

# Four principles for improved statistical ecology: a worked example of a wetland burning experiment

10/30/23

## Table of contents

<b>1 Principle 1. First, define a focused research question, then plan sampling and analysis to answer it</b>	<b>2</b>
1.1 1A – Define your research question . . . . .	2
1.1.1 Predictions . . . . .	3
1.2 1B – Match data collection to research aims . . . . .	3
1.2.1 Field Sampling . . . . .	3
1.2.2 Manipulative Experiment . . . . .	5
1.3 1C – Plan analysis and consider registration . . . . .	6
1.3.1 Model . . . . .	6
1.3.2 Primary Effects of interest (with confidence intervals) . . . . .	6
1.3.3 Multiple testing . . . . .	6
<b>2 Principle 2. Develop a model that accounts for the distribution and dependence of your data</b>	<b>7</b>
2.1 2A – Model dependence . . . . .	7
2.1.1 Read in biomass data . . . . .	7
2.1.2 Mixed model . . . . .	7
2.2 2B – Check assumptions . . . . .	7
2.2.1 Residual v.s. fitted plot (marginal) . . . . .	7
2.2.2 Scale-location plot . . . . .	8
2.3 3A – Replace statistical significance with ecological relevance by emphasising effect sizes . . . . .	9
2.3.1 Results . . . . .	13
2.3.2 Discussion . . . . .	13

```
library(sjPlot)
library(ggplot2)
library(lme4)
library(lmerTest)
library(dplyr)
library(here)
library(emmeans)
library(randomizr)
library(tidyverse)

options(digits = 3, scipen = 5)
```

Underground mining is known to disrupt surface and groundwater flows which may affect nearby swamp communities. The researchers wanted to examine how differing water availability affected swamp plant communities, both alone and in combination with a fire disturbance. For this study, researchers planned to collect mesocosms from multiple swamps, then randomise them to water and fire treatments in a glasshouse to infer if water availability causes changes in biomass. Mesocosms, for the purpose of this study, are a column of soil and plants, collected by hammering PVC pipe (diameter of 150 mm and a depth of 250 mm) to ground level and extracted with trenching shovels. They would then placed in tubs in a glasshouse, and tub water levels were manipulated to simulate different levels of groundwater availability. A fire event was to be simulated by sequentially applying biomass removal (clipping), heat and smoke to half of the mesocosms in each water treatment after 20 months (see Figure 1 main manuscript).

The example below is a simplified portion of the full experiment and analysis in (Mason *et al.*, 2022)

## 1 Principle 1. First, define a focused research question, then plan sampling and analysis to answer it

### 1.1 1A – Define your research question

#### **i** FINER

- **Feasible** - There is enough funding and expertise to conduct a 3.5 year glasshouse experiment with about 250 mesocosms.
- **Interesting** - Swamp communities are likely to be sensitive to changes in the hydrological gradient and fire regime (Keith *et al.*, 2009; Keith *et al.*, 2022) and recent longwall underground coal extraction has caused subsidence, diminishing

water resources (Mason *et al.*, 2021).

- **Novel** - Current evidence is from observational studies, no controlled empirical manipulations have been undertaken, and this would add substantially to the evidence base.
- **Ethical** - Extraction of small mesocosms is unlikely to have a negative effect on the health of the ecosystem.
- **Relevant** - Underground coal mining is active in the Southern Coalfield of the Sydney Basin. Understanding the compounding effects of hydrological and fire disturbance will assist conservation of endangered upland swamp communities.

### 1.1.1 Predictions

- Mesocosms with low water resource availability will have reduced biomass over time compared with mesocosms with high water resource availability.
- Low water resource availability will compound fire effects (interaction).

Note - We will assume burnt vegetation will have greater reduction in biomass than unburnt vegetation, and this is not of interest.

