Data Wrangling

For Stoat control in Beech forest paper

Anthony Davidson

2019-05-16

# Overview

New Zealand beech forests exhibit boom-bust dynamics (Figure 1). Where, beech trees mast in spatial synchronised but annually variable years dependant (Wardle 1991). Mice populations have invaded these systems and studies have shown that populations response numerically to changes in resources (beech seed) and mice have been modelled under a range of both functional and numerical responses (King 1983).

This report investigates the estimation of the raw observations. The data is in a similar structure as historically used for estimation of food availability (Choquenot and Ruscoe 2000; Ruscoe, Goldsmith, and Choquenot 2001; Ruscoe et al. 2005).

## Final figure/s

At some point …

# Introduction and Methods

*Generally look at thesis drafts* [*here*](%22https://www.ssnhub.com/%22)*.*

This RMarkdown file is intend to be extended on for different species of trees, fruit and seed. These files are the supporting documetation to the data-wrangling code in the Beech forest manuscript [here](%22https://www.ssnhub.com/invasive-species-research.html/%22).

Bbut there is also many other supporting resources and less formal tutorials on my [website](https://www.ssnhub.com) and the feed below:

## Invasive species notes

## Tidyverse

To aviod issues with factors and data management I have used the following pre-print publications authored by the creators of dplyr and ggplot2. Tidyverse is a group of packages that together make use a tidy approach effective.

Here is some more information about the [tidyverse approach](%22https://www.ssnhub.com/a-tidy-world.html%22)

## Meta-data

* [Current draft](https://www.dropbox.com/home/phd-drafts-anthony/beech-forest-dynamics/drafts/Davidson_2019_BeechForest.html)
* [Style sheet](https://www.dropbox.com/home/phd-drafts-anthony/beech-forest-dynamics/Styles_manual_sheet.md/)
* [Introduction](https://www.dropbox.com/sh/5h4mp67p7u6t1lj/AAAQVKS4qnvu2oQLu53JQUofa?dl=0)
* [Bayesian methods](https://www.dropbox.com/home/phd-drafts-anthony/beech-forest-dynamics/A1_full_bayesian_model.pdf)
* [Figures](https://www.dropbox.com/home/phd-drafts-anthony/beech-forest-dynamics/figs)
* [Functional response](https://www.dropbox.com/home/phd-drafts-anthony/beech-forest-dynamics/Davidson_2019_BeechForest_Appendix.pdf)
* References: *coming online soon*

All other data and resources to render project from raw data (copied from my private GIT repository) can be found on [dropbox](https://www.dropbox.com/home/phd-drafts-anthony) but will be available here when I have incorporated the co-authors feedback

## Variable structure

I have defined the following parameters that help to keep the type of variable the same throughout the joining of the raw data and the model outputs.

### Group levels

THis has been hard to do with as little manual mistakes in factor levels as possible. The levelling structure should be the same throughout this document as descripted in beech data above.

*But in short*

#### Labels/levels

*all mixed up* **still**

#month levels to get it right  
month\_levels <- c("Jan", "Feb", "Mar", "Apr", "May", "Jun","Jul", "Aug", "Sep", "Oct", "Nov", "Dec")  
  
# full set of dates  
true.date <- as.Date(c("1999-05-01","1999-08-01","1999-11-01","2000-02-01","2000-05-01","2000-08-01","2000-11-01","2001-02-01","2001-05-01","2001-08-01", "2001-11-01","2002-05-01","2002-11-01","2003-02-01","2003-05-01","2003-08-01","2003-11-01","2004-02-01","2004-05-01","2004-08-01"))  
  
# Grid labels  
labels <- c("egl M1" = "Grid one",   
 "egl M2" = "Grid two",   
 "egl MR1" = "Grid three",  
 "egl MR2" = "Grid four",  
 "hol M1" = "Grid five",  
 "hol M2" = "Grid six",  
 "hol MR1" = "Grid seven",  
 "hol MR2" = "Grid eight")  
  
# Valley labels  
labels2 <- c(egl = "Eglinton valley", hol = "Hollyford valley")

##### Full treatment

**6 levels**

treat.six.levels <- c("hol no control rats.removed",  
 "hol control rats.present",  
 "hol control rats.removed",  
 "egl control rats.present",  
 "egl control rats.removed" ,  
 "hol no control rats.present")

In this the structure of the variables for the seed dataset have to be joint and visualised we need to insure that there are the same number of replicates in this dataset as the output dataset (from the JAGS model).

