

PDG Documentation, version 3

Sylvie Muratorio, September 2017

Contents

1)	PDG general purpose and main originalities	2
1.1.	<i>General purpose</i>	2
1.2.	<i>Modeling plastic and genetic variation of life-history traits</i>	2
1.3.	<i>Modeling mortality</i>	4
	At adult stage	4
	At seedlings stage	5
2)	General principle of use	6
2.1.	<i>Initialization of the simulations</i>	6
2.2.	<i>Running a simulation</i>	6
2.3.	<i>Sequence of events within a year</i>	8
2.4.	<i>Simulation options</i>	8
2.5.	<i>Results of simulations</i>	9
2.6.	<i>Main limitations</i>	9
3)	Model architecture and main features	11
4)	References	14
Appendix 1: Parametrization of genotypes, allelic effects and heritability in PDG		15
1)	The quantitative genetic framework	15
1.1.	From genotype to phenotype:	15
1.2.	Allelic effects, genetic variance and heritability	16
2)	Relation with the variables in Capsis GENETICS library and in PDG model	16
2.1.	Individual genotypes:	16
2.2.	Description of the population-level genetic composition	17
2.3.	Population-level allele effects and heritability:	18
2.4.	Individual genetic, environmental and phenotypic values	19
3)	Parametrization of genotypes, allelic effects and heritability in PDG inventory file	20
Appendix 2 : inventory, climate and parameter files		24
1)	Inventory file	24
1.1.	General parameters	24
1.2.	GENETICS library definitions	28
1.3.	Plot and cells definition	30
1.4.	Individual trees definition	32
1.5.	R script to generate inventory file	33
2)	Climate file	34
3)	Species parameters	34
Appendix 3: Output files		35
1)	yearlyResults.csv	35
Appendix 4 : Some detailed examples		37

1) PDG general purpose and main originalities

1.1. General purpose

The Physio-Demo-Genetics model (denoted PDG) is an individual-based, spatially explicit mechanistic simulation model, coupling (1) a physiological module simulating individual tree responses to the environment (the CASTANEA model, Dufrêne *et al.* 2005); (2) a demographic module simulating tree survival, reproduction and pollen and seed dispersal; and (3) a quantitative genetics module controlling the heritability of key life history traits. The initial objectives of PDG were twofold: first, PDG aims at simulating the temporal and spatial variation of life-history traits in a tree population within a heterogeneous environment, while accounting for genetic and plastic components of this variation; we focus here on those functional traits that mediate individual tree response to climate, and whose values are computed by the CASTANEA model. Secondly, PDG allows to investigate how this inter-individual variation affect adaptive tree population dynamics across generations. PDG model runs on a yearly time step, although its physiological module (CASTANEA) runs on a daily time step. PDG model was used to investigate the plastic and genetic components of the variations of the timing of budburst along an altitudinal gradient of *Fagus sylvatica* (the European beech), in the reference publication of PDG (Oddou-Muratorio & Davi 2014).

PDG takes advantage of the modularity of the Capsis platform. The physiological module is based on the CASTANEA library, implemented by Hendrik Davi in Capsis. The quantitative genetics module is based on the GENETICS library implemented by Ingrid Seynave and Christian Pichot in Capsis (Seinave & Pichot 2004). The framework of the demographic module is similar to the one used in other previous models of Capsis (eg Luberon, Alisier).

This documentation first details the main originalities of PDG, that is the modeling of plastic and genetic variation in functional traits (paragraph 1.2 below) and the mortality process (paragraph 1.3 below). In paragraph 2, the general principle of use of PDG is described, while paragraph 3 describe the architecture and implementation of PDG.

1.2. Modeling plastic and genetic variation of life-history traits

Presently, PDG allows two life-history traits to be considered as both plastically and genetically variable: the timing of budburst (TBB, as in the reference publication) and water use efficiency (WUE). Potentially, any other trait modeled in CASTANEA can also be considered, but this will require informatics development by the user. PDG uses a classical quantitative genetic framework to model phenotypic and genetic variation for TBB and WUE.