#### **i** PICO

- **Population** - Upland swamp plant communities of the Sydney Basin bioregion, Australia
- **Intervention** - Two interventions: water availability (low, medium, high) and fire (yes, no), factorial.
- **Comparison** - The control group is high water unburnt mesocosms, as this group is considered to reflect the natural (unburnt) state of the swamps, though differences between all treatments are of interest.
- **Outcome** - Biomass

## 1.2 1B – Match data collection to research aims

The sampling is summarised in Fig. 1

### 1.2.1 Field Sampling

1. Researchers collected 31 mesocosms comprising above-ground and root biomass along with soil material from both cyperoid heath and Ti-tree thicket vegetation communities

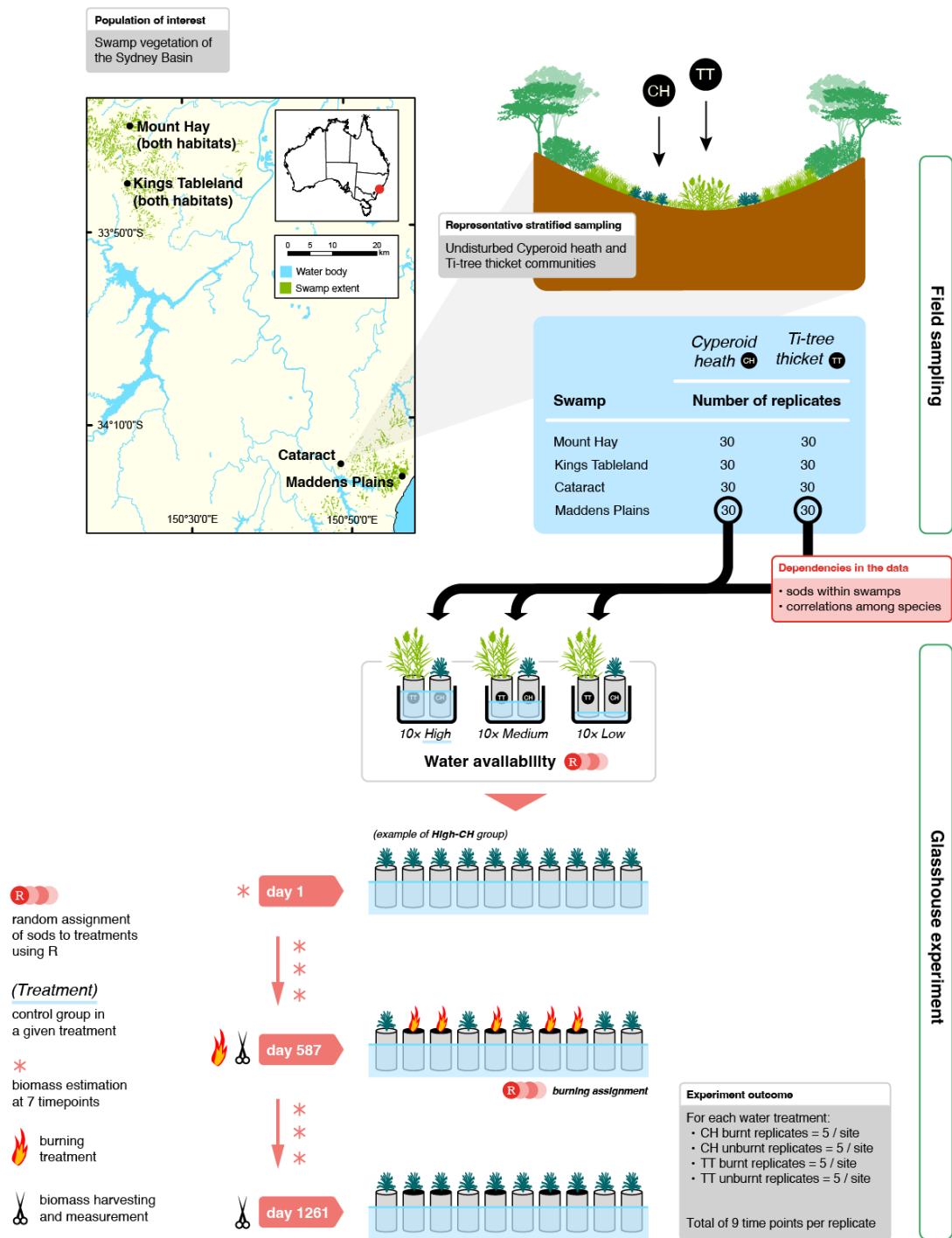


Figure 1: Schematic illustration of the wetland sampling strategy.

in each of four upland swamps in the Sydney Basin.