##### Still coming out of wash

COntrol labels - Stoats yes/no

Conditions labals - Rats yes/no

### Seasonal factorisation

In the data the seasonal parameter is month where:

|  |  |  |
| --- | --- | --- |
| Month | Season | Factor level |
| February | Summer | 1 |
| May | Autumn | 2 |
| August | Winter | 3 |
| November | Spring | 4 |

# Seeds

This is important because the following bias are known and need to be accounted for before analysising data.

1. Seeds without the kernal have no energy value and need to be removed.
2. Different beech species produce different sized seeds and this has an effect on the “available seed”.
3. As our models build in complexity we will include a variety of food resources and therefore need to model these relationships in the future

## Importing data

The orginal data was in excel format and manually covered to csv as below. This has had the non-viable material removed and only whole kernals of seed for three species of beech tree were reported.

### Seed Measurement

The data is very clean with all the information but to make life simplier for the analysis and bindng to output data to be correct.

#### Method - metre sq (m2)

1. Bucket radius = 0.125m
2. Bucket area = 0.049087385m2
3. 4 buckets per grid = total catch area = 0.19635 m2
4. So seeds /m2 =# Seeds counted/0.1963

Therefore, if seeds were counted on a grid on a particular trip, this will equate to .

### Find extra trips and grids

### Remove additional data

## Final seed dataset

This is the full raw data reduction for the publication and in all the analysis. This data is lagged by a single trip (as a proxy for seed to “intake rate” relationship). Descriptive plots can be found in appendix.

# Mice

The mouse dataset is generated from the full CR model found [here](%22https://www.ssnhub.com/jags-bayesian-model.html%22)

## Import data

The seed data above needs to be joined to the mouse dataset below. I will step through this process to insure that there are no errors in the merging of the two datasets.

### JAGS output

The output data of the model is in JAGS format. This means that it needs to be converted into a dataframe format. There are many different ways to do this. I have choosen the following tidy approach.

We save and sort the bayesian model outputs in a .rds file and extract these results using:

*JAGSui output* (Kellner 2018).

#### Naming variable

#### Extract only N

## Restructuring dataset

### Adding grouping variables

### Reduced seed joining dataset

Check the structure of this to make sure that the variables being joined are being correctly merged.

# A tibble: 144 x 6  
# Groups: grid, year [44]  
 grid trip year month cum.seed valley  
 <chr> <dbl> <dbl> <chr> <dbl> <chr>   
 1 egl M1 1 1999 May 591. egl   
 2 egl M1 2 1999 Aug 621. egl   
 3 egl M1 3 1999 Nov 632. egl   
 4 egl M1 4 2000 Feb 20.4 egl   
 5 egl M1 5 2000 May 1324. egl   
 6 egl M1 6 2000 Aug 1701. egl   
 7 egl M1 7 2000 Nov 1716. egl   
 8 egl M1 8 2001 Feb 0 egl   
 9 egl M1 9 2001 May 0 egl   
10 egl M1 10 2001 Aug 0 egl   
# ... with 134 more rows

### Joining seed and mice

# Stoat control

**control variable**

# Rat removal

**treatment variable**

# More grouping options

# Re-leveling dataset

Just a final check that everything has worked.

# Saved datasets

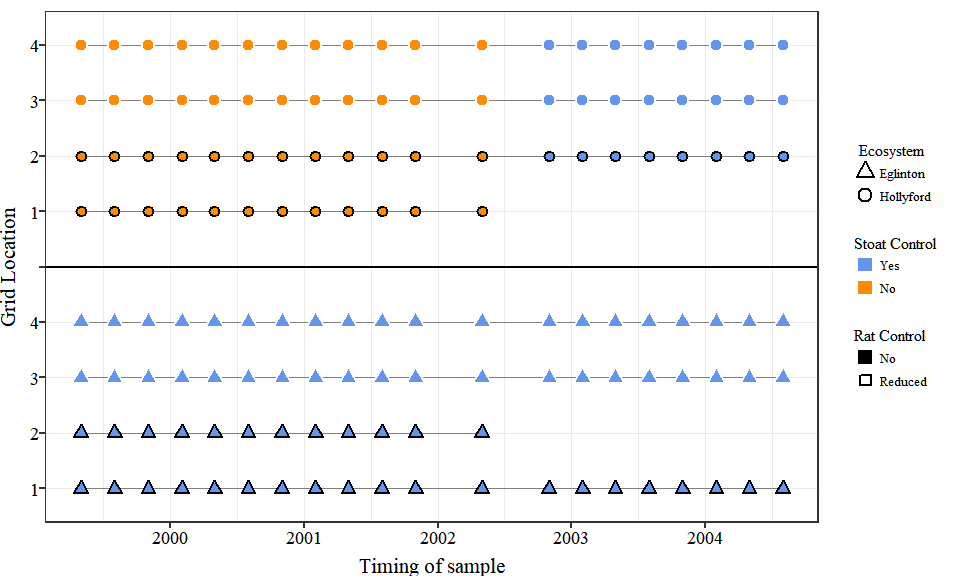
I will save the final dataset for the analysis of the Beech foret data for the publication as so: The final plots. All extended plots are in the Rcode/figures/. These plots focus on simplifying the general trends and confirming any previous predictions.

