Final report Francesca Grisafi

[Duration of project: 1](#_Toc101448002)

[Project members: 1](#_Toc101448003)

[Aims 1](#_Toc101448004)

[Hazelnut description 2](#_Toc101448005)

[Material and methods 2](#_Toc101448006)

[Procedures 2](#_Toc101448007)

[Results 3](#_Toc101448008)

[Exploratory analysis 3](#_Toc101448009)

[GLMs 8](#_Toc101448010)

[To Do: 15](#_Toc101448011)

# Duration of project:

October 2021 – March 2022

# Project members:

* Francesca Grisafi\_ PhD student
* Evelyne Costes
* Frederic Boudon
* Jean Baptiste Durant

# Aims

The goal of my period in Montpellier was to **analyze** and **code** a coarse architectural model of hazelnut. The model should address several questions:

1. How 1-year-old shoot is composed?

* There are some zones of the same type of bud?
* There are differences in the composition according to the length of the shoot?

1. What is the behavior of lateral shoots?

* How many of them developed?
* Where?
* From which bud? (vegetative or mixed?)
* How can we deal with multiple buds and multiple lateral shoots per node?

The way those questions are solved is through data **analysis,** in **Rstudio** from a different point of view (i.e. shoot scale, metamer scale, and bud scale) and with different statistical methods and models (i.e. exploratory analysis, Glms, Markovian models). Then, the resulting models were used in **Lpy** to **code** the growth of the plant.

# Hazelnut description

Hazelnut (*Corylus avellana)* is a plant with a bushy shape. In a 1-year old proleptic shoot, with an **alternate phyllotaxy**, are present different types of entities:

* **Mixed bud**: it has both vegetative and reproductive parts. It will burst into a vegetative shoot with the inflorescence (and then, nuts) in the apical position.
* **Vegetative bud**: it has just the vegetative part. It will burst into a vegetative shoot.
* **Blind node**: node with no buds presence
* **Sylleptic shoot**: this is a short shoot that develops in the same year of the parent. They also have, with an alternate phyllotaxy, the presence of Mixed and Vegetative buds. At the apical position, they have the **male inflorescence: the catkin.**

# Material and methods

A sampling of **120** 1year old shoots was collected **in 2020 in Perugia** (Italy). At the same moment of the collection, some **biometrical measures (i.e. diameter, length of the shoots, number of nodes)** were made as well as the qualitative and quantitative measures, at the node level, of the type of entities we could observe. Shoots were then aggregated into four **length classes (short, medium, long, very long)** according to their length in cm.

In 2021 the same analysis was repeated on the children’s shoots that were developed on the 1-year-old shoots targeted the year before.

In Montpellier, those data were analyzed and the results were used to code the first draft of hazelnut FSPM.

## Procedures

The first problem that was faced was the **definition of hazelnut elements** (e.g. sylleptic/proleptic shoots, mixed/catkins, lateral/apical, single/multiple buds)**.** Thus**,** common names were defined to describe the tree. Thus, the original dataset was re-designed into three ones according to the scale they were describing: shoot scale dataset; metamer scale dataset; bud scale dataset. Thus in the last two, the apical metamer/bud was distinguished from the lateral ones. The first lead to the **succession of the** shoot while the seconds lead to **lateral branches.**

First, **exploratory analyses** were performed, in **RStudio,** to understand the relationships between different variables inside the dataset (e.g.. length of the children ~ length of the parent, number of nodes ~ length of the parent, type of buds ~ rank node). Linear regression models, t-test, and proportion test were used to evaluate correlations or differences in mean/proportions. Exploratory analyses gave a first idea of the architectural features of hazelnut. Thus un diagram of the architectural dynamics was drawn. Following it, **generic linear models** and **multinomial regressions** were performed in **RStudio.** These models lead to equations that then were used in **Lpy** to link the different objects of the coarse architectural model.

