Correlations

Javier Porras

2023-12-19

Context

The non-destructive methodology we are currently following (based on (**Jabbot?**)) had not been yet decided in the first 3 samplings (sampling 0, 1 and 2). While now we are taking data of height (H), circumference at base level (Cb) and circumference at medium height level (Ch), in the 3 first samplings we only gathered information about H. To apply the models proposed by (**Jabbot?**) and Perronne et al. (2020) we need the 3 morphological measurements. Hence, we have to find a way to fill these gaps of data.

Approach

So far, excluding these 3 samplings, we have around 3 thousand data points, which include a measurement of each of the 3 morphological data for a certain individual from a species. Hence, we can use these data to check if there is any significant correlation between H and Cb and Cm. To do this, we can create linear models between these parameters per species and use $R - squared$ and $p - value$ as statistical indicators to check the significance of correlations. Therefore, we would have: - lm($Cb \sim H$, flora). Where we can see the correlation between H and Cb. - lm($Cm \sim H$, flora). Where we can see the correlation between H and Cm.

After filtering data, we have 53 species to be subjected to the linear regression. We will choose only the data points stored in the database for these species.

However, there are species that were only present during the first three samplings. The final number of species of which we have available data are **34**. The lacking 19 species are the following ones: apar, asteraceacardosa, avfr, cardo, dile, g13, hola, homu, loco, mepo, ni cardo, ni_orquidea, opap, pobu, tr, trdu, trnu, valo, vear

Correlations

As a result of applying the linear model regressions between H and Cb and Cm, we have found that barely 3 species have a significant linear correlation. Besides, only 1 out of these 3 species have a substantial amount of measurements included (Table 1.)

species	num_measurements	cb_R_squared	cb_P_value	cm_R_squared	cm_P_value
brst	3	0.9901884	0.0631631	0.9888797	0.0672583
libi	50	0.7614052	0.0000000	0.8705340	0.0000000
gafr	6	0.7609468	0.0234116	0.8660776	0.0070508
avst	18	0.2454422	0.0365615	0.4282153	0.0032140
tono	11	0.2339902	0.1316797	0.7537008	0.0005291
bepe	63	0.2253497	0.0000846	0.1949912	0.0002914
pore	186	0.1924614	0.0000000	0.1859120	0.0000000

Table 1. Top 7 Correlations based on R-squared values for Cb

Since we have found no proper correlations for Cb or Cm, we decided to try with the expression X. X stands for the part of the equation provided by Perronne et al. (2020) that aggregates the three morphological values: H, Cb and Cm. The circumferences are transformed in *area* values.

So we have the general equation:

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

Where X should be:

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

Therefore:

$$B = dx^{z}$$

d and z are respectively 1.96 and 2/3 (see Perronne et al. (2020))

As a result, we have found an improvement in the number of species where there is an acceptable linear correlation (R-squared > 0.5). There is a total of 7 of these species, as we can see in Table 2.

species	x_R_squared	x_P_value	num_measurements
brst	0.9878137	0.0704211	3
visa	0.9527183	0.0000000	34
cadi	0.6483155	0.0000000	31
gafr	0.6305867	0.0592305	6
melu	0.6096626	0.0000000	40
libi	0.6048553	0.0000000	50
tono	0.5618299	0.0079102	11
rabu	0.4107980	0.0000012	47
pore	0.3307080	0.0000000	186
poan	0.3043711	0.0012942	31

Table 2. Top 10 Correlations based on R-squared values for X

Further questions and tasks

- Can we do any "gap-filling" with these data?
- If we cannot, how should we proceed with the first 3 samplings?

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.