## Study design

CR-design is complex and when treatments and conditions change through time it can be very hard to explain the data collection methods well. We have therefore constructed a diagram of the study design process.

### Data

### Plot



### Saving

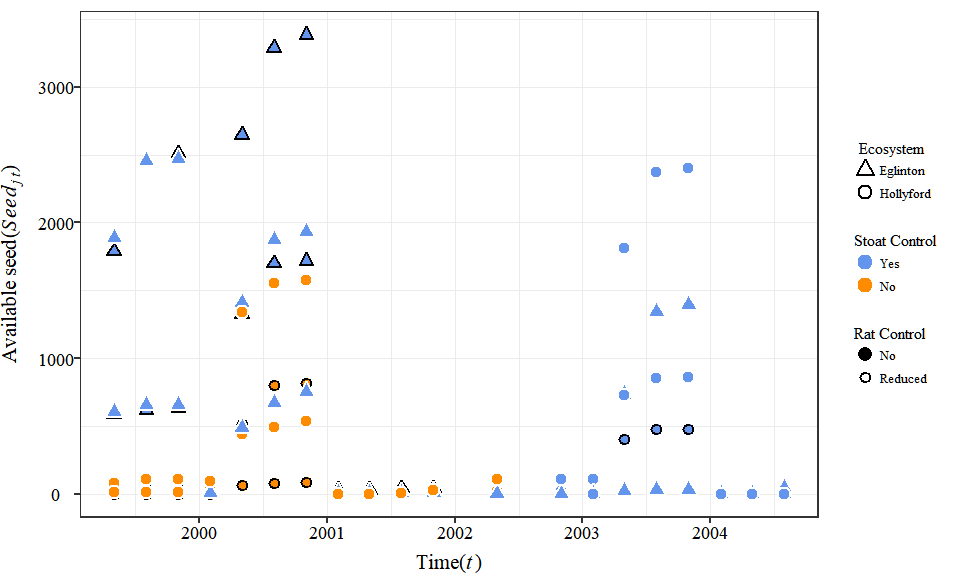
png   
 2

