

LTER_allometry

CML & CAK

2022-09-22

Introduction

We selected the and_vertebrates database from the LTER site (https://lter.github.io/lterdatasampler/reference/and_vertebrates.html):

The dataset includes count and size data for cutthroat trout and salamanders in clear cut or old growth sections of Mack Creek, Andrews Forest LTER.

```
## here() starts at /home/carlos/LTER_CML_CAK

## -- Attaching packages ----- tidyverse 1.3.2 --
## v ggplot2 3.3.6     v purrr    0.3.4
## v tibble   3.1.7     v dplyr    1.0.9
## v tidyr    1.2.0     v stringr  1.4.0
## v readr    2.1.2     v forcats 0.5.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()   masks stats::lag()
##
## Attaching package: 'plotly'
##
##
## The following object is masked from 'package:ggplot2':
##
##       last_plot
##
##
## The following object is masked from 'package:stats':
##
##       filter
##
##
## The following object is masked from 'package:graphics':
##
##       layout
##
##
## Attaching package: 'lubridate'
##
##
## The following objects are masked from 'package:base':
```

```

## date, intersect, setdiff, union
##
## # Attaching packages: easystats 0.5.2 (red = needs update)
## x insight      0.18.3    x datawizard   0.6.0
## √ bayestestR   0.13.0    x performance  0.9.2
## x parameters   0.18.2    x effectsize    0.7.0.5
## √ modelbased   0.8.5    x correlation   0.8.2
## x see         0.7.2    √ report       0.5.5
##
## Restart the R-Session and update packages in red with `easystats::easystats_update()``.