## Results

### Exploratory analysis

#### Annual shoot (year n)

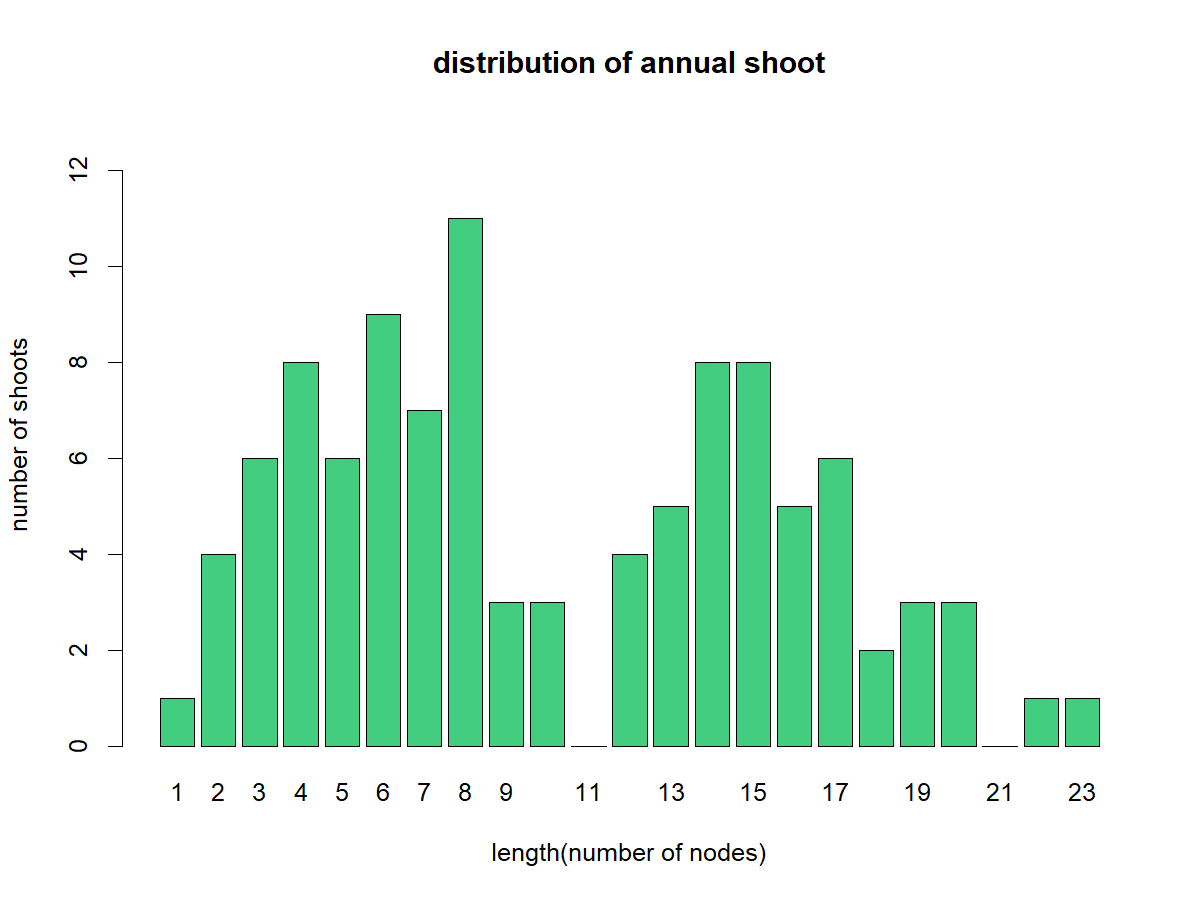
The distribution of length (number of nodes) was **bimodal**. This could initially suggest merging some class lengths (Figure 1: e.g. Sh+Me, Lo+VLo).

Figure 1: distribution of shoots per length (number of nodes). Performed in the annual shoot of year n.