## Seed

### Data

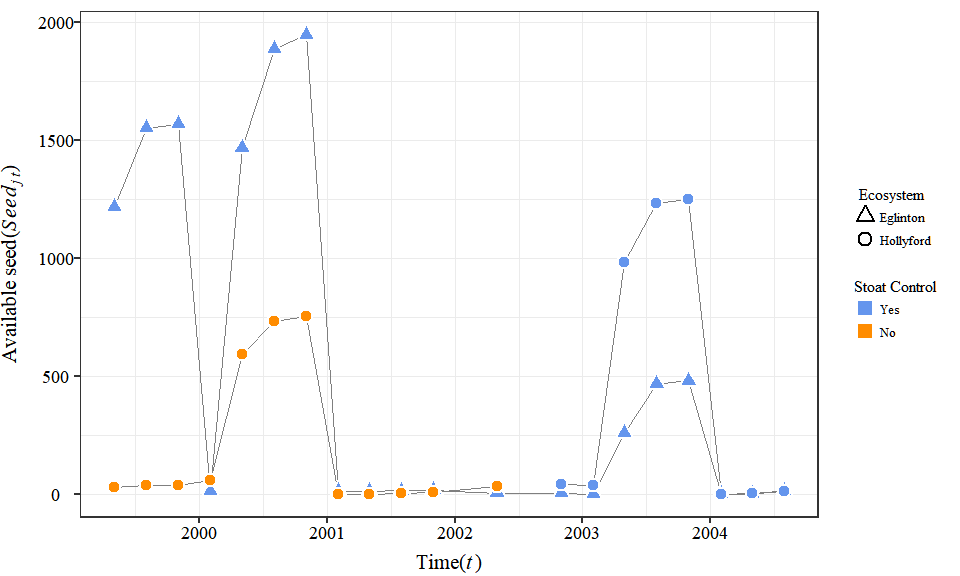
### Plot

#### All seed data

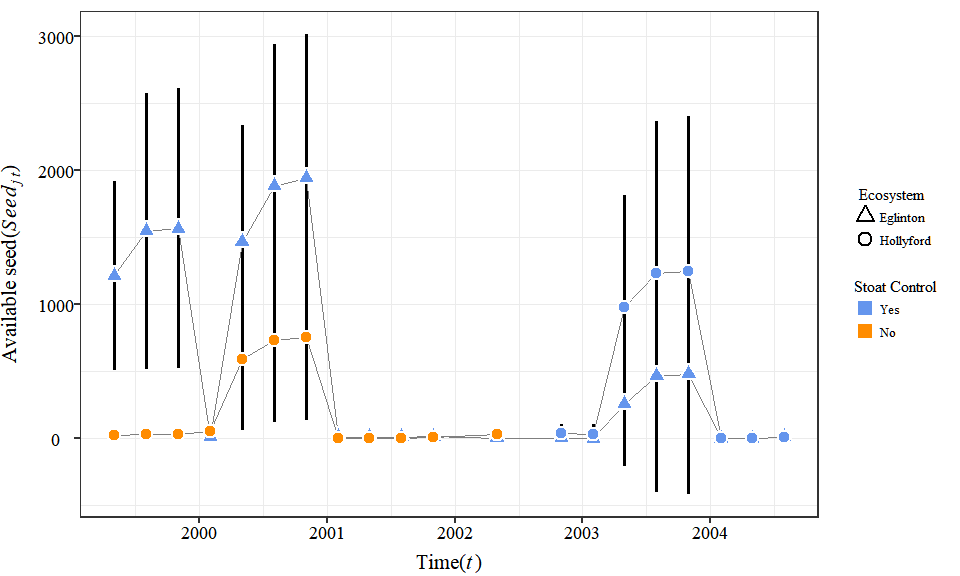


#### Summarising

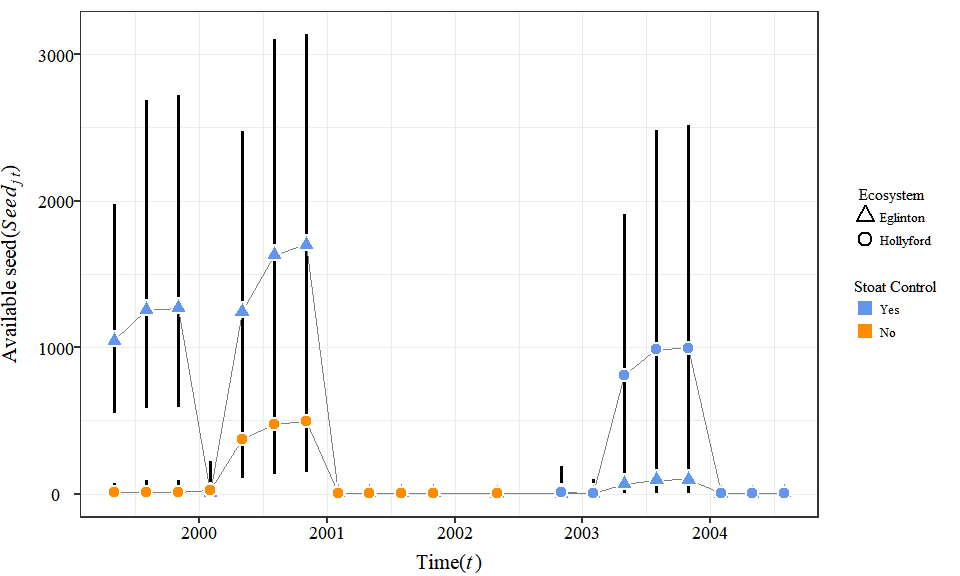
I have done this by grouping dataset into the three key treatments (with and without stoat control in the Hollyford Valley and stoat control in the Eglinton Valley).