##
## Loading required package: carData
##
## lattice theme set by effectsTheme()
## See ?effectsTheme for details.
##
## This is DHARMA 0.4.6. For overview type '?DHARMA'. For recent changes, type news(package = 'DHARMA')
##
## New names:
## Rows: 32209 Columns: 17
## -- Column specification -----
## Delimiter: ","
## chr (7): sitecode, section, reach, unittype, species, clip, notes
## dbl (9): ...1, year, pass, unitnum, vert_index, pitnumber, length_1_mm, len...
## date (1): sampledate
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.

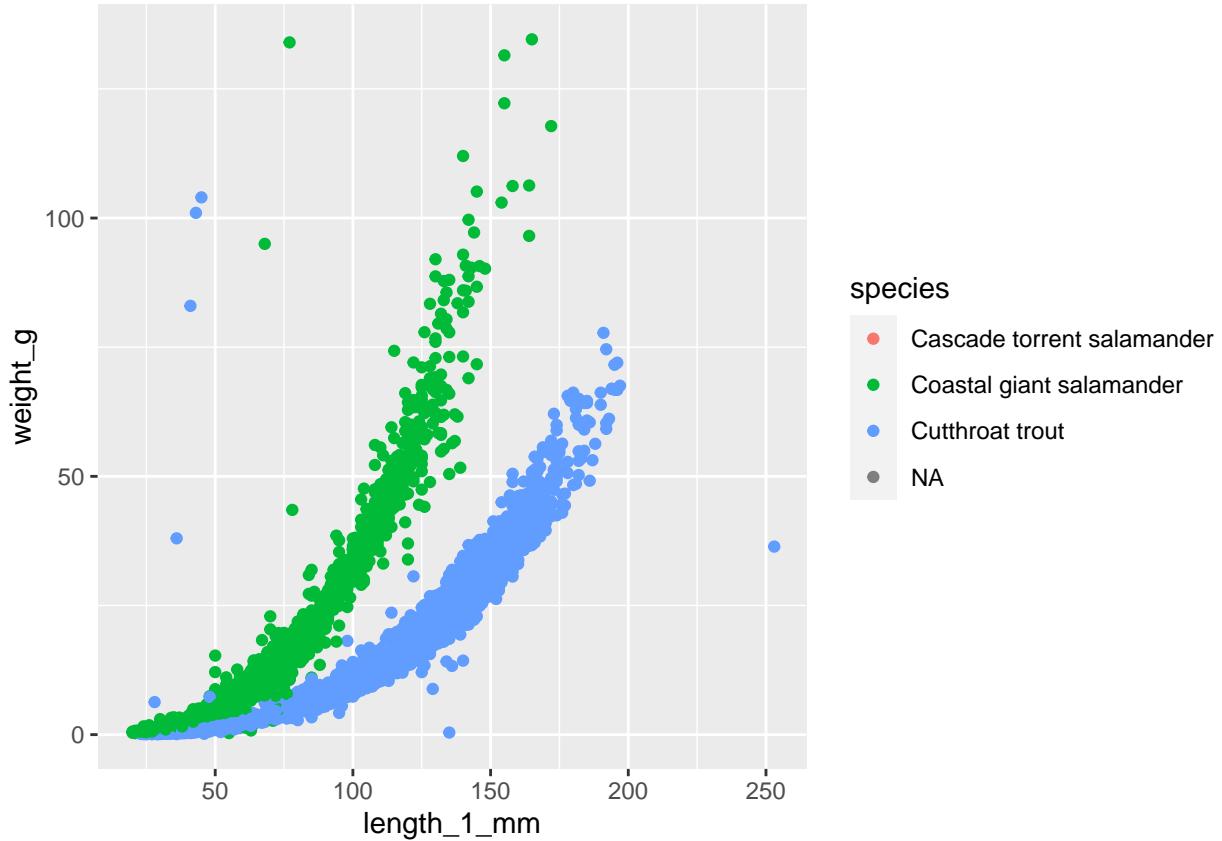
## # A tibble: 6 x 17
## ...1 year sitecode section reach pass unitnum unittype vert_index pitnumber
## <dbl> <dbl> <chr>   <chr>  <dbl> <dbl> <chr>   <dbl> <dbl>
## 1     1 1987 MACKCC-L CC     L     1     1 R     1     NA
## 2     2 1987 MACKCC-L CC     L     1     1 R     2     NA
## 3     3 1987 MACKCC-L CC     L     1     1 R     3     NA
## 4     4 1987 MACKCC-L CC     L     1     1 R     4     NA
## 5     5 1987 MACKCC-L CC     L     1     1 R     5     NA
## 6     6 1987 MACKCC-L CC     L     1     1 R     6     NA
## # ... with 7 more variables: species <chr>, length_1_mm <dbl>,
## #   length_2_mm <dbl>, weight_g <dbl>, clip <chr>, sampledate <date>,
## #   notes <chr>