2. Above-ground vegetation was clipped to approximately 10% of the original biomass and mesocosms were bagged and then transported to the glasshouse. Mesocosms from each site were allowed to acclimatise for a 28 days, with regular watering prior to treatment allocation.

### 1.2.2 Manipulative Experiment

Mesocosms were randomised to high, medium or low water availability treatment levels, and burnt or unburnt fire treatment levels.

To do this we need the `randomizr` package, and a list of mesocosms.

```
mesocosms <- read.csv(here("data", "mesocosms.csv"))
head(mesocosms)
```

	Mesocosm	swamp	veg
1	CAT_CH_01	CA	ch
2	CAT_CH_02	CA	ch
3	CAT_CH_03	CA	ch
4	CAT_CH_04	CA	ch
5	CAT_CH_05	CA	ch
6	CAT_CH_06	CA	ch

1. The simplest way to randomize is complete random assignment, where each mesocosm is allocated to each treatment with some probability. If we want equal probabilities for each treatment then with 6 treatments (3 water x 2 fire) we use `prob_each = rep(1/6,6)`

```
mesocosms$treatment = complete_ra(N = nrow(mesocosms), prob_each = rep(1/6,6),
                                   conditions = c("H_u", "M_u", "L_u", "H_b", "M_b", "L_b"))
head(mesocosms)
```

	Mesocosm	swamp	veg	treatment
1	CAT_CH_01	CA	ch	L_b
2	CAT_CH_02	CA	ch	L_b
3	CAT_CH_03	CA	ch	L_u
4	CAT_CH_04	CA	ch	L_b
5	CAT_CH_05	CA	ch	M_b
6	CAT_CH_06	CA	ch	L_b

We can check how many mesocosms are in each treatment.

```
table(mesocosms$treatment)
```

```
H_u M_u L_u H_b M_b L_b
42  41  41  42  41  41
```

2. For the duration of the experiment, water levels of treatment tubs were maintained at 70, 155 and 240 mm below the mesocosm surface for the high, medium and low water availability levels, respectively.
3. A single fire event was simulated in mesocosms allocated to the burnt treatment by sequentially clipping, heating and applying smoked water to these mesocosms.
4. Biomass was measured at 2 years (immediately prior to burning) for burnt sods and 4 years (conclusion of experiment) for all sods.

### 1.3 1C – Plan analysis and consider registration

#### 1.3.1 Model

- (Generalised) linear mixed model with biomass as outcome, fixed effects for swamp and vegetation type to control for these, and random effect for mesocosm (swamp) (to account for repeated measures), as well as fixed effects for time (categorical), water (low, medium, high), and fire (burnt, unburnt).
- Outcome type (normal, log transformed etc.), and relationship type (linear, quadratic) to be determined with reference to residual plots.

#### 1.3.2 Primary Effects of interest (with confidence intervals)

- Effect of **water availability** on change in biomass on unburnt mesocosms, controlling for swamp and vegetation type. Planned contrasts of change between water availability levels at 2 years (immediately prior to burning) to 4 years (conclusion of experiment).
- Effect of **burning** on difference in biomass among **water availability levels**, controlling for swamp, vegetation type and water availability treatment. Planned contrast between burnt and unburnt at 4 years (conclusion of experiment),

#### 1.3.3 Multiple testing

- Primary effects p-values and confidence intervals will be adjusted for multiple testing using the multivariate t distribution (`method = "mvt"` in `emmeans`)

## 2 Principle 2. Develop a model that accounts for the distribution and dependence of your data

### 2.1 2A – Model dependence

We have three sources of dependence

1. Mesocosms come from four swamps, and different swamps may have different biomass. We therefore include a fixed effect of swamp.
2. Mesocosms are from two vegetation communities (Ti-tree thicket and Cyperoid heath), and these communities may differ in biomass. So we will include a fixed effect for vegetation community.
3. Mesocosms will be measured multiple times, and these repeated measurements will be correlated. So we will include a random intercept for mesocosm.