##### Raw confidence intervals

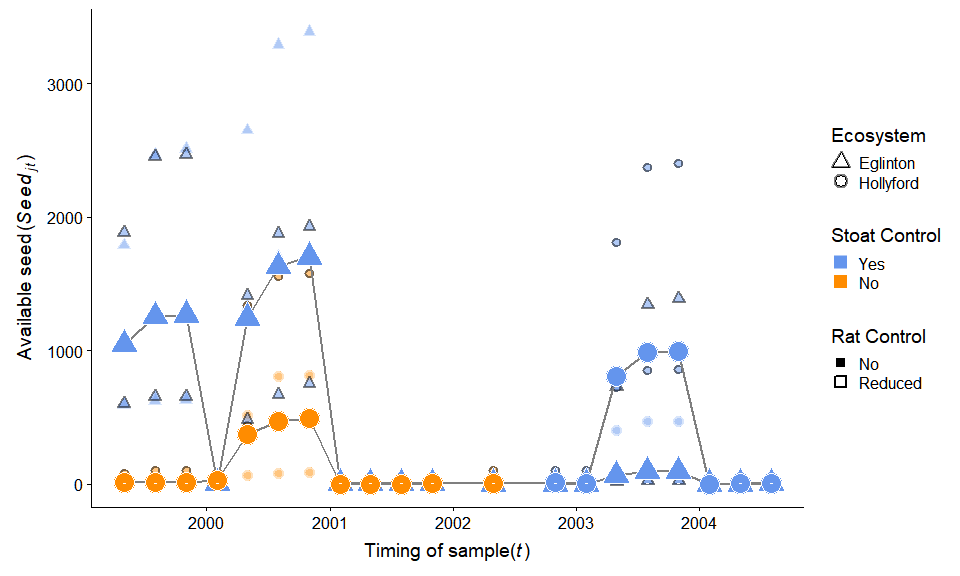


##### Log transformation to remove less and zero seed…

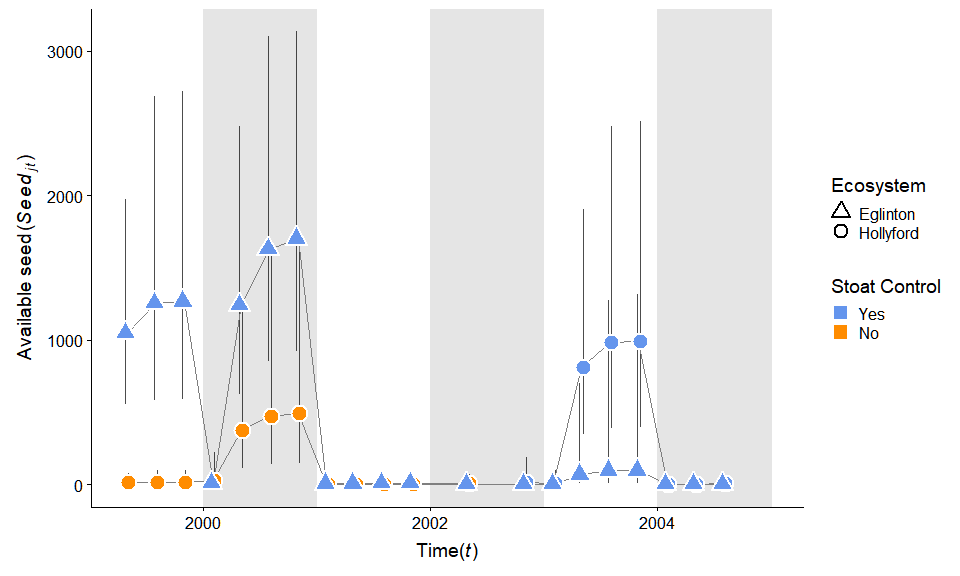


##### Combing plots

Orginal data seems too messy so have reverted to the summary plot above.



