Sensitivity analysis of model parameters

2024-01-17

Model

The allometric model described by is the following:

$$B = d\left[\frac{H}{2}\left(A_{basal} + A_{half}\right)\right]^z,$$

where B is the biomass calculated by the allometric equation, d is the bulk density of the plant species, H is the measured height of the plant, A_{basal} and A_{half} are the basal area and the area at half the height of the plant, respectively; and z is the power coefficient of a power law.

Problem

In Perronne et al. (2020), the authors assign a value of 1.96 to d and a value of 2/3 to z; however, we do not know how those parameters interact with each other.

Approach

To solve this, we will perform a sensitivity analysis in which we modify the values of d and z within a certain range. Importantly, to simplify the allometric equation and do a clear sensitivity analyses that works independently of the field data, we opt to include the parameter x as:

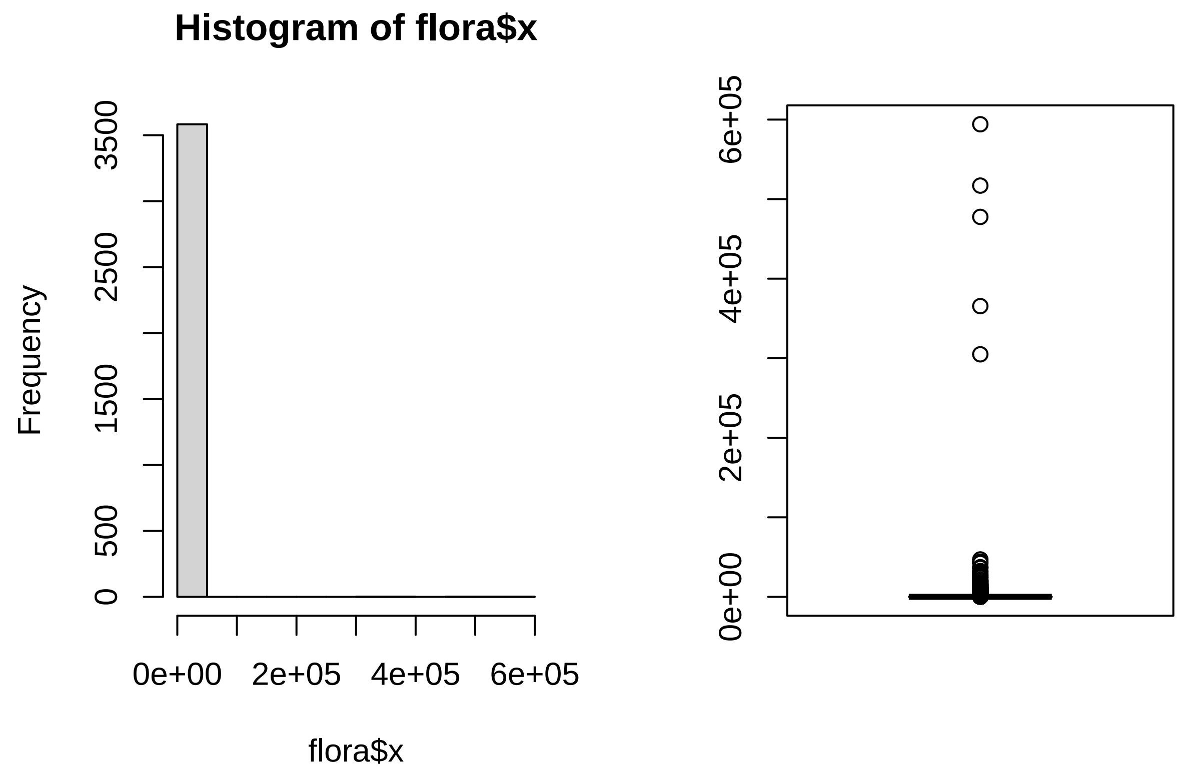
$$x = \left[\frac{H}{2}\left(A_{basal} + A_{half}\right)\right],$$

meaning that all allometric measurements are summarized now by x. Thus, the equation used for our sensitivity analysis is the following:

$$B = dx^z$$

Distribution of X and outliers management

First, we have to take a look to an histogram and a boxplot of X distribution:



As we can see, there is a problem with X outliers. If we take a look to the quantiles:

- 0%: 7.0685835×10^{-4}
- 25%: 2.4774796
- 50%: 26.3029845
- 75%: 91.8915851
- 100%: 5.9411444×10^{5}

Up to 75% there is no a significant jump in the magnitude, but, in the last quantile (75%-100%) there is an increase of 10^5. We can increase resolution between the last quantiles:

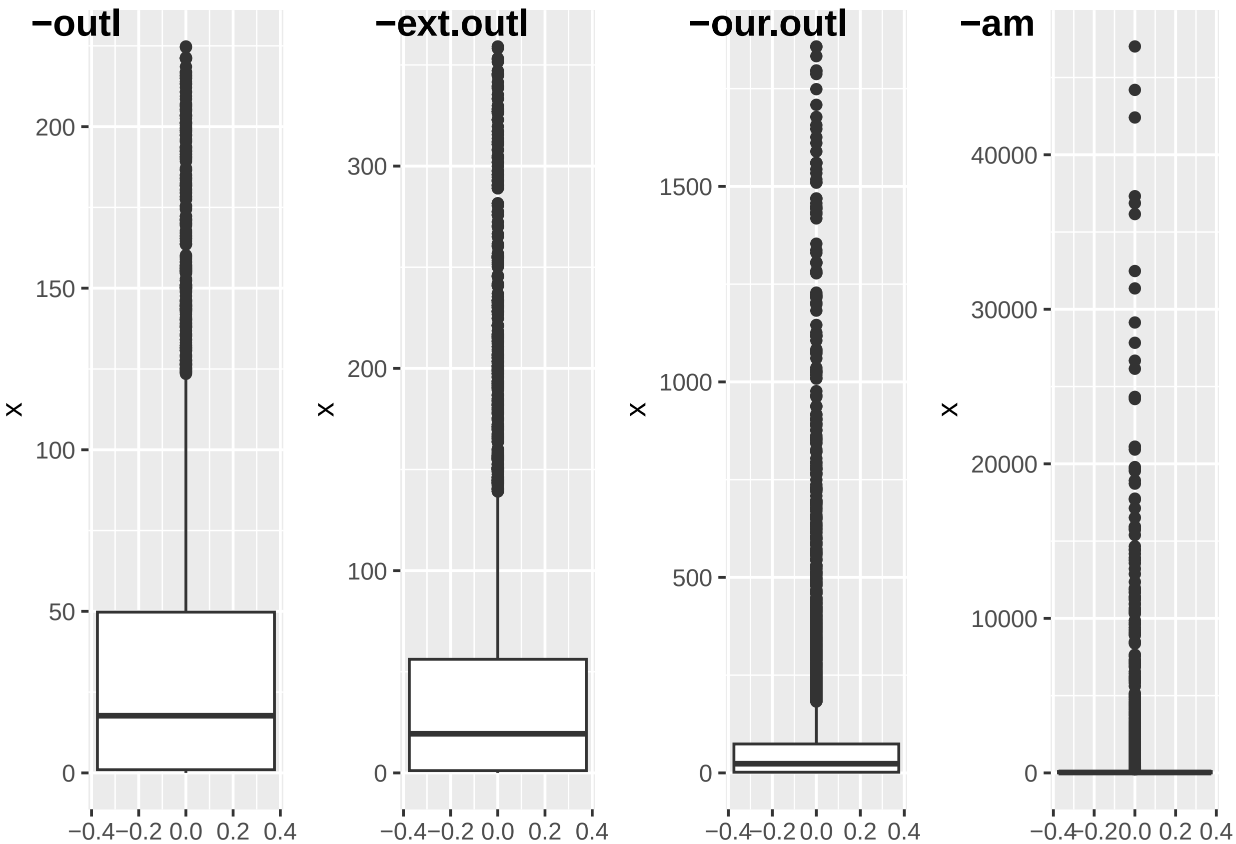
- 80%: 136.4629309
- 85%: 221.249117
- 90%: 461.6460449
- 95%: 1820.313359
- 96%: 2914.2591552
- 97%: 4576.3963066
- 98%: 8927.6887817
- 99%: 1.5436943×10^{4}