```

Plot the raw data

In this case, we are interested in modeling the length-mass relationships for cutthroat trout and salamanders in Mack Creek:

```
## Warning: Removed 13279 rows containing missing values (geom_point).
```



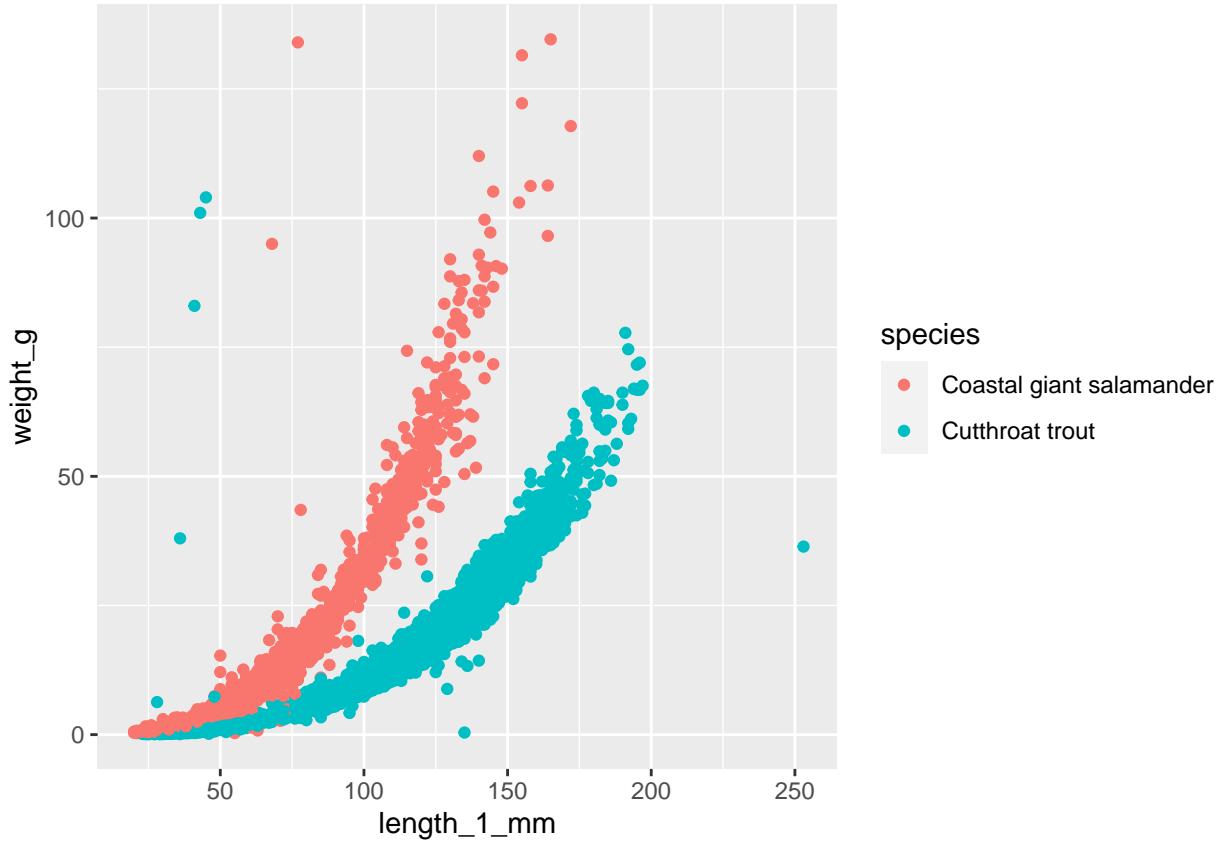
There are three species sampled but the Cascade torrent salamander is almost absent from the dataset, so we decided to ignore this species from our analysis.

First we subsampled the dataset and then plotted it:

```
data_species2 <- ourdata %>%
  subset(species != "Cascade torrent salamander")

data_species2$species <- as.factor(data_species2$species)
ggplot(data_species2, aes(x = length_1_mm,
                          y = weight_g,
                          color = species)) +
  geom_point()

## Warning: Removed 13270 rows containing missing values (geom_point).
```