##### Final plot



### Saving

#### Data

#### Figure

png   
 2

## Mice

### Data

Control

Valley

Date

mean.s

sd.s

se.s

lcl.s

ucl.s

gp.treat

N

Rats

Yes

Eglinton

1999-05-01

52.52743

27.308946

26.762767

25.76466

79.29020

Eglinton Yes

52.52743

Full

Yes

Eglinton

1999-08-01

165.56351

59.608474

58.416304

107.14721

223.97981

Eglinton Yes

165.56351

Full

Yes

Eglinton

1999-11-01

105.38420

36.525855

35.795338

69.58886

141.17954

Eglinton Yes

105.38420

Full

Yes

Eglinton

2000-02-01

25.23375

6.496038

6.366117

18.86763

31.59986

Eglinton Yes

25.23375

Full

Yes

Eglinton

2000-05-01

55.47290

27.866482

27.309153

28.16375

82.78205

Eglinton Yes

55.47290

Full

Yes

Eglinton

2000-08-01

128.91481

33.675592

33.002080

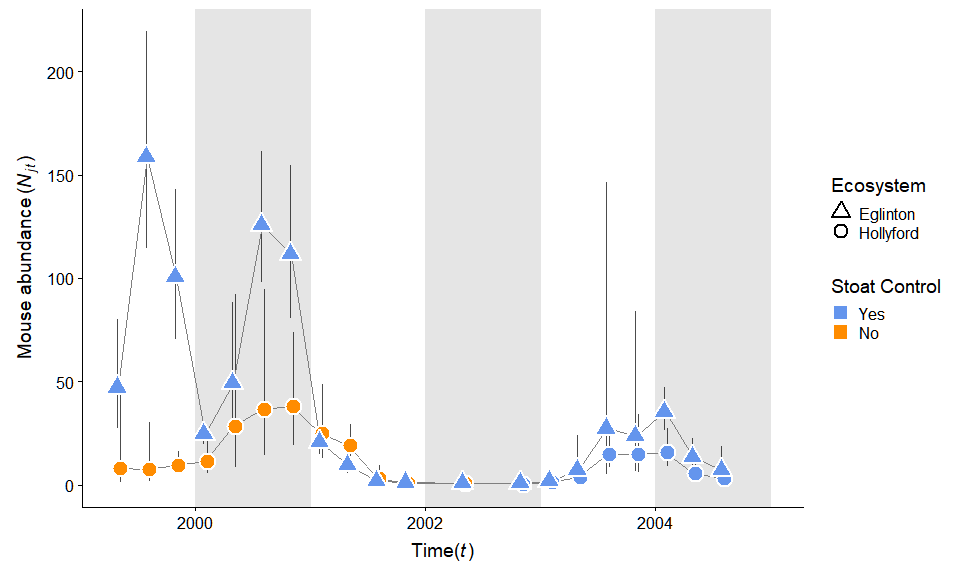
95.91273

161.91689

Eglinton Yes

128.91481

Full



### Saving

png   
 2

## Rats

### Data

Observations: 144  
Variables: 35  
$ N <dbl> 85.185920, 141.346453, 125.542907, 23.769273, 8...  
$ se.N <dbl> 17.258336, 24.813696, 15.717429, 1.661358, 18.5...  
$ lcl.N <dbl> 53, 103, 100, 22, 57, 88, 93, 25, 13, 0, 1, 1, ...  
$ ucl.N <dbl> 114, 200, 162, 28, 125, 179, 164, 43, 26, 5, 5,...  
$ var <chr> "N[1,1]", "N[2,1]", "N[3,1]", "N[4,1]", "N[5,1]...  
$ grid <chr> "egl M1", "egl M1", "egl M1", "egl M1", "egl M1...  
$ trip <chr> "1", "2", "3", "4", "5", "6", "7", "8", "9", "1...  
$ grid.n <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,...  
$ trip.no.x <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, ...  
$ valley.x <chr> "egl", "egl", "egl", "egl", "egl", "egl", "egl"...  
$ control <chr> "control", "control", "control", "control", "co...  
$ Valley <fct> Eglinton, Eglinton, Eglinton, Eglinton, Eglinto...  
$ year <dbl> 1999, 1999, 1999, 2000, 2000, 2000, 2000, 2001,...  
$ month <chr> "May", "Aug", "Nov", "Feb", "May", "Aug", "Nov"...  
$ cum.seed <dbl> 590.781767, 621.339445, 631.525337, 20.371785, ...  
$ seed.account.N <dbl> 6.9352044, 4.3958616, 5.0303546, 0.8570639, 14....  
$ log.seed <dbl> 2.7714271, 2.7933289, 2.8003908, 1.3090291, 3.1...  
$ valley.rep <chr> "egl", "egl", "egl", "egl", "egl", "egl", "egl"...  
$ grid.rats <fct> M1, M1, M1, M1, M1, M1, M1, M1, M1, M1, M1, M1,...  
$ Conditions <chr> "rats.removed", "rats.removed", "rats.removed",...  
$ grouping.1 <chr> "treat.highN", "treat.highN", "treat.highN", "t...  
$ grouping.2 <chr> "treat.highN", "treat.highN", "treat.highN", "t...  
$ grouping.3 <chr> "treat.highN", "treat.highN", "treat.highN", "t...  
$ grouping.4 <chr> "treat.highN", "treat.highN", "treat.highN", "t...  
$ true.date <date> 1999-05-01, 1999-08-01, 1999-11-01, 2000-02-01...  
$ treat.six <chr> "egl control rats.removed", "egl control rats.r...  
$ Rats <fct> Reduced, Reduced, Reduced, Reduced, Reduced, Re...  
$ Control <fct> Yes, Yes, Yes, Yes, Yes, Yes, Yes, Yes, Yes, Ye...  
$ Date <date> 1999-05-01, 1999-08-01, 1999-11-01, 2000-02-01...  
$ Treatments <fct> egl control rats.removed, egl control rats.remo...  
$ Prediction <chr> NA, NA, NA, NA, NA, "A", NA, NA, NA, "B", NA, N...  
$ valley.y <chr> "egl", "egl", "egl", "egl", "egl", "egl", "egl"...  
$ trip.no.y <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, NA, NA, NA, NA, NA, ...  
$ n <dbl> 1, 4, 14, 2, 5, 2, 5, 3, 2, NA, NA, NA, NA, NA,...  
$ rat.mna <dbl> 1, 4, 14, 2, 5, 2, 5, 3, 2, NA, NA, NA, NA, NA,...