#### 2.1.1 Read in biomass data

```
bio_data <- read.csv(here("data","wetland_biomass.csv")) %>%  
  mutate(water = factor(water, levels = c("H","M","L")))
```

#### 2.1.2 Mixed model

```
bio_live <- lmer(log(biomass + 1) ~ swamp + veg + factor(days) * water + factor(days) * fi  
  water*fire +  
  (1 | Mesocosm),  
  data = bio_data)
```

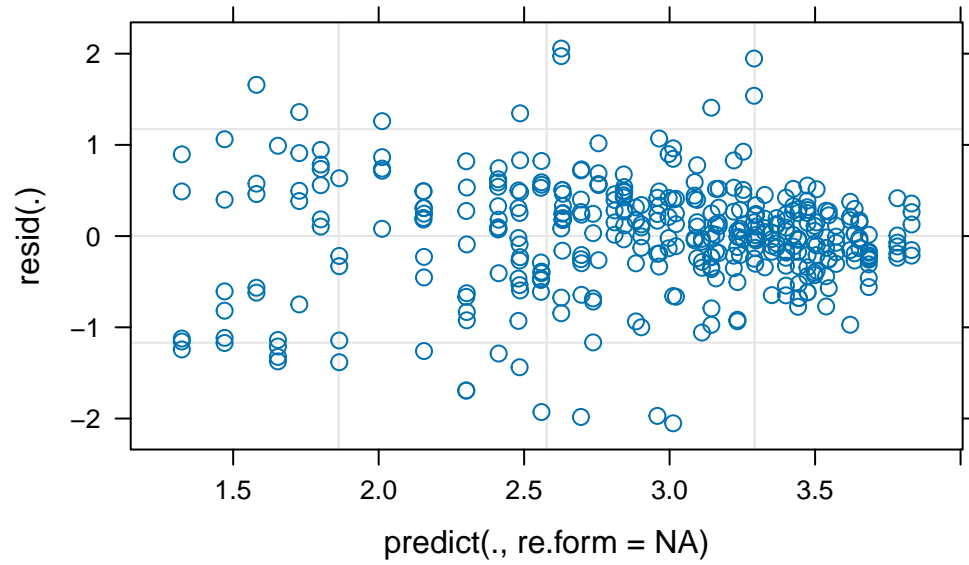
The warning about rank deficiency means is expected. It is because we don't have measurements of burnt swamps prior to the burning treatment at 2 years.

### 2.2 2B – Check assumptions

#### 2.2.1 Residual v.s. fitted plot (marginal)

Note - `re.form = NA` gives marginal residuals.

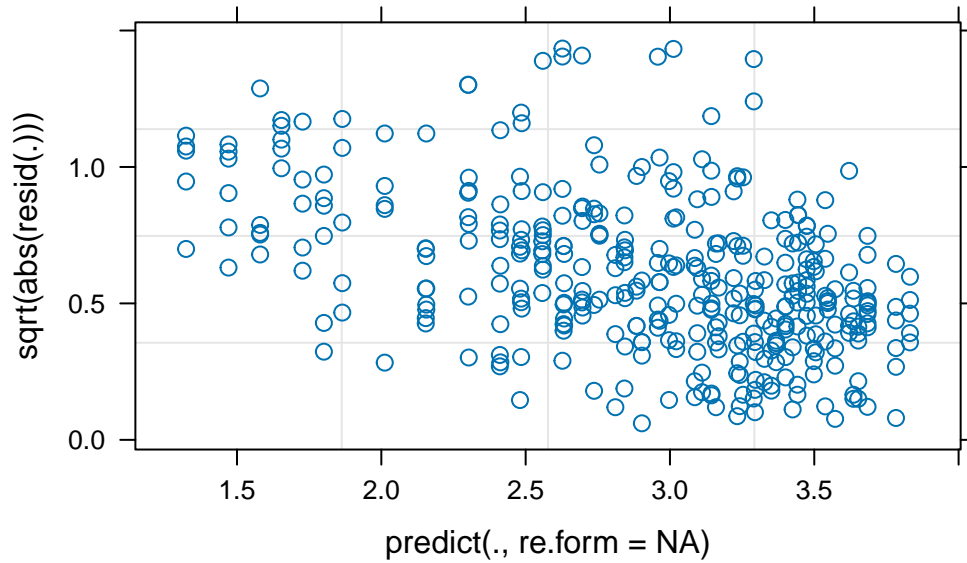
```
plot(bio_live, resid(. ) ~ predict(., re.form = NA))
```