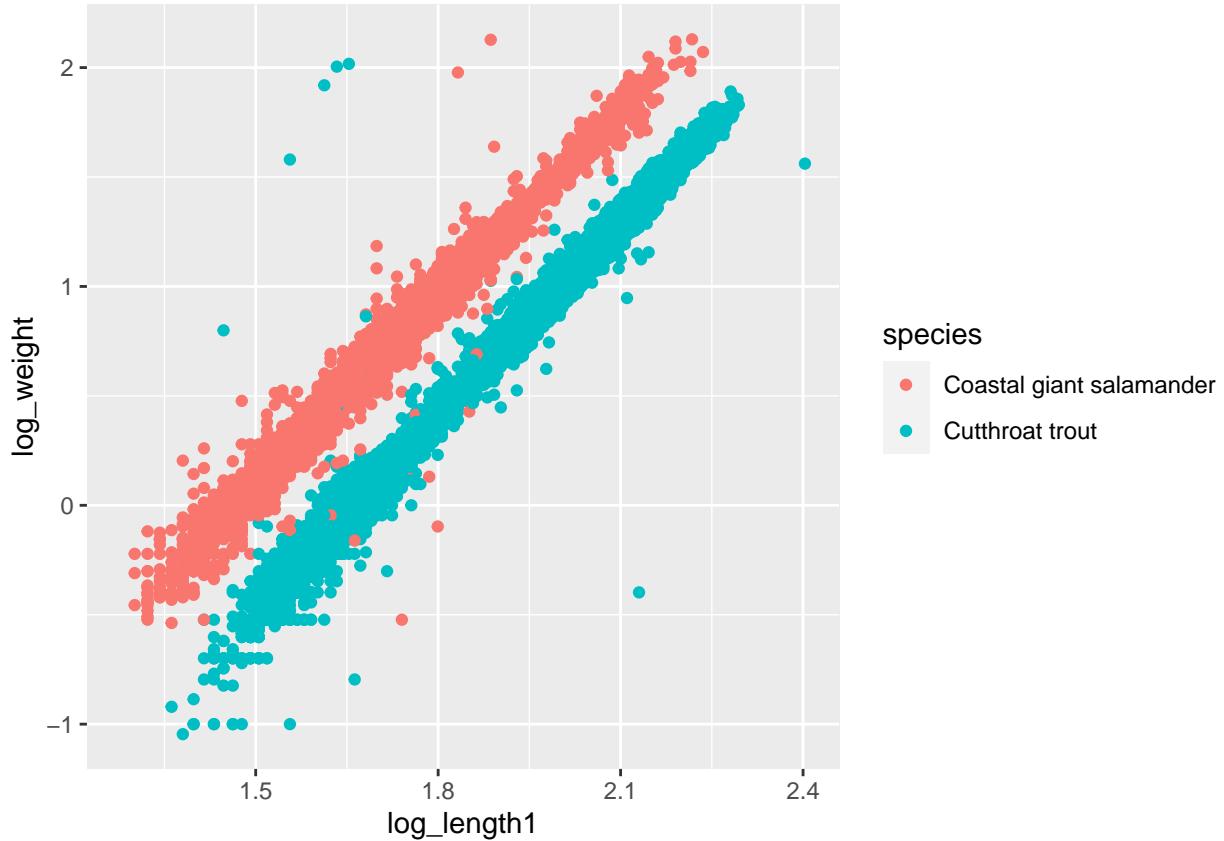
The relationship between length and weight of these species is exponential, thus we decided to transform our variables into logarithm (we checked log2 for length and log10 for biomass)

The reason for log10 transform the biomass is because we have many small values (near zero) and wanted to enlarge the distribution “scale”

```
data_species_log <- data_species2
data_species_log$log_length1 <- log10(data_species_log$length_1_mm)
data_species_log$log_weight <- log10(data_species_log$weight_g)

ggplot(data_species_log, aes(x = log_length1,
                             y = log_weight,
                             color = species)) +
  geom_point()

## Warning: Removed 13270 rows containing missing values (geom_point).
```