Control

Valley

Date

mean.rat

sd.rat

se.rat

lcl.rat

ucl.rat

gp.treat

N

Rats

Yes

Eglinton

1999-05-01

1.333333

0.5773503

0.5658033

0.7675301

1.899137

Eglinton Yes

1.333333

Full

Yes

Eglinton

1999-08-01

2.250000

1.2583057

1.2331396

1.0168604

3.483140

Eglinton Yes

2.250000

Full

Yes

Eglinton

1999-11-01

8.000000

4.5460606

4.4551394

3.5448606

12.455139

Eglinton Yes

8.000000

Full

Yes

Eglinton

2000-02-01

1.500000

0.7071068

0.6929646

0.8070354

2.192965

Eglinton Yes

1.500000

Full

Yes

Eglinton

2000-05-01

2.666667

2.0816660

2.0400327

0.6266340

4.706699

Eglinton Yes

2.666667

Full

Yes

Eglinton

2000-08-01

1.500000

0.5773503

0.5658033

0.9341967

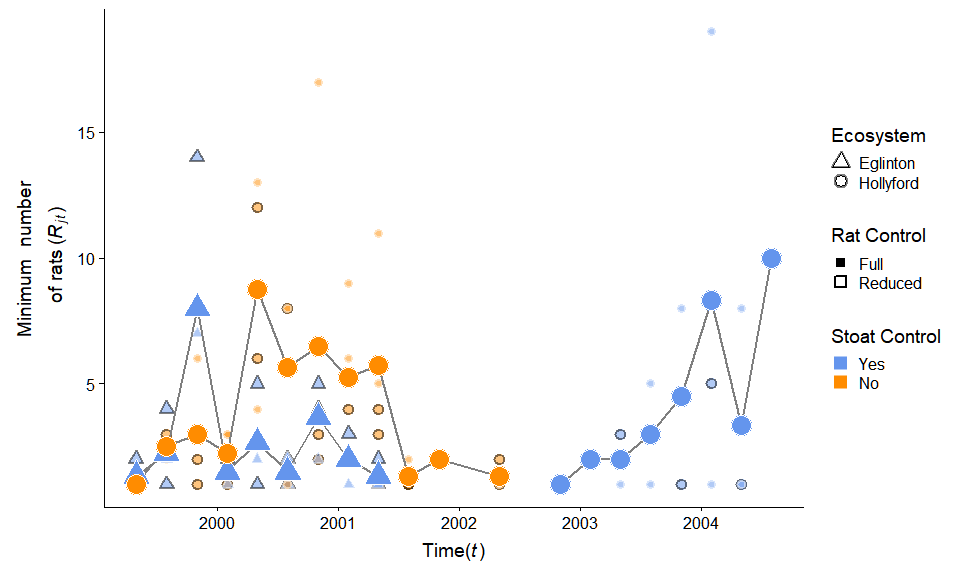
2.065803

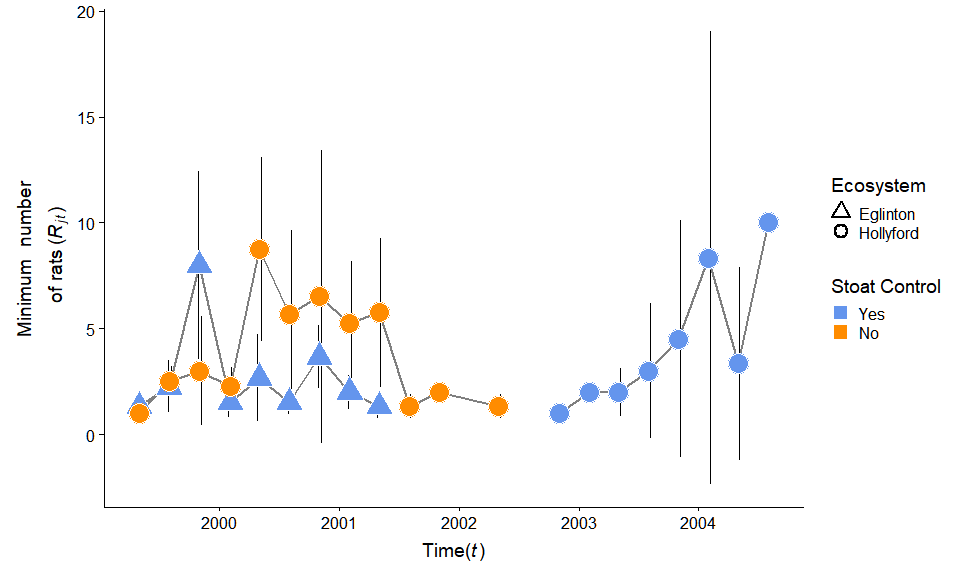
Eglinton Yes

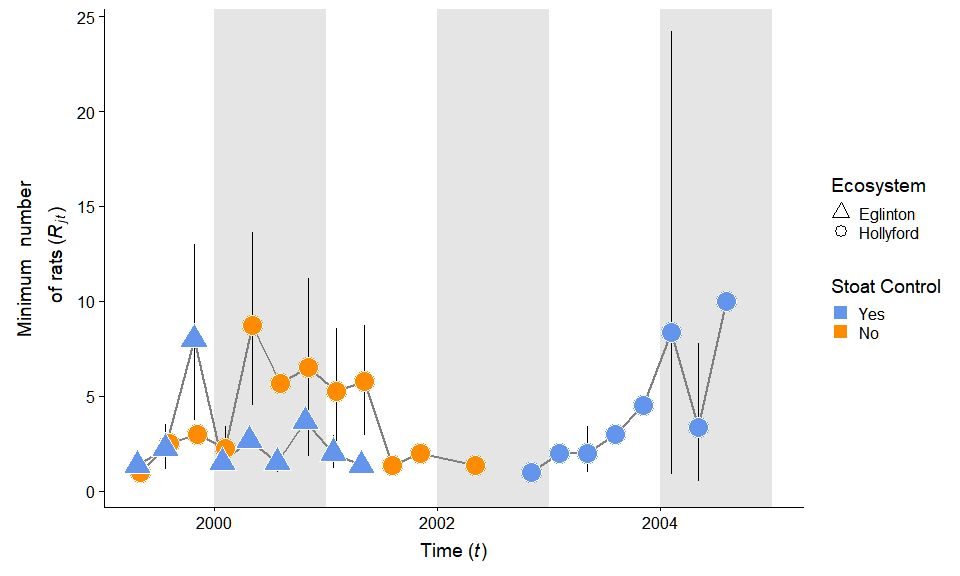
1.500000

Full

### Plot







### Saving

png   
 2