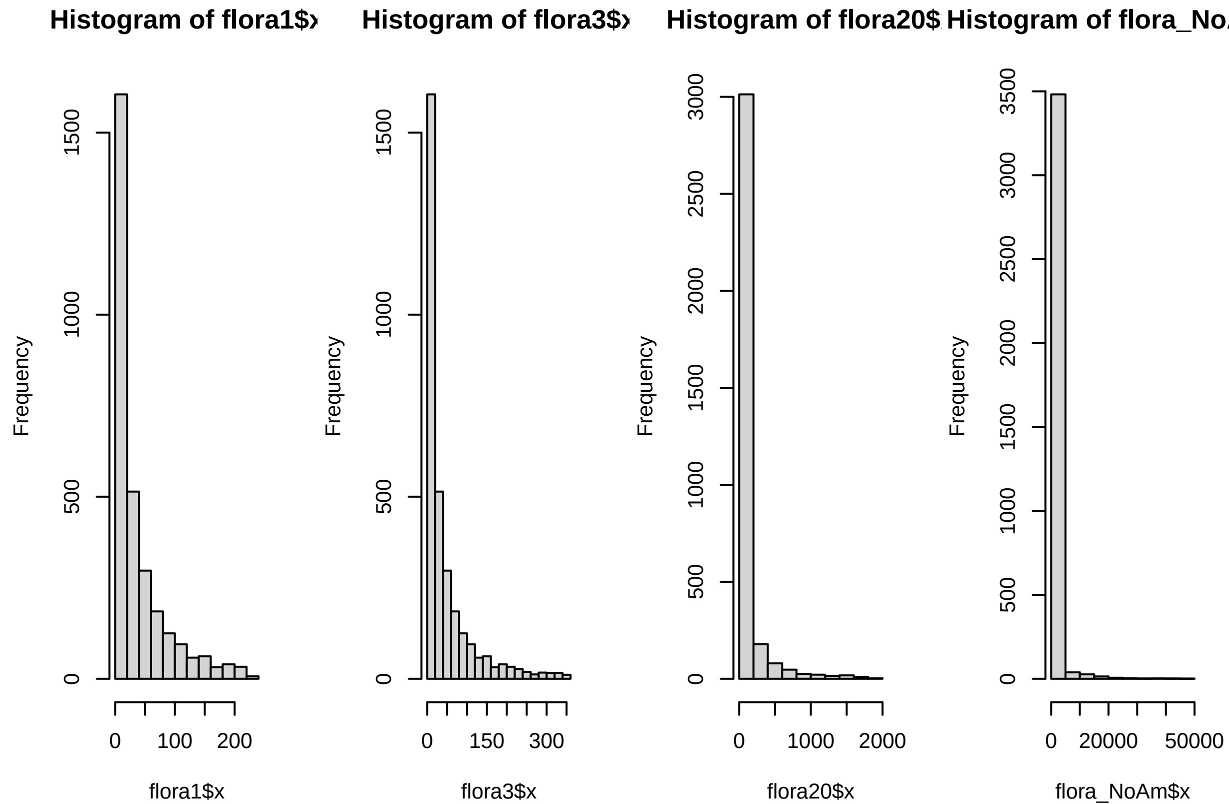
Hence, we have a series of outliers that are preventing us of observing the distribution of x. To manage outliers, we have several options:

- Removing the *outliers* that fall above $Q3 + 1.5$ůIQR. We would remove everything above $x = 226.01$, which are 535 data points out of 3588 (14.91 %).

- Removing the *extreme outliers* that fall above $Q3 + 3ůIQR$.We would remove everything above $x =$ 360.13, which are 424 data points out of 3588 (11.82 %).
- Removing outliers above a specific value that we consider more appropiate. For example, removing every data that falls above $Q3 + 20ůIQR$ is removing everything above 1880.17. Which are 177 data points out of 3588 (4.93 %).
- Removing specific species. e.g. removing the species "Amaranthus sp." ("am"). We would remove 5 datapoints, but the maximum value would be 4.7009909×10^4.

The distribution of data once we reject different levels of outliers would be as follow:





Boxplots and histograms of X

Questions about outliers

- What is the ecological meaning of the outliers? And what would be the meaning of their removal?
- If we do remove outliers, which group of outliers should we take out?

d and z influence on x

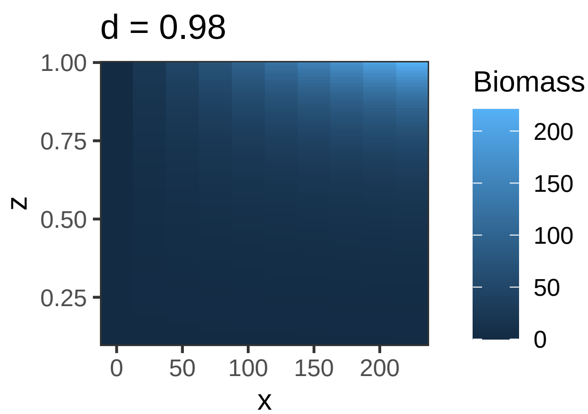
At this point, we assign the following values to the different parameters in the model:

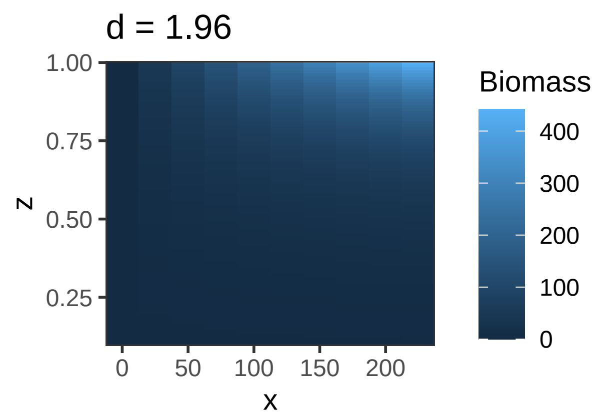
- d. 1.96 is the value proposed by Perronne et al. (2020). By the structure of the model, we know that d will have an influence on the scale of biomass. Hence, we will only apply half of that value and its double. Therefore: for d, values of 0.98, 1.96, 3.92;
- x. We have seen how distribution of x is influenced by the pressence of outliers. We will have 3 groups of x. 1) for $x1$ (without outliers), 0, 25, 50, 75, 100, 125, 150, 175, 200, 225; 2) for $xsmall$ (for the smallest values of x, going from the minimum value until the 25% quantile), 0, 0.15, 0.3, 0.45, 0.6, 0.75, 0.9, 1.05, 1.2, 1.35, 1.5, 1.65, 1.8, 1.95; for $xout$, outliers of x (outliers).
- z. The value proposed for z by the authors is 2/3. Since it is a power coefficient, we need to apply more values of z in order to understand its influence on the biomass. Hence, we will have a list of values: 0.1, 0.11, 0.12, 0.13, 0.14, 0.15, 0.16, 0.17, 0.18, 0.19, 0.2, 0.21, 0.22, 0.23, 0.24, 0.25, 0.26, 0.27, 0.28, 0.29, 0.3, 0.31, 0.32, 0.33, 0.34, 0.35, 0.36, 0.37, 0.38, 0.39, 0.4, 0.41, 0.42, 0.43, 0.44, 0.45, 0.46, 0.47, 0.48, 0.49, 0.5, 0.51, 0.52, 0.53, 0.54, 0.55, 0.56, 0.57, 0.58, 0.59, 0.6, 0.61, 0.62, 0.63, 0.64, 0.65, 0.66, 0.67, 0.68, 0.69, 0.7, 0.71, 0.72, 0.73, 0.74, 0.75, 0.76, 0.77, 0.78, 0.79, 0.8, 0.81, 0.82, 0.83, 0.84, 0.85, 0.86, 0.87, 0.88, 0.89, 0.9, 0.91, 0.92, 0.93, 0.94, 0.95, 0.96, 0.97, 0.98, 0.99, 1.

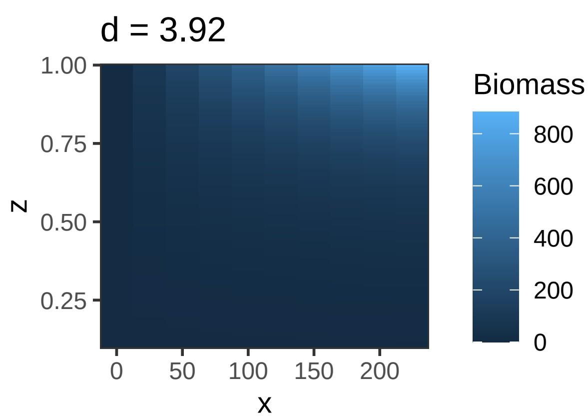
Dataviz

Now, some visualizations to understand the relationship between the different parameters.

x and z The relationship between x and z for values of d looks like this:





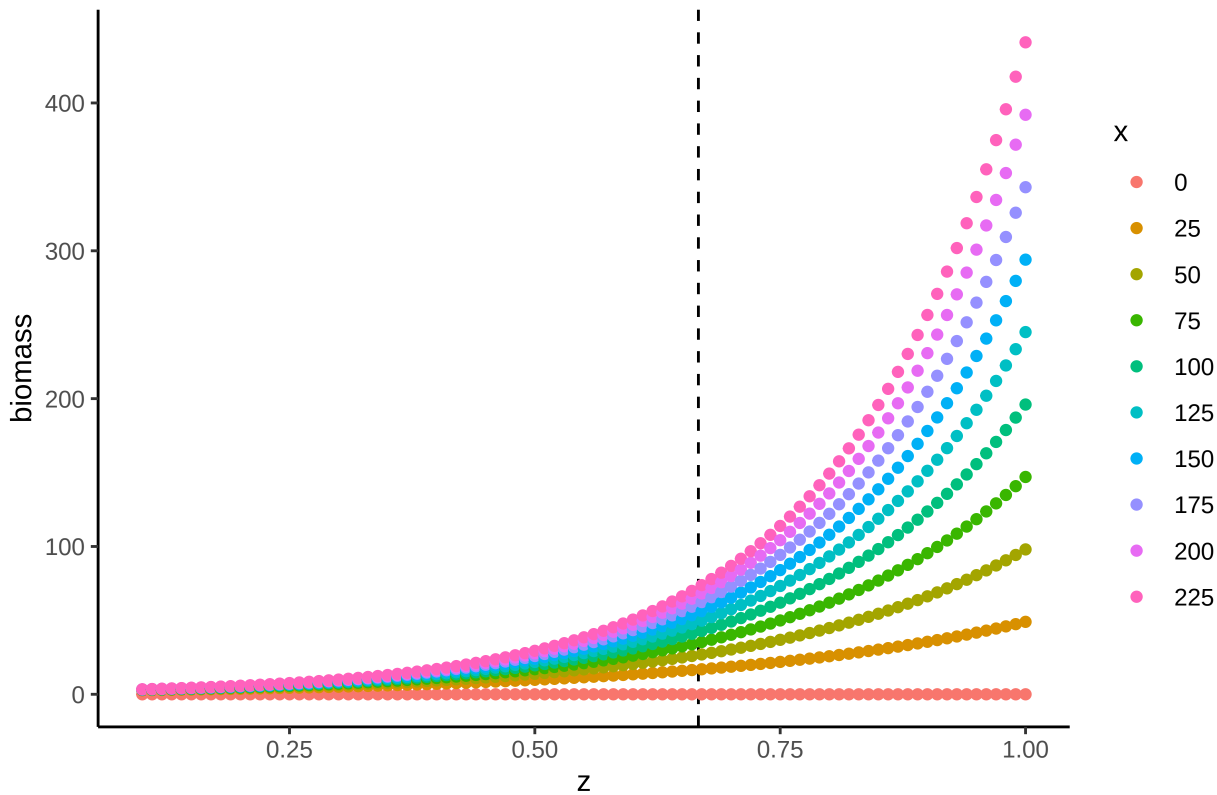


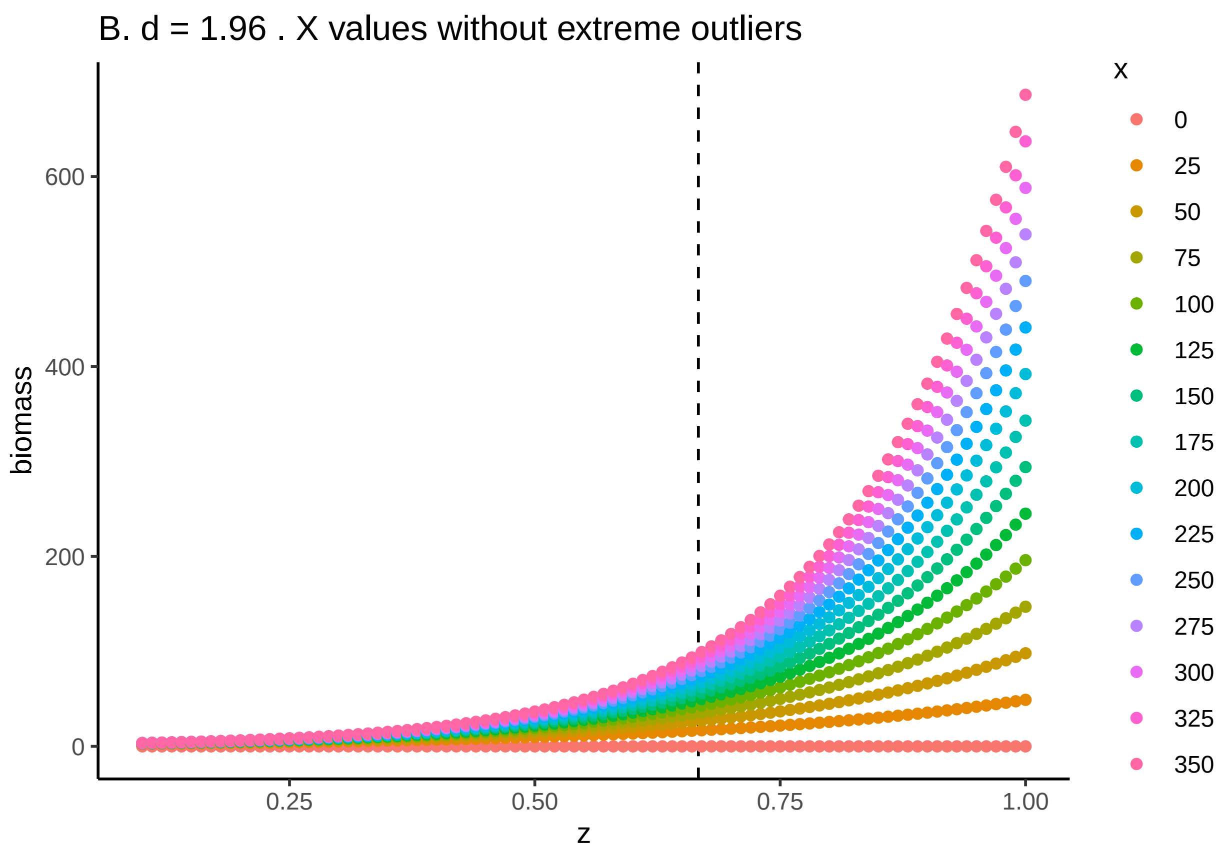
From these graphs we can conclude that: d have an effect on magnitude of biomass and z increases biomass proportionally to X. Since we do not want to exactly know the absolute values of biomass, but rather the relative contribution of each species, d value is not as relevant. However, the relationship between z and x deserves further consideration. In this regard, we can first take a look to x distribution.

