-library/4.1  
## [2] C:/Program Files/R/R-4.1.2/library  
##   
## ------------------------------------------------------------------------------

## 2.2 Examples

SEE tidyPipes MANUAL for general concept and approach. This repository contains example data and scripts needed to visualise and communicate multiple different projects including:

* PhDplanning2020/ contains the documentation and reporting for graduate research.
* flatttingRMDstlyle/ (a repository for documentign and recording reproducible reports for my flatting arrangements).

## 2.3 Statistical models

There are three general levels of assumptions we make when building any statistical model.

#### 2.3.0.1 1. Control to Treatment

“Can I make a representative comparison between my control and treatment groups?”

Ecological studies commonly focus on this issue. This can be very hard to verify in ecological studies as the environmental variables are not held constant. Ecologists have focused on randomisation and block design to deal with these issues.

#### 2.3.0.2 2. Sample to Population

“Is my data representative of the total population of interest?”

When estimating abundance of species this is a key aspect of estimating population parameters sub as abundance. Within this thesis we account for this using capture-recapture studies. This allows our models to quantify this uncertainty by looking at the detection curves of individuals within each population.

#### 2.3.0.3 3. Observation to Process

“Do my observations represent the underlying process I assume is being undertaken?”

My thesis focuses on this set of assumptions. By building Bayesian State Space models we can compare the effects of our assumed process model in comparison to the observed data using the Bayesian model.

# 3 Reproducible framework

This chapter creates a workflow that allows the ability to access and use the invasive species database in a reproducible way. I take the existing research analysis to build a reproducible research synthesis database of the theoretical relationships proposed in over 100 years of conservation research in NZ (e.g.  (**Ruscoe2001?**; **Ruscoe2005?**; **Tompkins2006?**; **Holland2015?**; **King1983?**).

## 3.1 [Davidson2020-Reproducibility]

**Title:** *Dealing with Reproducibility: Why and how to make a reproducible workflow for invasive species research in New Zealand*

**Abstract:** Science is currently undergoing a “Reproducibility crisis” (**Baker2016a?**) but the tools and methods to ilimate certian aspects of this crisis with respect to computating software and platforms used to conduct this research. I apply novel reproduciblity measures throughout this thesis to replicate theoretical processes and fit observed data to these models to provide a framework that can account for dynamic changes in observd outcomes of predator control.

**Status:** Frameworks working…need to write up.

**DRAFT:** Here is a current document of the draft (May2020)["./vigneettes/Reproducible-framework-for-estimating-New-Zealand-Forest-Dynamics.docx"]

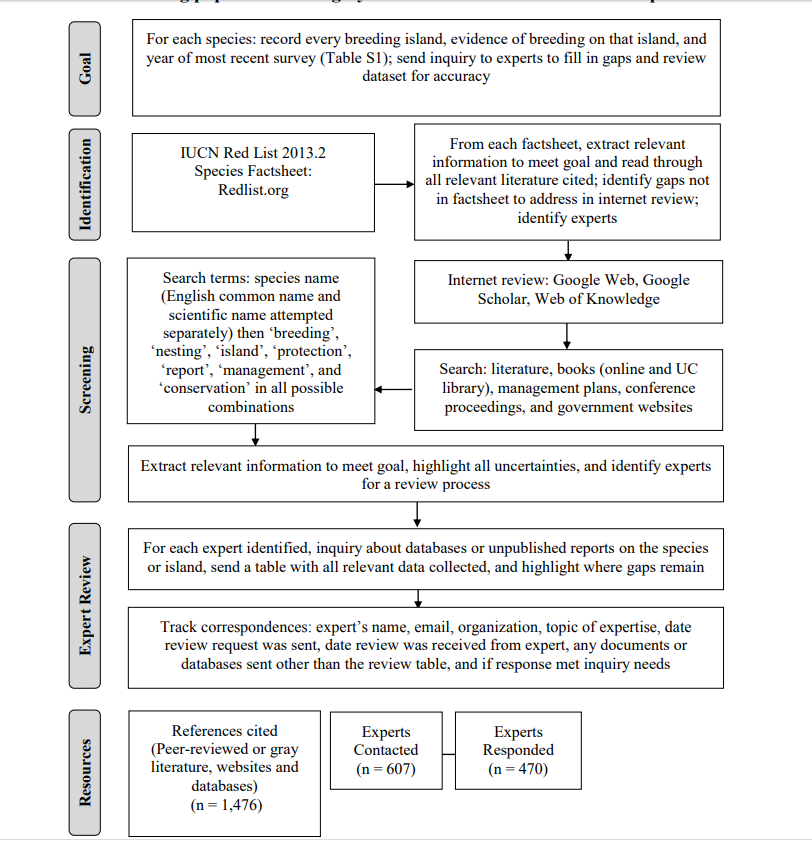
**NOTE:** This is already part of a standard literature review and if this does not get any further than that it is a bonus.

filepath <- "C:/test"  
render(file.path(filepath, "test.Rmd"))  
browseURL(file.path("file:/", filepath, "test.docx"))

# 4 Methods

We conducted a systematic review to idenifity all invasive species in NZ currently documented.