# References

Choquenot, David, and Wendy A Ruscoe. 2000. “Mouse Population Eruptions in New Zealand Forests: The Role of Population Density and Seedfall.” *Journal of Animal Ecology* 69: 1058–70.

Kellner, Ken. 2018. *jagsUI: A Wrapper Around ’Rjags’ to Streamline ’JAGS’ Analyses*.

King, Carolyn M. 1983. “The Relationships Between Beech (Nothofagus Sp.) Seedfall and Populations of Mice (Mus Musculus), and the Demographic and Dietary Responses of Stoats (Mustela Erminea), in Three New Zealand Forests.” *Journal of Animal Ecology* 52 (1): 141–66.

Ruscoe, Wendy A, Joseph S Elkinton, David Choquenot, and Robert B Allen. 2005. “Predation of Beech Seed by Mice: Effects of Numerical and Functional Responses.” *Journal of Animal Ecology* 74: 1005–19. <https://doi.org/10.1111/j.1365-2656.2005.00998.x>.

Ruscoe, Wendy A, Ruth Goldsmith, and David Choquenot. 2001. “A Comparison of Population Estimates and Abundance Indices for House Mice Inhabiting Beech Forests in New Zealand.” *Wildlife Research* 28: 173–78.

Wardle, P. 1991. *Vegetation of New Zealand*. Cambridge University Press.