Model for TBB: In CASTANEA, the timing of budburst (TBB) is usually simulated following (eq. 9- 11 in Dufrêne *et al.* 2005):

$$R_{\text{fcBB}} = \begin{cases} T - T_2 & \text{if } T > T_2 \text{ and } N > N_{\text{start1}} \\ 0 & \text{if } T \leq T_2 \text{ or } N < N_{\text{start1}} \end{cases} \quad (\text{Eq 1})$$

where R_{fcBB} is the rate of forcing for bud break, T the mean daily temperature, T_2 the base temperature, N the day of year and N_{START1} the date of onset of rest.

$$S_{\text{fcBB}} = \sum_{N_{\text{START1}}}^N R_{\text{fcBB}} \quad \text{if} \quad S_{\text{fcBB}} < F_{\text{critBB}} \quad (\text{Eq 2})$$

$$TBB = N \quad \text{if} \quad S_{\text{fcBB}} \geq F_{\text{critBB}} \quad (\text{Eq 3})$$

with S_{fcBB} the state of forcing, F_{critBB} the critical value of state of forcing for the transition from quiescence to the active period and TBB the day when bud break occurred. Note that T_2 , N_{start1} and F_{critBB} are parameters, while R_{fcBB} , S_{fcBB} and TBB are dynamic variables of CASTANEA.

The simple one-phase phenological sub-model described by the equations 1-3 assumes that the variation in TBB is driven only by forcing temperatures; this hypothesis is valid for our case study (*Fagus sylvatica* on Mont Ventoux), but should be considered with caution in other cases. Note that diverse refined phenology models (including chilling) were recently integrated from Phenofit into CASTANEA, and could allow different hypothesis to be made.

The main originality of PDG is to consider F_{CRITBB} parameter as genetically determined and variable among individuals. The value of F_{critBB} is determined by ten independent biallelic loci with purely additive effects (see Appendix 1). The inter-individual variation in TBB thereby results from the inter-individual variation of mean daily temperature (T in equation 1) and from genotypic variation in F_{critBB} . This is the classical frame of quantitative genetics; note that we did not explicitly introduce genotype-by-environment interaction.

Figure 1A illustrates the variation of TBB along a temperature gradient as simulated in PDG. A given genotype (one color curve on Figure 1A) unfolds its leaves earlier when environment gets hotter because the sum of temperature it receives (i.e. the state of forcing S_{freBB}) reaches faster F_{CRITBB} . On the other hand, in a given environment, late and early individuals can be observed as the result of genetic variation for F_{CRITBB} : late individuals typically require higher F_{CRITBB} than early individuals (two different color curves on Figure 1A).

Model for WUE: water use efficiency (WUE) can be defined at leaf level as the ratio of assimilation (A) to stomatal conductance (g_{sH_2O}):

$$WUE_i = A/g_{sH_2O} \text{ (Eq 4)}$$

In CASTANEA, three main equations determine assimilation A and g_{sH_2O} (eq 1,2 and 4 in Dufrêne *et al.* 2005), following the model from Ball *et al.* (1987). First the carbon dioxide demand is:

$$A = V_C - R_d \text{ . (Eq 5)}$$

where V_C is the carboxylation rate and R_d the respiration during the night. Second, the carbon dioxide supply writes:

$$A = g_{sCO_2}(C_S - C_I) \text{ . . (Eq 6)}$$

where g_{sCO_2} is the stomatal conductance for CO_2 and $(C_S - C_I)$ the gradient of CO_2 between evaporative site and leaf surface.

Third, the carbon dioxide control which allows the calculation of g_{sH_2O} writes:

$$g_{sH_2O} = \frac{g_0 + g_1 \times A \times RH}{C_S} \text{ (Eq 7)}$$

where is RH the relative humidity in the surrounding air, g_1 depends on soil water stress and g_0 correspond to cuticle conductance of the leaf.

Re-arranging eq 7 shows that g_1 , the slope of the relation between photosynthesis and stomatal conductance, is inversely related to WUE :

$$\frac{A}{g_{sH_2O} - g_0/C_S} = \frac{C_S}{g_1 \times RH} \text{ . (Eq 8)}$$

Assuming $g_{sH_2O} \gg \frac{g_0}{C_S}$, the higher g_1 , the lower WUE.