### 2.2.2 Scale-location plot

```
plot(bio_live, sqrt(abs(resid(. ))) ~ predict(., re.form = NA))
```



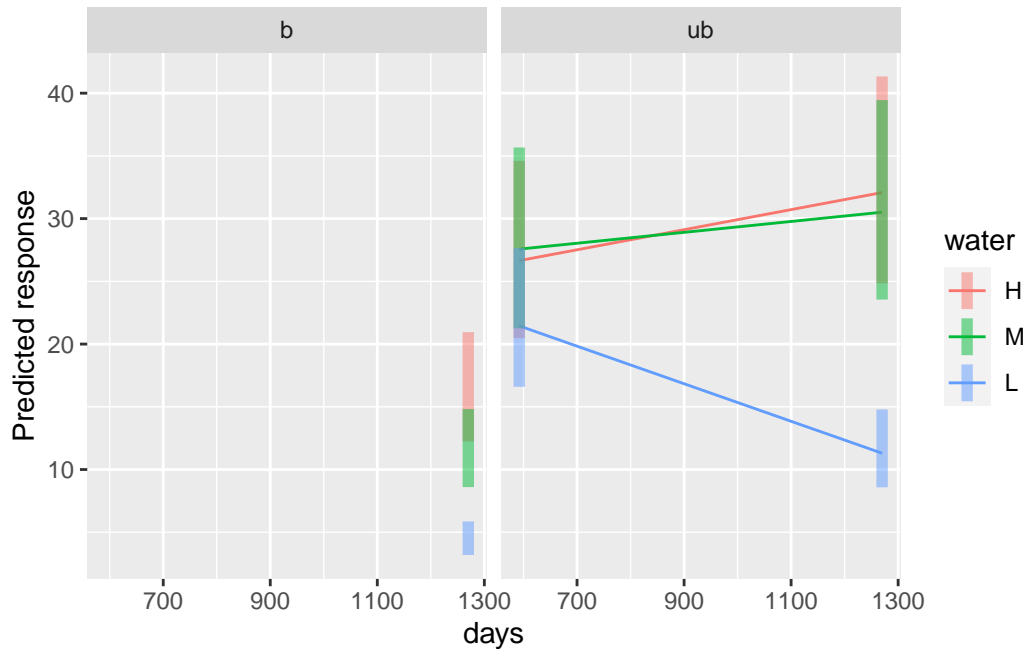


Assumptions of linearity and constant variance are approximately satisfied, though there is some hint of variance reducing with the mean.

## 2.3 3A – Replace statistical significance with ecological relevance by emphasising effect sizes

Start by plotting model estimates. The easiest way is to use `emmip` from `emmeans`.

```
emmmeans_plot = emmip(bio_live, ~ water ~ days | fire ,
                       CIs = TRUE, type = "response") # always include confidence intervals
emmmeans_plot
```



We can take the data from the previous plot and use them to make a prettier plot.