Chart, treemap chart

Description automatically generatedThe proportion of different buds/sylleptic shoots in proleptic shoots showed no difference in the frequency of vegetative and mixed buds. However, they were both statistically different from Sylleptic and blind node frequency. The proportion of vegetative and mixed buds inside sylleptic shoots was not statistically different and was ~ 40-50%. (Figure 2).

Figure 2: percentage of mixed buds and vegetative buds in sylleptic shoots.

Chart, bar chart

Description automatically generatedThe proportion of different buds/sylleptic shoots according to different parent class lengths, of proleptic shoots, showed a higher presence of blind nodes in short and medium shoots while a higher presence of mixed buds in long and very long shoots (Figure 3).

Figure 3: percentage of lateral buds/sylleptic per each parental class length. The difference in proportion per each bud/sylleptic, between classes, was tested using prop.test in RStudio.

Chart, line chart

Description automatically generatedThe proportion of different buds/sylleptic shoots according to parent rank node showed a higher presence of blind nodes in the basal part of the shoots while a higher presence of mixed buds in nodes near the tip of the shoot (Figure 4).

Figure 4: percentage of lateral buds in proleptic shoots according to rank node (1 is at the base of the shoot). Just nodes from 1 to 16 were analyzed because more frequent.

Chart, bar chart, histogram

Description automatically generatedHazelnut could have more than one bud per node. The analysis of multiple buds per rank node showed that the majority of ranks have just one bud while sometimes there is a rank with two buds and rarely with three, four, or nine (Figure 5).

Figure 5: percentage of multiple buds (1,2,3,4 or 9 buds in the same node) in proleptic shoots according to rank node (1 is at the base of the shoot). Just nodes from 1 to 16 were analyzed because more frequent.

Chart, bar chart

Description automatically generatedThe fate of lateral buds was investigated. In fact, mixed and vegetative buds could burst and become lateral shoots in year n+1. At the beginning of the shoot (nodes 1-4), there are more blind nodes and vegetative buds. This zone is followed by another (node 5-8) with the prevalence of sylleptic shoots. Nodes 9-16 are characterized by the prevalence of mixed and vegetative (Figure 6). This is in line with Figure 4. The combination of more than one bud, if present is in the middle part of the shoot and at the end.

Figure 6: percentage of different lateral buds combination in proleptic shoots according to rank node (1 is at the base of the shoot). Just nodes from 1 to 16 were analyzed because more frequent.

The same analysis is meaningless in sylleptic shoots because some data are missing: sylleptic length, and number of nodes in sylleptic. Due to that, the number of combinations inside sylleptic seems bigger than it is in reality. This is further work that could be interesting to invest in.

#### Lateral shoot (year n+1)

Table

Description automatically generatedIn year *n+1*, at each metamer of the **parent shoot** (annual shoot), at least one **lateral shoot** is produced from **vegetive buds and/or mixed lateral buds**. (Table 1)

Table 1: description of lateral shoots born from buds in proleptic or sylleptic parents.

Chart, bar chart

Description automatically generatedThe relationship between the length of **proleptic parent** and the length of lateral children showed that the majority of new shoots are short shoots. Some medium shoots are observable from buds in very long (>40cm) parent shoots (Table 1,Figure 7, Figure 8).

Figure 7: percentage of lateral child (from buds in PROLEPTIC) class length per each parental class length. The difference in the proportion of Medium lateral children was tested using prop.test (RStudio).

Chart, histogram

Description automatically generatedLonger shoots (“Me” class, yellow-colored) are more present in higher rank nodes (i.e. longer parental shoots) (Figure 8). It is also noticeable how the majority of new shoots are single “Sh”, a few double “Sh” are present at the beginning-middle part of the parent shoot (Figure 7).

Figure 8: percentage of different lateral shoots combination in proleptic shoots according to parental rank node (1 is at the base of the shoot). Just nodes from 1 to 16 were analyzed because more frequent.