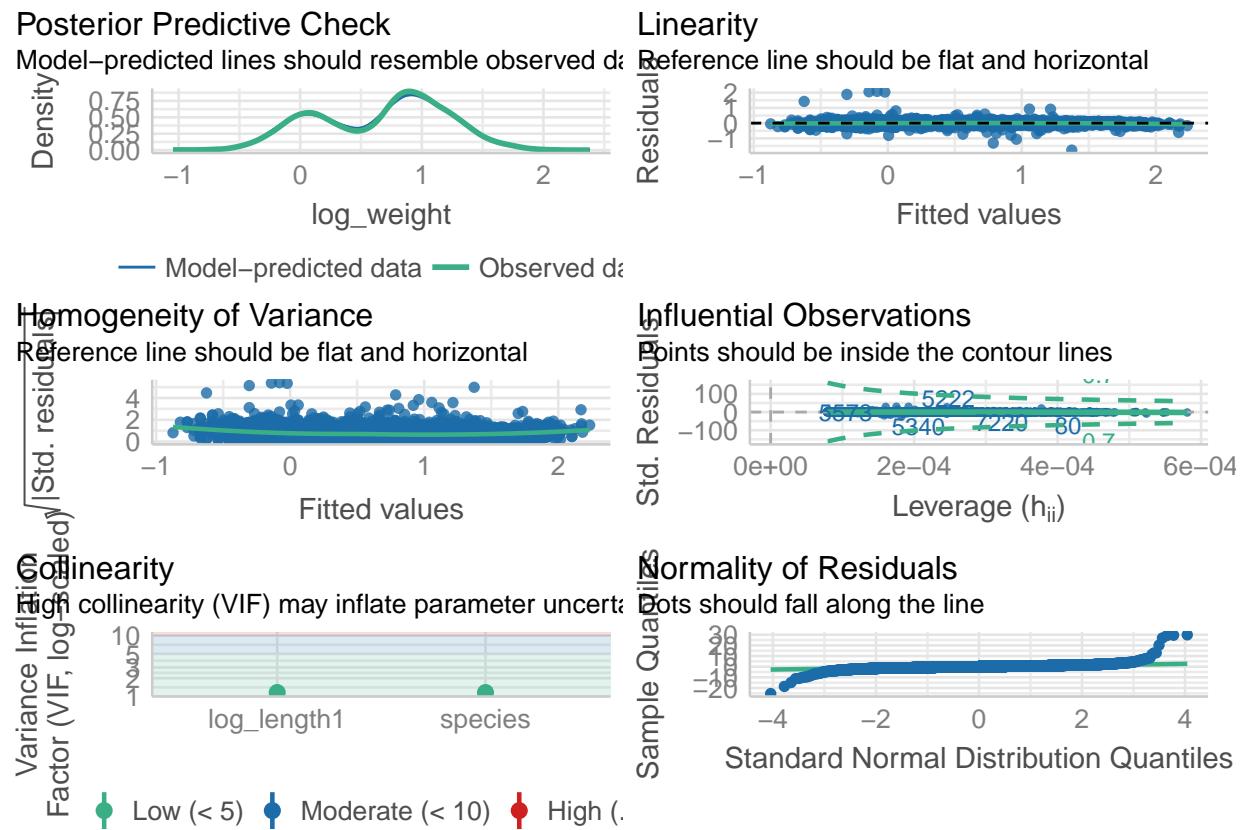
Result: It worked. The relationship is almost linear. We now model the allometric relationship between the length and biomass of these species and check whether the curves per species are different. We will check if the models “follow” the assumptions :

```
lm_log_species <- lm(log_weight ~ log_length1 + species, data=data_species_log)
summary(lm_log_species)
```

```
##
## Call:
## lm(formula = log_weight ~ log_length1 + species, data = data_species_log)
##
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -1.77068 -0.03244  0.00066  0.03173  2.08377 
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) -4.299328  0.005009 -858.3 <2e-16 ***
## log_length1  2.922709  0.002860 1021.8 <2e-16 ***
## speciesCutthroat trout -0.554276  0.001184 -468.0 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.07131 on 18918 degrees of freedom
##   (13270 observations deleted due to missingness)
## Multiple R-squared:  0.9823, Adjusted R-squared:  0.9823
```

```
## F-statistic: 5.263e+05 on 2 and 18918 DF, p-value: < 2.2e-16
```

```
check_model(lm_log_species)
```



[COMENTARLO CARLOS] Hola Cristina!!! Ahí va mi comentario: A la vista de los resultados del summary y de la comprobación de las asunciones del modelo con ‘checkmodel’, podemos decir que el modelo lineal propuesto es válido. Encontramos una clara relación positiva entre la longitud y el peso en ambas especies.

[NOTA para Cristina: he hecho para probar el modelo lineal sin incluir el factor especie, así: lm_log_species_only <- lm(log_weight ~ log_length1, data=data_species_log) Efectivamente sale peor, el R cuadrado sale 0.778, por lo que explica peor los datos.

```
lm_log_species_i <- lm(log_weight ~ log_length1 * species, data=data_species_log)
summary(lm_log_species_i)
```