In CASTANEA, g_1 is assumed to decrease linearly when soil water storage decreases. The effect of soil water stress on photosynthesis is mediated through g_1 :

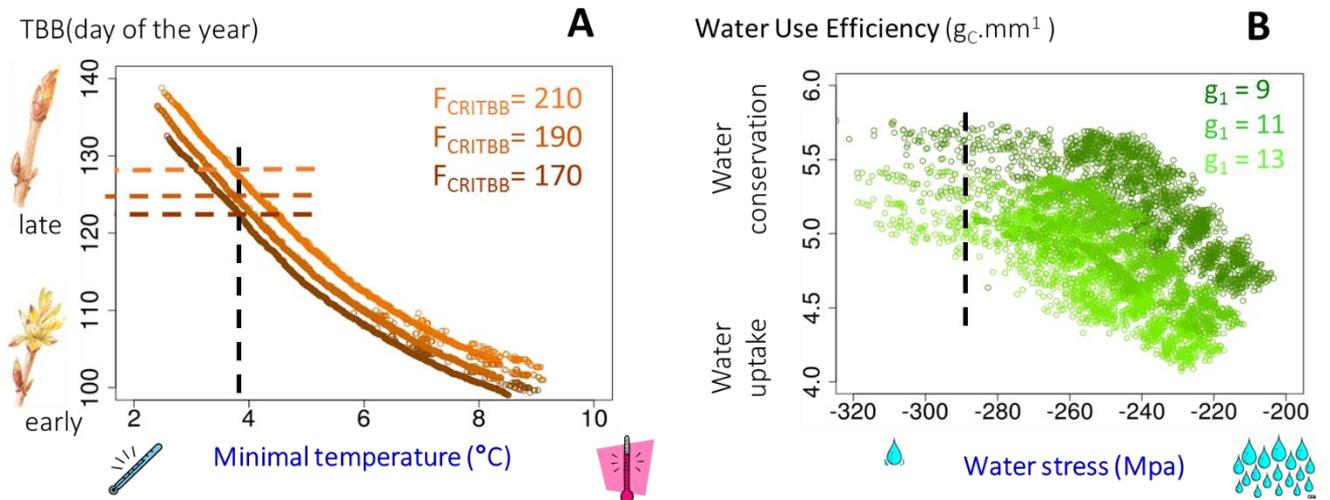
$$g_1 = (g_{1max} - g_{1min}) \times reduc + g_{1min}. \text{ (Eq 9)}$$

Where g_{1max} and g_{1min} are the maximal and minimal values taken by g_1 , and “reduc” is a soil water stress index varying between 0 (maximal soil water stress) and 1 (no water stress).

In PDG, we consider g_{1max} as a parameter genetically determined and variable among individuals. The value of g_{1max} is determined by ten independent biallelic loci with purely additive effects (see Appendix 1). The inter-individual variation in WUE thereby results from the inter-individual variation of soil water stress (reduc in eq 9) and from genotypic variation in g_{1max} . This is again the classical frame of quantitative genetics; note that as for TBB, we did not explicitly introduce genotype-by-environment interaction.

Figure 1B illustrates the variation of WUE along a water stress gradient as simulated in PDG. A given genotype (one color curve on Figure 1B) favors water conservation when water stress increases because g_1 decreases as “reduc” increases. On another hand, in a given environment, individuals favoring either water uptake or water conservation can be observed as the result of genetic variation for g_{1max} : individuals with high g_{1max} typically favor water uptake while individuals with low g_{1max} typically favor water conservation (two different color curves on Figure 1B).

Figure 1: PDG-simulated relationship between each considered life history traits (A : Timing of Budburst, TBB; B: Water Use Efficiency, WUE) and the pertinent environmental variable (A : Temperature; B: Water stress index) for different genotypes (A : different F_{CRITBB} ; B: different g_{1max})



1.3. Modeling mortality

At adult stage

The reserves produced by photosynthesis at a daily time step are allocated to growth and the level of reserve together with tree water potential are used to compute the probability of an adult tree to die **after each year of growth**. There are three different “causes” of mortality, associated with three different processes, thresholds, and probabilities:

- Below a **critical value of carbon reserve (CS_{tres})** at the end of the year, there is a non-zero probability for the tree to die because of carbon starvation (P_{mortCS})

- Below a **critical value of carbon reserve at budflush (BF_{tres})**, there is a non-zero probability for the tree to die because of carbon starvation (P_{mortBF})
- Below a **critical minimal daily value of water potential ($P_{min_{tres}}$)** there is a non-zero probability for the tree to die because of hydraulic failure (P_{mortHF})

Note that in each case, mortality is a stochastic process (for each tree for which a value is below the threshold, a random number is drawn and compared to P_{mortCS} , P_{mortBF} or P_{mortHF}).