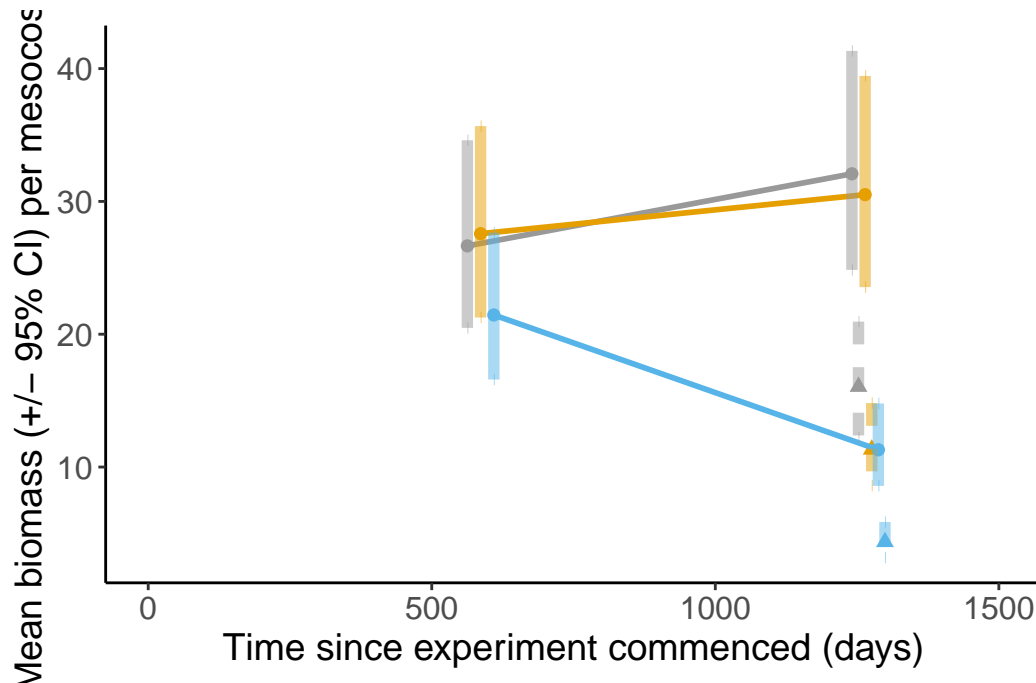
```
pos = position_dodge(width=70) # dodge water

#colour blind friendly palette
cbPalette <- c("#999999", "#E69F00", "#56B4E9", "#009E73", "#F0E442", "#0072B2", "#D55E00")

plot_df <- emmeans_plot$data %>%
  mutate(fire = relevel(fire,"ub"))

ggplot(plot_df, aes(days, yvar, color = water, shape = fire, linetype = fire),
       position = position_dodge2(width = 0.1)) +
  geom_point(position = pos, size = 2) +
  geom_path(position = pos, size = 1) +
  geom_errorbar(aes(ymin = LCL, ymax = UCL), position = pos, alpha = 0.5, width = 0, size = 1) +
  theme_classic() +
  xlim(0,1500) +
  xlab("Time since experiment commenced (days)") +
  ylab("Mean biomass (+/- 95% CI) per mesocosm (g)") +
  scale_colour_manual(values=cbPalette) +
  theme(legend.position = "none",
       axis.text = element_text(size = 12),
```

```
axis.title = element_text( size = 14))
```



Next we calculate the desired effects and their confidence intervals, again using `emmeans`.

Estimate the marginal means for all water treatments, at just the times of interest (592 days and 1270 days), for unburnt mesocosms using the `emmeans` function.

```
# water by time changes
em_water <- emmeans(bio_live, ~ water + days , type = "response",
  at = list(days = c( 1270, 592), fire = "ub"))
```

Note - You may get a warning, but it can be ignored.

Calculate all pairwise interactions between the two.

```
water_changes <- contrast(em_water, interaction = "pairwise")
water_changes
```

water_pairwise	days_pairwise	estimate	SE	df	t.ratio	p.value
H - M	1270 - 592	0.081	0.254	341	0.320	0.7490
H - L	1270 - 592	0.781	0.253	341	3.093	0.0020
M - L	1270 - 592	0.700	0.253	341	2.771	0.0060

Results are averaged over the levels of: swamp, veg  
 Degrees-of-freedom method: kenward-roger

Also estimate the means for all water and fire treatment combinations at the last time point (1270 days).

```
# fire by water changes
em_fire <- emmeans(bio_live, ~ water + fire , type = "response",
  at = list(days = c(1270)))
```

Again calculate all pairwise interactions.

```
fire_changes <- contrast(em_fire, interaction = "pairwise")
fire_changes
```

water_pairwise	fire_pairwise	estimate	SE	df	t.ratio	p.value
H - M	b - ub	0.276	0.254	341	1.087	0.2780
H - L	b - ub	0.168	0.253	342	0.663	0.5080
M - L	b - ub	-0.108	0.253	342	-0.428	0.6690

Results are averaged over the levels of: swamp, veg  
 Degrees-of-freedom method: kenward-roger