The same analysis is meaningless in sylleptic shoots because some data are missing: sylleptic length, and number of nodes in sylleptic. Due to that, the number of combinations inside sylleptic seems bigger than it is in reality. This is further work that could be interesting to invest in.

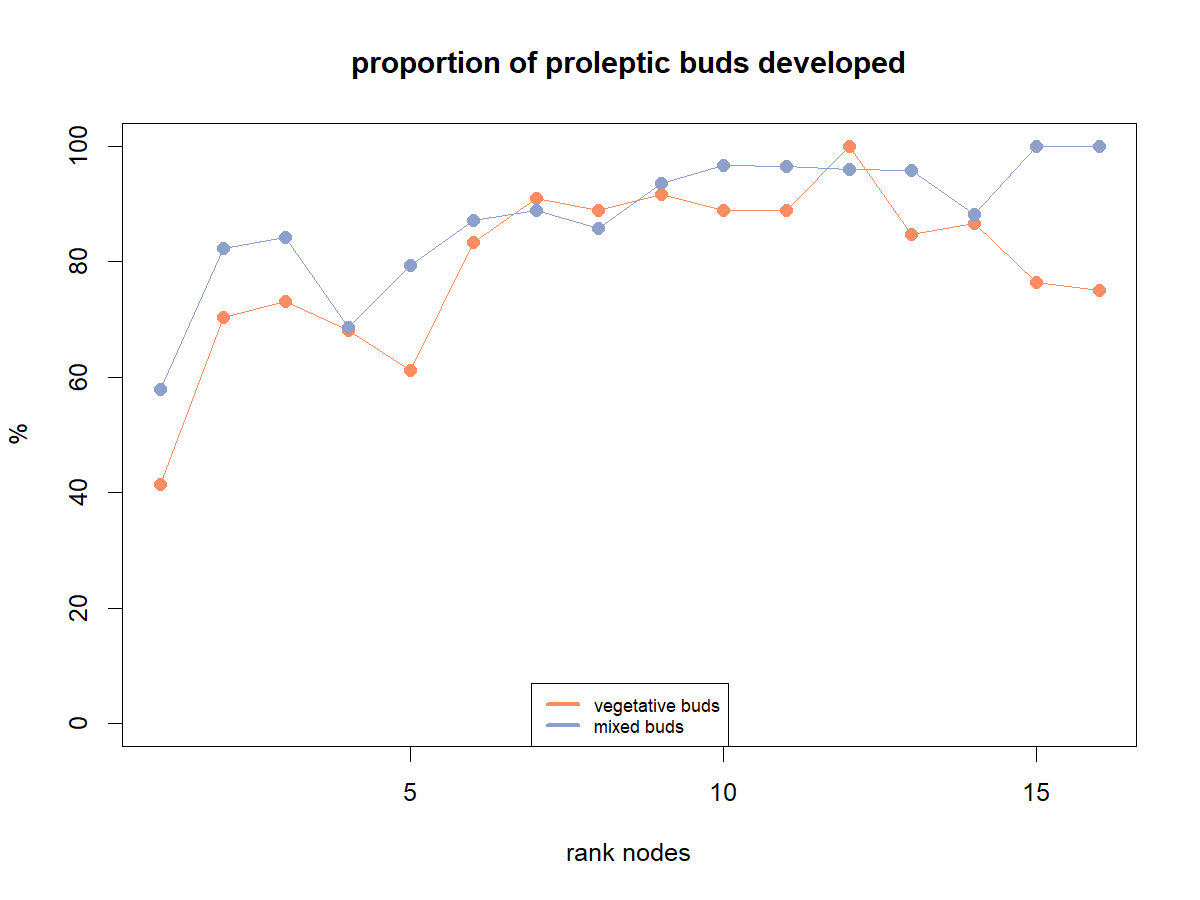
The percentage of bursting M or V increases with the rank node (Figure 9).