At seedlings stage

Seedlings mortality occurs only once, at the seedlings-to-adult transition, and can be mediated either by (1) neutral (genotype-independent) density-dependence or (2) genotype-driven fitness differences among seedlings.

In both cases, a given maximal number of young adults per cell is allowed ($nbAdultPerCell$, parameter chosen by the user).

- In the first case (density-dependence), seedlings too close from each other are randomly killed. Our density-dependence mortality process uses as criteria the size of the canopy at adult age (parameter $maxCanopyRadius$, computed during simulations) and the maximum percentage of overlapping canopy ($maxOverlappingPercentage$, parameter chosen by the user). Random mortality occurs until the target number of young adults per cell is reached.
- In the second case, seedlings are sorted according to their fitness (typically estimated through their height) and the less fit seedlings are killed until the target number of young adults per cell is reached.

2) General principle of use

2.1. Initialization of the simulations

To run PDG, the user must provide:

- A **species parameter file** specifying SOME *Castanea* parameters, ie all the parameters defined at tree species level. This file is named *CastaneaSpecies3.txt*, its values should not be changed and it is located in the directory *data/physiodemogenetics/*
- An **inventory file** containing:
 - The parameters related to the simulation design
 - The parameters of the demographic module
 - The parameters of the genetic module, in particular the genetic and environmental controls of F_{CRITBB} and $g_{1\text{max}}$ (heritability, allelic effects), as well as on the genetic map of the considered species.
 - The parameters of CASTANEA which are not defined in the *CastaneaSpecies3.txt* file, ie all the parameters related to simulation option in CASTANEA, as well as parameters related to plot (soil etc..)
 - The definition of the plot (list of cells with their properties),
 - A list of initial trees with their properties at the beginning of the simulation (year y_0); Note that all the trees are individual trees (no multipleGenotype allowed); depending on their age, trees are considered either as adult or seedlings (by default, the age at which tree become adult, *adultAge*, is 40 years old). In the present version of PDG, all the trees belong to the same single species (see Appendix 2 for the list of possible species to consider).The inventory file should be located in the directory *data/physiodemogenetics*.
- A **climate file** containing the values of six climatic variables (global radiation, rainfall, wind speed, air humidity and temperature) on a daily basis. This file should be located in the directory *data/physiodemogenetics/climate*.

All these files, as well as the list of parameters to control are detailed in Appendix 2.

2.2. Running a simulation

Originally, PDG was developed to run simulation in even-aged populations, containing either only “active” adult trees, or “active” seedlings (“active” meaning growing and eventually reproducing trees). For such even-aged simulations, the user can run “evolution” from an initial year y_0 to a next year (y_n) using 3 different modes:

- **mode “g”** (for growth): all the trees are considered as adult, and growth and mortality are computed for all the trees of the stand using CASTANEA. Note that neither growth nor mortality are affected by competition (each tree grow as a dominant tree). Although adult trees can produce seeds, there is no recruitment of young seedlings in evolution with mode “g” step.
- **mode “r”** (for reproduction): all the trees are considered as adult and reproductive; growth and mortality are computed for all the trees of the stand using CASTANEA; Seeds are dispersed and eventually germinate, and the newly established young seedlings do not grow/die.
- **mode “s”** (for seedlings): all the trees are considered as seedlings, and growth is simulated only for a subset of them using CASTANEA. Here, we assume that CASTANEA model is valid for seedlings, while it was calibrated for adult trees. The only difference between mode “g” and “s” (besides the subsampling of growing individual) is that seedlings cannot die (critical values CS_{tres} , BF_{tres} and $P_{\text{min}_{\text{tres}}}$ are set to non-limiting values). There is no recruitment of young seedlings during a “s” step (no adult trees present). At the end of

phase “s”, the dimensions of all the seedlings that did not grow are updated based on the dimension of the seedlings that grew. Then, seedlings density-dependent mortality is simulated, considering either (1) neutral (genotype-independent) mortality or (2) genotype-driven mortality.

For each evolution step in mode g, r or s, the user has to indicate which climate file is used, and specify the initial year, month and day from which evolution begins.