Each of the above analyses will answer our primary questions, however we would like to control for multiple testing: for this we can use the `rbind` function.

```
# combine these for multiple testing adjustment
combined_effects <- rbind(water_changes, fire_changes, adjust = "mvt")

#confint(combined_effects) #confidence intervals
#summary(combined_effects) # for p-values

#combined table
cbind(summary(combined_effects)[,c(1:4,8)], confint(combined_effects)[, 7:8]) %>%
  mutate_at(c(4, 6, 7), round, 1) %>%
  mutate_at(5, round,3)
```

	water_pairwise	days_pairwise	fire_pairwise	estimate	p.value	lower.CL	upper.CL
1	H - M	1270 - 592	.	0.1	0.995	-0.6	0.7
2	H - L	1270 - 592	.	0.8	0.011	0.1	1.4

3	M - L	1270 - 592	.	0.7	0.030	0.0	1.4
4	H - M	.	b - ub	0.3	0.716	-0.4	0.9
5	H - L	.	b - ub	0.2	0.933	-0.5	0.8
6	M - L	.	b - ub	-0.1	0.986	-0.8	0.5

Note: these are slightly different from the analysis in the manuscript, as multiple testing adjustments were applied to all 6 contrasts simultaneously.

### 2.3.1 Results

Differences in biomass between high and low water unburnt mesocosms more than doubled (relative change = 2.2 (95% CI: 1.1 – 4.2)) between two and four years. Similarly, differences in biomass between unburnt low and medium water mesocosms doubled (relative change = 2.0; 95% CI: 1.1 – 3.9), but there was no evidence of differences in biomass changes between high and medium water mesocosms (relative change = 1.1; 95% CI: 0.6 – 2.1). We did not find any evidence of an interaction between fire and water treatments.

### 2.3.2 Discussion

In this experimental glasshouse simulation, reduced water availability leads to large reductions in biomass in wetland communities relative to communities simulating undisturbed swamps (high water availability), however we did not find any evidence of a synergistic effect of fire disturbance on biomass.

The intent of the experiment was to demonstrate a causal relationship between water availability and richness, biomass and composition of wetland species, and every effort was made to meet the causal assumptions for experiments (Kimmel *et al.*, 2021). Some interference may have happen as above ground biomass increased over the course of the experiment, potentially shading neighbouring mesocosms. The amount of shading depends on height (and so biomass), which we know differed between treatments, and could have led to bias over time. To combat this mesocosms were infrequently re-randomised across tubs within water treatment levels. Some mesocosms subsided in the PVC casing after collection. This was problematic as the treatment effect relied on water depth in the tubs. To limit any effect (multiple treatment), the lower section of the subsided mesocosms was packed with a 50:50 mix of nursery-sourced river sand and peat moss to facilitate capillary action up the casing. In total, 101 mesocosms required packing, and these were randomised to all treatments. Soil moisture was measured throughout the experiment, and clear differences between water treatment levels were found, as intended by the experimental design. However, this glasshouse setup does not directly mimic field moisture or nutrient transport profiles and may affect soil moisture, species richness and composition outcomes.

- Keith D, Benson D, Baird I, Watts L, Simpson C, Krogh M, *et al.* (2022). [Interactions between anthropogenic stressors and recurring perturbations mediate ecosystem resilience or collapse](#). *Conservation Biology* **37**: e13995.
- Keith DA, Rodoreda S, Bedward M (2009). [Decadal change in wetland-woodland boundaries during the late 20th century reflects climatic trends](#). *Global Change Biology* **16**: 2300–2306.
- Kimmel K, Dee LE, Avolio ML, Ferraro PJ (2021). [Causal assumptions and causal inference in ecological experiments](#). *Trends in Ecology & Evolution* **36**: 1141–1152.
- Mason TJ, Krogh M, Popovic GC, Glamore W, Keith DA (2021). [Persistent effects of underground longwall coal mining on freshwater wetland hydrology](#). *Science of The Total Environment* **772**: 144772.
- Mason TJ, Popovic GC, McGillicuddy M, Keith DA (2022). [Effects of hydrological change in fire-prone wetland vegetation: An empirical simulation](#). *Journal of Ecology* **111**: 1050–1062.