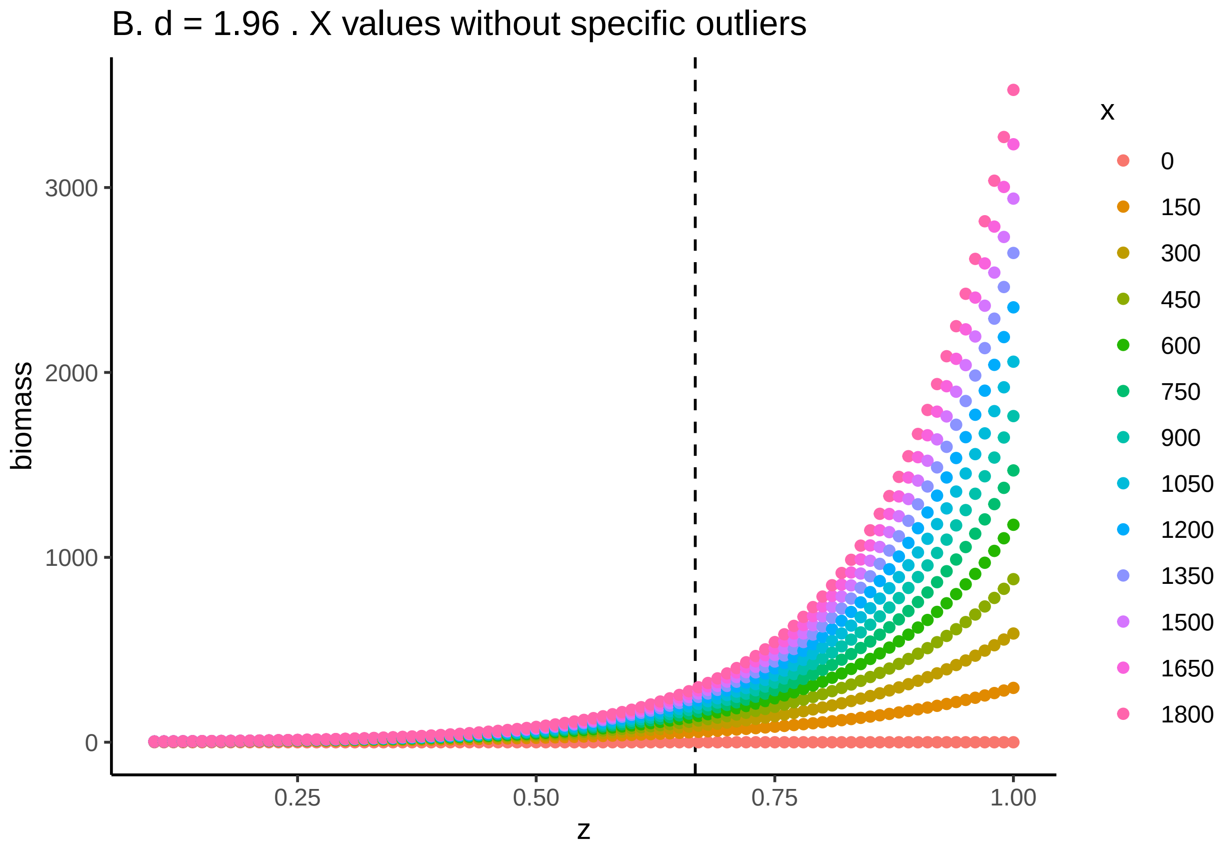
Expression "X" and its interaction with Z