Figure 9: percentage of buds (vegetative or mixed) that sprouted, according to parental rank node (1 is at the base of the shoot). Just nodes from 1 to 16 were analyzed because more frequent.

*Apical shoot (year n+1)*

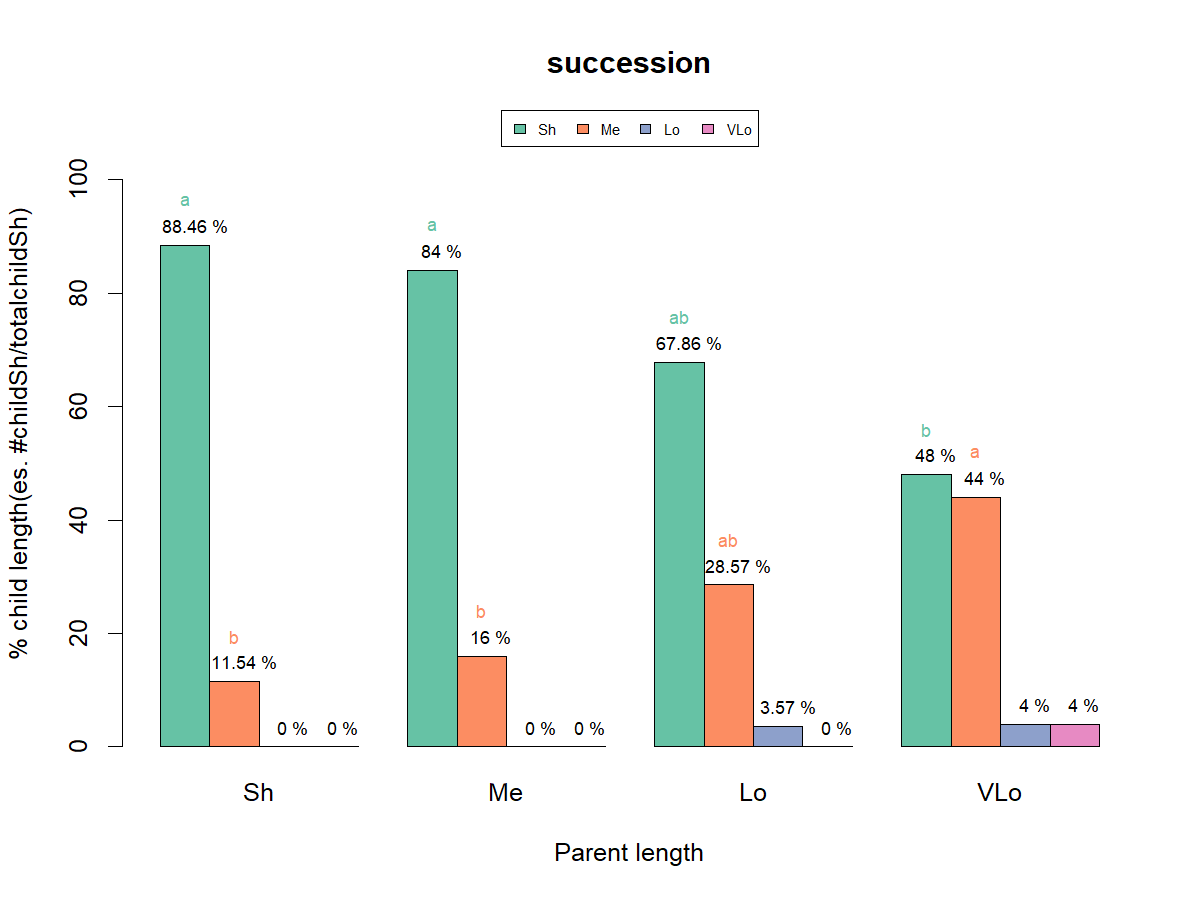
All of the buds sprout into shoots and the relationship between the length of **proleptic parent** and the length of apical children showed that the majority of successive shoots are short shoots for shorter parents, while their length increase with the increase of parent length (Figure 10)

Figure 10: percentage of apical child class length per each parental class length. The difference in proportion was tested using prop.test (RStudio).