These different modes of evolution can be combined. For instance, using an inventory file where all the initial trees have an age of 10 years old (seedlings), and the evolution sequence “s30 g50 r60” leads to the following evolution:

- From year $y_0=0$ to year $y_n=30$, a subset of the trees from the initial stand are grown as seedlings (ie without mortality);
- At year 30, the dimensions of all seedlings are updated as function of the dimensions of the seedlings that grew, and mortality of seedlings can occur. As the seedlings that survived are 40 years old, they switch to the category “adult trees”.
- From year 30 to year 50, all of the adult trees grow/die up (reaching age 60).
- From year 50 to year 60, all of the adult trees grow/die/reproduce (reaching age 70), and a cohort of new seedlings of age 10 at year 60 will eventually be recruited in the stand.

Important note: it is the responsibility of the user to take care that the values of trees characteristics given in the inventory file (for instance their age), the parameters values (for instance adultAge) and the sequence of evolution are consistent one with each other.

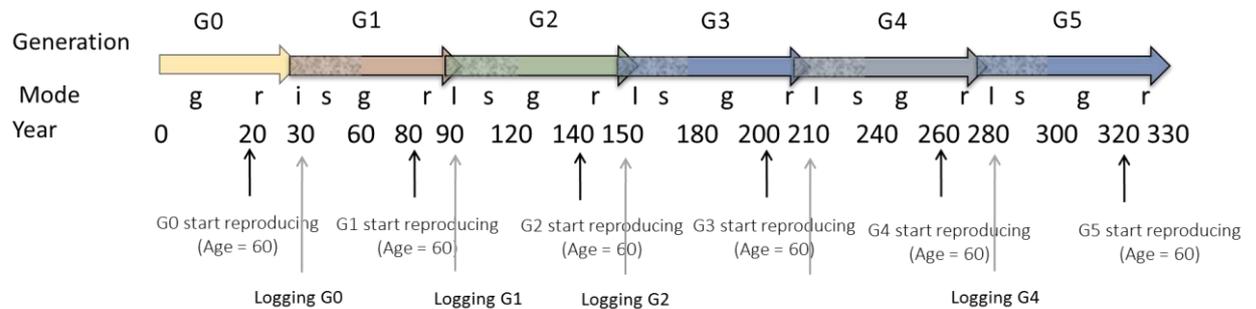
There is a fourth possible mode, “i” (for intervention), which consist in cutting all the adult trees (with age >40) at a given year y_n . Mode “i” thus corresponds to an instant operation, and not to an evolution (no climate required) By combining these four modes, the user can simulate evolution across successive and non-overlapping generations. Table 1 summarizes a typical sequence of evolution simulating 5 non-overlapping generations in an even-aged context.

Table 1: sequence of evolution simulated when using the instruction “g20 r30 i30 s60 g80 r90 i90 s120 g140 r150 i150 s180 g200 r210 i210 s240 g260 r270 i270 s300 g320 r330”

From	To	mode	Type	generation	age
0	20	g	Adult evolution without regeneration	G0	40-60
20	30	r	Adult evolution with regeneration	G0	60-70
30		i	Cut of G0		
30	60	s	Seedlings evolution	G1	10-40
60	80	g	Adult evolution without regeneration	G1	40-60
80	90	r	Adult evolution with regeneration	G1	60-70
90		i	Cut of G1		
90	120	s	Seedlings evolution	G2	10-40
120	140	g	Adult evolution without regeneration	G2	40-60
140	150	r	Adult evolution with regeneration	G2	60-70
150		i	Cut of G2		
150	180	s	Seedlings evolution	G3	10-40
180	200	g	Adult evolution without regeneration	G3	40-60
200	210	r	Adult evolution with regeneration	G3	60-70
210		i	Cut of G3		
210	240	s	Seedlings evolution	G4	10-40
240	260	g	Adult evolution without regeneration	G4	40-60
260	270	r	Adult evolution with regeneration	G4	60-70

270		i	Cut of G4		
270	300	s	Seedlings evolution	G5	10-40
300	320	g	Adult evolution without regeneration	G5	40-60
320	330	r	Adult evolution with regeneration	G5	60-70

Figure 2: Scheme of the evolution simulated when using the instruction “g20 r30 i30 s60 g80 r90 i90 s120 g140 r150 i150 s180 g200 r210 i210 s240 g260 r270 i270 s300 g320 r330”



Another version of PDG for uneven-aged stand is under development, and involves the coupling with the SamsaraLight library.