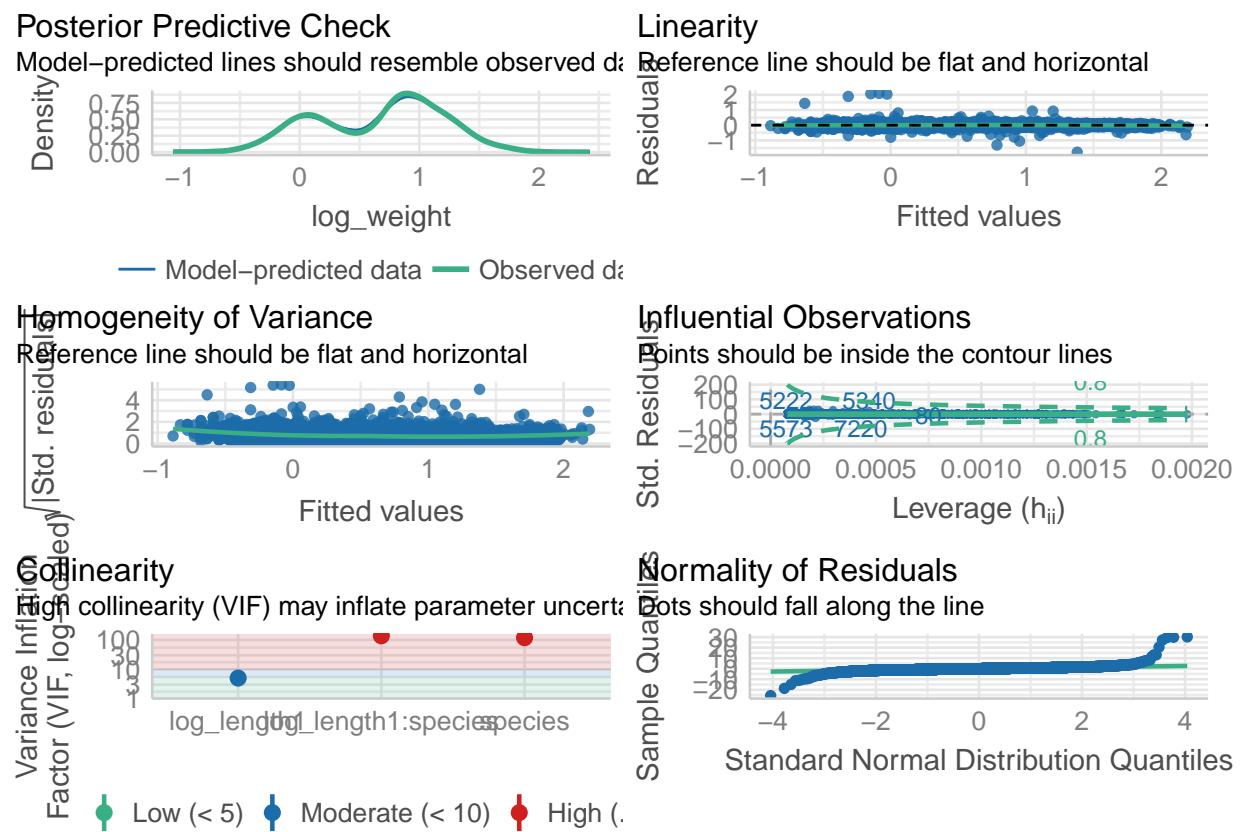
```
##
## Call:
## lm(formula = log_weight ~ log_length1 * species, data = data_species_log)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.77696 -0.03225  0.00075  0.03096  2.08983
##
## Coefficients:
## (Intercept)          Estimate Std. Error t value Pr(>|t|)
## -4.157994    0.010196 -407.81    <2e-16 ***
## log_length1          Estimate Std. Error t value Pr(>|t|)
##  0.03096    0.00075   40.98    <2e-16 ***
## species              Estimate Std. Error t value Pr(>|t|)
##  0.03096    0.00075   40.98    <2e-16 ***
```

```

## log_length1              2.840679  0.005895  481.87 <2e-16 ***
## speciesCutthroat trout -0.742230  0.011893 -62.41 <2e-16 ***
## log_length1:speciesCutthroat trout 0.106861  0.006728  15.88 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.07084 on 18917 degrees of freedom
##   (13270 observations deleted due to missingness)
## Multiple R-squared:  0.9826, Adjusted R-squared:  0.9826
## F-statistic: 3.556e+05 on 3 and 18917 DF,  p-value: < 2.2e-16

```

```
check_model(lm_log_species_i)
```



[COMENTARLO CARLOS] A la vista del resultado del modelo con iteración, podemos decir que efectivamente existe cierta interacción con la especie, lo que indica que la pendiente de la recta del modelo varía según la especie, aunque esta variación debe ser pequeña porque el R cuadrado tiene sólo una pequeña mejoría respecto al modelo anterior.

[¡¡Paper listo para envío a revista!!]