* we did this using figure 1:



Collection of NZ invasive mammal species present currently (2020)

* a generic conceptual model for all of the possible interactions and environment types in NZ. Something like below:

**Language choices**

* Invasive (either native or introduced??)
* Introduced (pre?? Dog and Kiore examples…)
* Native (NZ birds)
* Invaders
* Alien species
  + alien species are second to habitat loss… (**wilcove1998?**)

**Modelling choices**

* Functional and numerical responses…
* Observations and Processes
* Invidividual, Population and community
* Stochastic and deterministic
* Temporial and Enviromental
* Genetics?
* Demographic stochasticity
* Estimation or Prediction..

**Bayesian Modelling**

* Why?
* Bias?
* incorperate proirs
* Reproducibility is key
* Approach to creating a reproducible workflow
  1. What exactly does a scientist do? (FLow diagram)
  2. Making packages to make these steps more transperent
  3. REsulting repositorys to do this (containerizations>>!)

## 4.1 Work-flow diagram

Package options drake, workflowr

# 5 Case studies

I then apply a Bayesian modelling framework to two case studies, incorperating increased complexity from beech forests (Chapter Two) t mixed forest dynamics (Chapter Three) in NZ forests. Each case study consists of high quality CR datasets. The CR study design allows me to encorperate both the proposed ecological processes (e.g. predation) know to drive populations and the observation error (e.g. estimating population size) from the research synthesis.

## 5.1 Beech Forests

Compare the importance of bottom-up and top-down processes in regulating invasive species in New Zealand forests.

### 5.1.1 [Davidson2020-Beech-forests]

**Title:** Merky forests. What can we expect from stoat control in NZ forests? - Food vs. Predation.

**Status:** [1st draft with Richard](https://www.dropbox.com/s/m5hte0n2vyl1dt2/Davidson_2019_BeechForest_19022019.docx?dl=0)

**Abstract:** This research paper clarifies the discrepancy between two previous modelling papers; both suggesting that mesopredator release of rodents is possible in New Zealand forest systems and; several field studies that have presented limited but conflicting support for increases in mouse abundance following pest control. We used an experimental design to test the differences between the two publications and found that there is no evidence to suggest mice will become more abundant after predator removal. Additional options and WR guidelines can be found here.

## 5.2 Mixed Podocarp forests

A more complex set of interacting invasive species.

### 5.2.1 [Davidson2020-MPD-forests]

**Title:** Merky forests. What can we expect from stoat control in NZ forests? - Food vs. Predation.

**Status:** [Analysis underway](https://www.dropbox.com/s/fm57ns1jndmkmq1/Davidson_2019_mpd_manuscript.docx?dl=0)

# 6 Consistency with predictions

### 6.0.1 Intro predictions:

1. during non-mast years when little seed is available,
2. at the peak of mouse abundance (during winter and spring in mast years),
3. mouse populations should increase in size more rapidly in response to increased seed availability in mast years with stoat control than without;
4. mouse populations should decline from peak abundance more slowly in mast years with stoat control than without.

## 6.1 Methods predictions:

1. Lower abundance in non-mast years;
2. Higher peak abundance in mast years;
3. A faster rate of increase in response to high seed availability in late summer during mast years;
4. A slower rate decline from peak abundance during mast years;
5. Predictions A-D should hold only when both stoat and rats are controlled. Results predictions:

## 6.2 Manual References

* Mislan, K. A. S., Jeffrey M. Heer, and Ethan P. White. 2016. “Elevating The Status of Code in Ecology.” Trends in Ecology & Evolution 31 (1): 4–7. <https://doi.org/10.1016/j.tree.2015.11.006>.
* Ouzzani, Mourad, Hossam Hammady, Zbys Fedorowicz, and Ahmed Elmagarmid. 2016. “Rayyana Web and Mobile App for Systematic Reviews.” Systematic Reviews 5 (1): 210. <https://doi.org/10.1186/s13643-016-0384-4>.
* Stodden, Victoria, Peixuan Guo, and Zhaokun Ma. 2013. “Toward Reproducible Computational Research: An Empirical Analysis of Data and Code Policy Adoption by Journals.” PLOS ONE 8 (6): e67111. <https://doi.org/10.1371/journal.pone.0067111>.
* Team, R Core. 2018. “R: A Language and Environment for Statistical Computing.” Vienna, Austria: R Foundation for Statistical Computing.