### GLMs

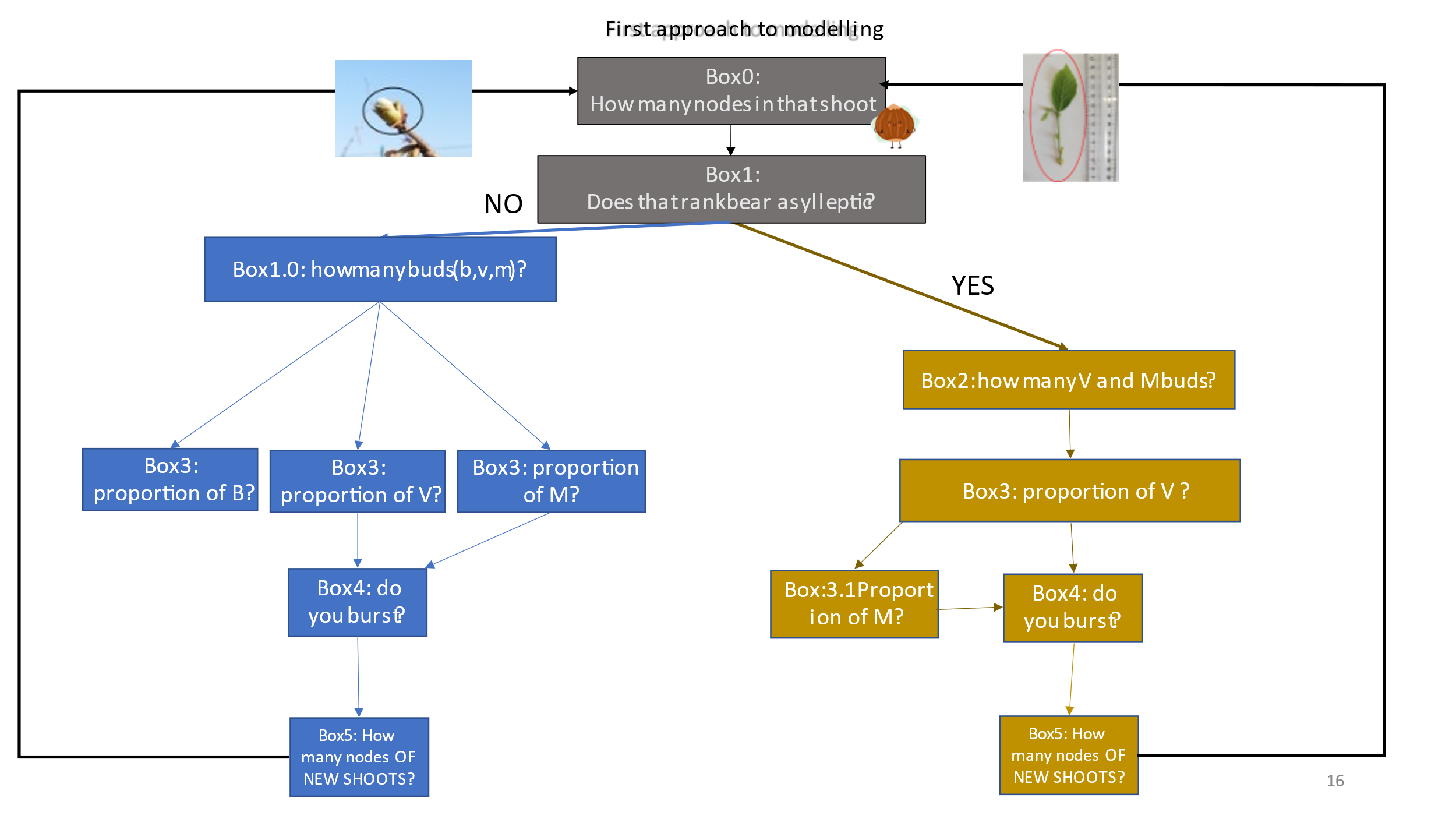
From the exploratory analysis was possible to draw the first logical scheme that drives the growth of hazelnut plant (Figure 11).