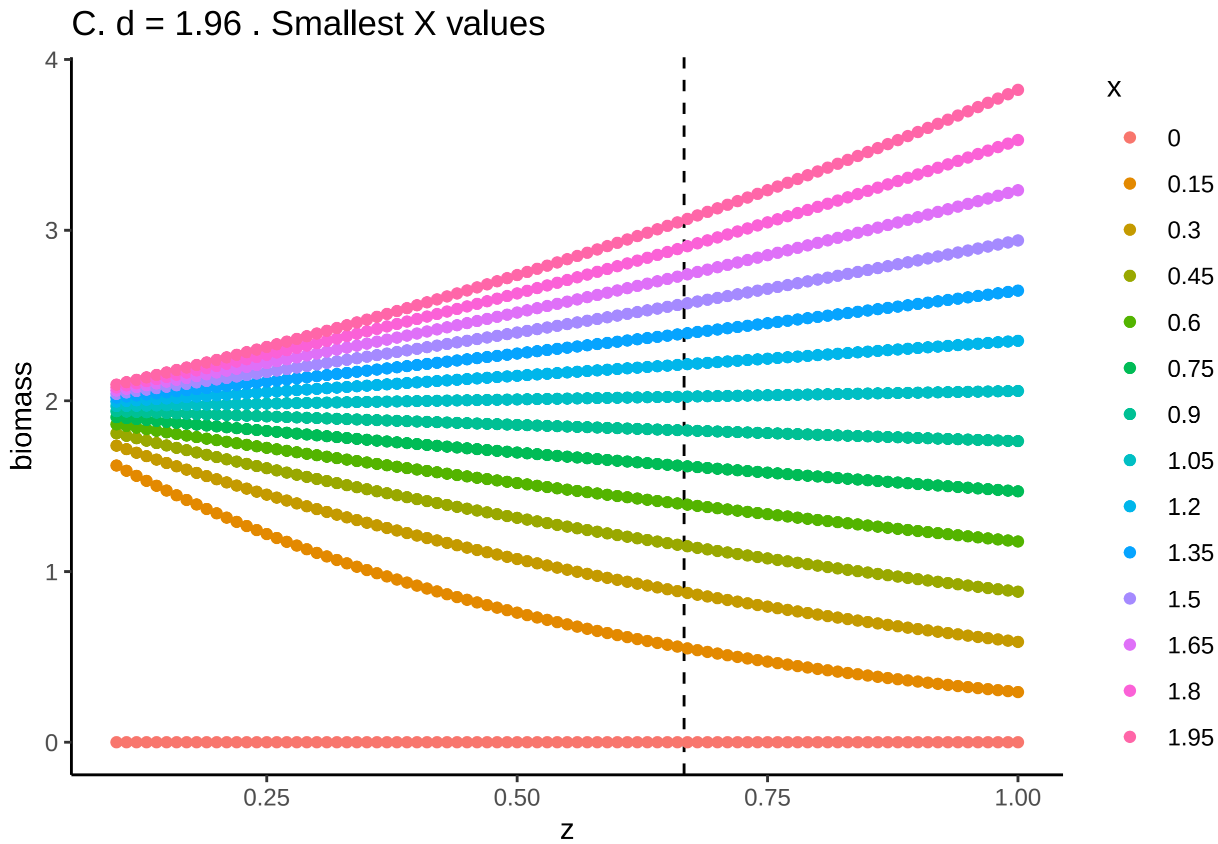
To check z influence on x we can display a graph where we can see how biomass change as a function of changes in z, for different values of x. In the following graphs we can see:

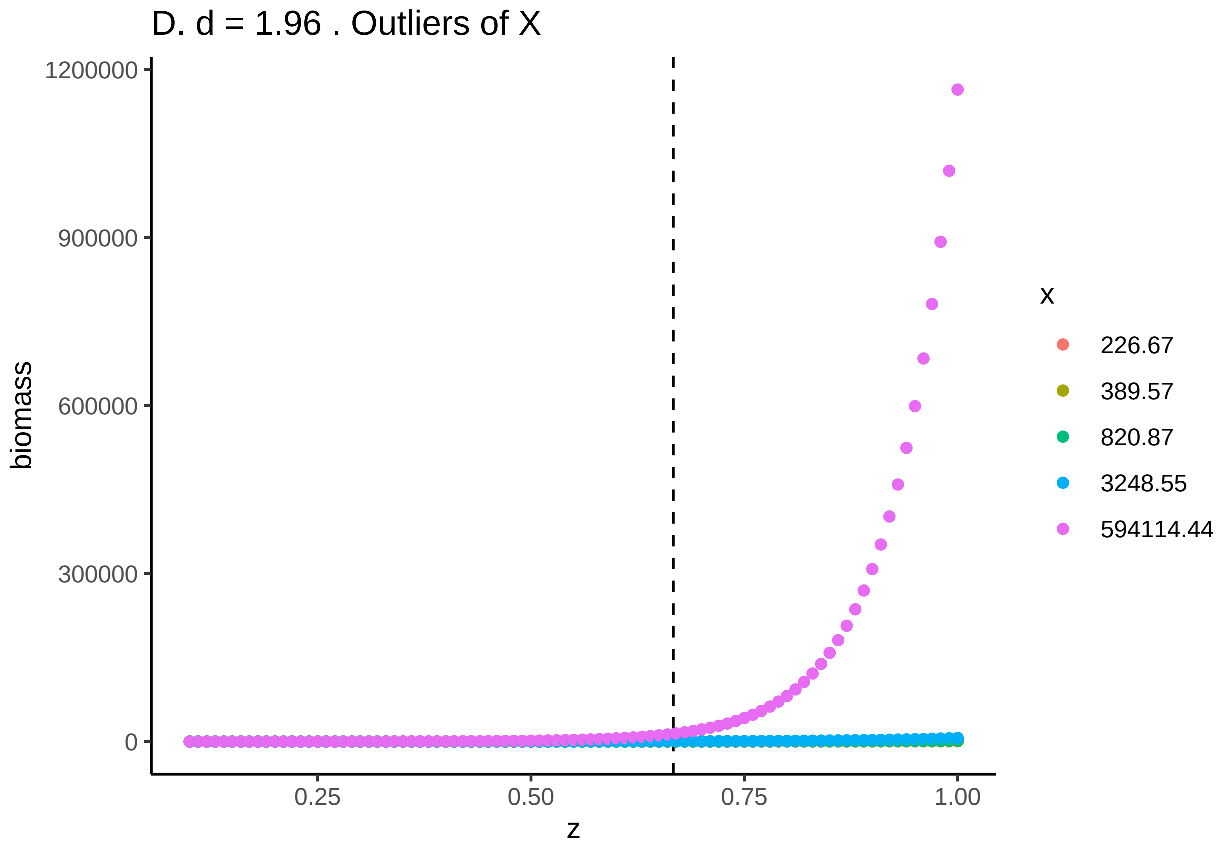
A. d = 1.96 . X values without outliers




B. d = 1.96 . X values without extreme outliers
biomass
600
400
200
0
0.25
0.50
0.75
1.00
z
x
0
25
50
75
100
125
150
175
200
225
250
275
300
325
350


B. d = 1.96 . X values without specific outliers
biomass
3000
2000
1000
0
0.25
0.50
0.75
1.00
z
X
0
150
300
450
600
750
900
1050
1200
1350
1500
1650
1800


C. d = 1.96 . Smallest X values
4
3
2
1
0
biomass
0.25
0.50
0.75
1.00
z
x
0
0.15
0.3
0.45
0.6
0.75
0.9
1.05
1.2
1.35
1.5
1.65
1.8
1.95



The dotted line represents the value of $z = 2/3$. For the highest values of x, a bigger z means a positive influence on biomass. Meanwhile, for those x values below 1, z has a negative influence on biomass. A 25 % of x datapoints in the database are below 0.

Conclusions

- $d = 1.96$. d affects the magnitude (scale) of biomass, no matter the size of the species. It would influence the fact that we got mg, cg, g or kg of biomass. Since we are itnerested in the relative values, it is no so relevant.
- $z = 2/3$. If it were higher, the differences estimated for biomass between small species and big species would be "too much" (when is too much?). If it were lower, some differences wouldn't be appreciated. hence, either we do not use a z value or we use the proposed value.

Further questions

- Based on z influence on X and on data distribution, which outliers should we take out? Extreme outliers or just outliers?
- If we do remove outliers, which ecological/conceptual implications would have?

References

Perronne, R., F. Jabot, and J. Pottier. 2020. Inter- and intraspecific variability of plant individual growth and its role on species ranking in grasslands. Journal of Plant Ecology 13:378–386.