Figure 11: logical diagram to model hazelnut growth. Gray boxes are the common ones. Blue boxes are followed in the case of bud in proleptic shoots, while yellow boxes are followed in the case of bud in sylleptic shoots.

The first step is to answer the question “**how many nodes has that shoot?”** Then **“does that rank bear a sylleptic?”**. If the answer is **YES** the simulation will follow the blu boxes, alternative, it will follow the yellow. In both cases the questions regarding:

* Number of buds;
* Proportion of buds;
* Probability of bursting of M and V;
* Number of nodes of new shoots

This last question is the one that closes the loop and connects the new shoots' information with the next steps. For each box, the dependent variable was explained by some predictors using *glm,* *multinomial regression,* or *non linear models*in RStudio. Each model tested the relation between the dependent variable (i.e. Number of MandV buds) with many predictors (i.e. Parent length, number of nodes, distance from median node). Every time that one parameter resulted not significant it was deleted from the equation and the model was run again. When the best model was found (e.g. all the predictors had at least 1 star) permutation models with, significance level put as 0.001, were done. Then the final model was chosen according to the best AIC.

#### Box0: how many nodes are in that shoot?

Chart, scatter chart

Description automatically generatedThe relationship between the length of annual shoots and the number of nodes was significant (p<0.001) (Figure 12).

Figure 12: relationship between length(cm) and number of nodes in annual shoots. length(node)=1.81\*length(cm)0.57.

#### Box1: does that rank bear a sylleptic?

Chart, histogram

Description automatically generatedBecause sylleptic shoots were almost in the median zone along with the parent shoot (Figure 4), the distance from each node to the median one was computed per shoot. Then, the absolute value for this parameter was calculated. The percentage of sylleptic shoots decreases with distance from the median node (Figure 13)

Figure 13: probability of having sylleptic shoots according to distance from parental median node. y= 1/1+e(-(0.4315-0.2672\*x)).

#### Box2\_sylleptic: how many V and M buds?

Chart

Description automatically generatedThe average sum of mixed and vegetative buds, in sylleptic shoots, was related to parent length and distance from parental median node (Figure 14)

Figure 14: average sum of mixed and vegetative buds in sylleptic shoots related to parent length(cm) and distance from median node. Y=e(0.613354)+e0.016119\*parent\_length(cm)+e-0.132113\*median\_distance.

#### Box3\_sylleptic: what is the proportion of V?

Analysis with glm was conducted. They reported a correlation with parent length in nodes (**Error! Reference source not found.**)

However, the model underestimated the proportion of vegetative buds for short shoots (left part of the graph Figure 15). Moreover, the parameter parent length is not the length of the sylleptic but of the parent that bears it. Because of those reasons the model was discarded and was assumed that the proportion of V buds was constant (0.55±0.02) as suggested in Figure 2. Accordingly, the proportion of M buds was computed as 1-%V.

*Chart, histogram

Description automatically generatedBox4\_sylleptic: do you burst?*

Figure 15: graph is drawn using coefficients of glm in Figure 19. The pink line represents predicted values, while blue lines are real ones.

Text

Description automatically generatedThe probability of bursting was computed with a glm with interaction (Figure 16). This strategy was chosen because the probability of bursting was influenced by the fate of the bud itself and the presence of other M or V buds in the same sylleptic (Figure 16).

Figure 16: summary of glm with interaction to compute the relationship between the probability of bursting and the chosen predictors (i.e. fate, other M buds and fate, other V buds and fate).

Chart

Description automatically generatedThe results show that if the bud is V the probability of bursting decreases in the presence of other V and if the fate is M the probability of sprouting decreases if there are other M in the same sylleptic (Figure 17).

Figure 17: graph showing the results of glm. The left side shows that if the bud is M and there are other M, the probability of having a new shoot (i.e. probability of bud bursting) decreases steeply than if there are other V. The opposite happens if the bud Is V

#### Box5\_sylleptic: how many nodes of new shoots?

Chart

Description automatically generatedNumber of nodes of new shoots was related to distance from median node (Figure 18).

Figure 18: Relationshio between new nodes and distance from median node. in Sylleptic shoots. y=3.36+0.17x (R2=0.02; p<0.005)

#### Box2\_proleptic: how many B, V, and M buds?

The total number of buds (B, M, V) in proleptic shoots was not related to anything.Text

Description automatically generated Thus, we decided to take the number of buds as a random number in a Poisson distribution with λ equal to exp(0.11689) and k equal to 1 (because the node we are investigating is just 1) (Figure 18)

Figure 18: null glm to have the intercept. The exponential of the coefficient is λ in the Poisson distribution.

#### Box3\_proleptic: what is the proportion of V, M, and B?

Text

Description automatically generatedTo answer this question a multinomial logistic regression was done (Figure 19). Rank node was the best variable to describe the proportion of V, M, and B.

Figure 19: multinomial regression to evaluate the proportion of V, M, and B, according to the parental rank node.

The results were not satisfying (Figure 20). Especially for the final rank nodes where predicted M buds were more than predicted V, while, using real data V is more than M (Figure 20). **Is it because I made multinomial regression considering rank node as a continuous variable(type=integer) instead of categorical (type= factor)** ??

*Chart, bar chart, histogram

Description automatically generatedBox4\_proleptic: do you burst?*

Figure 20: proportion of V, B, and M buds according to rank node, in proleptic shoots. Pr(B)=e(0.88-0.43\*x)/(1+e(0.88-0.43\*x)+e(-0.05+0.03\*x)); Pr(M)=e(-0.05+0.03\*x)/(1+e(0.88-0.43\*x)+e(-0.05+0.03\*x)); Pr(V)=1-Pr(B)-Pr(M).

**Text

Description automatically generated**The probability of bursting was computed with a glm with interaction(Figure 21). This strategy was chosen because the probability of bursting was influenced by the fate of the bud itself and the presence of other buds (M or V)in the same node.

Figure 21: summary of glm with interaction to compute the relationship between the probability of bursting and the chosen predictors (i.e. fate, sibling buds (other M+ other V in the same node).

*Graphical user interface

Description automatically generated*The results show that if the bud is V or M, the probability of bursting decreases in the presence of other buds (dibbling buds) at the same node. The V buds highly decrease the probability of bursting When distance increase. The probability of bursting increases, both for V and M when the length increase (Figure 22).

Figure 22: graph showing the results of glm. On the left upper side, it is possible to see that the bursting of both V and M buds is affected by the presence of sibling buds in the same node. The distance from the median rank node affects more V buds, while lengt

*Box5\_proleptic: how many nodes of new shoots?*

Graphical user interface, chart

Description automatically generatedThe number of nodes of new shoots is related by the cumulative effect of parent length and fate ()

Figure 30: Relationship between number of nodes of new shoots and parent length.fateV: y=2.39+0.05\*length; fateM: 3.34+0.02\*length. (R2=0.10. p<0.001)

Lpy

In github: \HazelnutFSPM\LpyModel\hazelnut.lpy

# To Do:

1. ~~Analysis fro box5 for proleptic and sylleptic~~
2. Evaluate the length distribution on hazelnut
3. Glms for apicals
4. Continue lpy