2.3. Sequence of events within a year

Within a given year, the sequence of successive events depends on the mode of the simulation, and are coordinated by the following classes:

- PDGGrowerAdult for simulations in mode “g”
- PDGGrowerAdultWithRegeneration for simulations in mode “r”
- PDGGrowerSeedlings for simulations in mode “s”

These sequences of events are summarized in Table 2 below.

Table 2: Sequence of events within a given year of simulation

Event	Mode	Main Classe(s) involved
Choice of the subset of seedlings that will grow	s	PDGFmGrowth
Growth of the trees (adult or seedlings)	g,r,s	PDGFmGrowth, PDGFmCell
Actualization of the dimensions of waiting seedlings	s (last year only)	PDGFmGrowth
Reproduction	r (every year)	PDGSimpleRegeneration, PDGSpread
Intervention	i	PDGModel

2.4. Simulation options

PDG can be run either in graphical mode (typically for test or development), or in script mode (advisable for routine simulations).

Command line for script mode in Linux:

Pre-evolution : it is theoretically possible to run a neutral pre-evolution, with the objective to bring the initial population to a migration-drift equilibrium. To that aim, the parameter

preEvolutionTime should be set to a non-zero integer value (for instance 5) specifying the number of generations of pre-evolution to be run. Preevolution includes only mating (no growth nor mortality) across non-overlapping generations. It was implemented in the context of the study of Oddou-Muratorio and Davi 2014, and has several constraints (see Appendix 2). Presently, we rather advice to use initial stand already at migration-drift equilibrium, or possibly at selection-migration-drift equilibrium using other simulation software to generate inventory files (see Appendix 2).

2.5. Results of simulations

If the inventory file is located there: data/physiodemogenetics/inventory.inv, the main results are stored in the directory data/physiodemogenetics/output-inventory, as several text files. The following files are written automatically:

- The main result file is called “**yearlyResult**” (csv file), and can be loaded in R. It contains for each tree and year the output of the CASTANEA simulations
- A file called “**understanding**” contains information printed at the beginning (value of some key parameters) or on the course of simulation, such as the number of individuals present on the scene at each year, etc... Useful to detect problems
- The realized fitness of each adult tree is computed each year as the number of offspring they sire as male or female parent (note that regeneration occurs only in mode “r”). These values are stored in the file “**YearlyIndivFitness**”, together with several parental “extended phenotypes” (FCRITBB and glmax values, as well as growth primary production (GPP), evapotranspiration (ETR), water use efficiency (computed as GPP/ETR), ring width and seed production

Additional files can be exported in script mode; in the script file clusterScript, the following files are filled at the beginning of a mode “r”:

- The realized fitness of each adult tree from generation X is computed as the number of offspring alive at generation X+1 (at the beginning of mode “r”). These fitness values are stored in the file “**IndivFitness**”, and are more “integrated” than the values of the file “**YearlyIndivFitness**”, as they include the effect of mortality.
- Realized distances between each mother-parent and its offspring (ie seed dispersal distance) or between male and female parents (ie pollen dispersal distance) are stored in the file “**RealisedDisp**”.
- **AllelicFrequencies** contains for each individual (in line) and each SNP locus (in column), the dose of the first allele (0 : homozygote 22; 0.5 = heterozygote; 1 : homozygote 11)
- Several files are exported (typically at the beginning of “r” step) contain the adult tree genotypes and position to be analyzed with classical population genetic software (for instance Genepop, Spagedi).

Inventory files can be exported under the name GXAdult.inv (typically at the beginning of each generation X); they can be used to launch new simulations.

All these files are described in Appendix 3.

2.6. Main limitations

The present version of PDG has several limitations:

- Each tree bears its own CASTANEA model, which simulates carbon and water fluxes independently from the other trees of the stands; so each tree behaves as a dominant tree.
- There is neither mortality due to competition, nor mortality due to perturbation such as storms.

- A single tree species can be simulated (no mixture of different species)
- Simulations are aimed to follow an “even-aged trajectory”, such as can be achieved through silviculture; but the whole simulated stand correspond to a single management unit and there is no gene flow from other stands.
- Typically, to run 5 generations in a stand of 2500 individuals, the simulation time is 2.5 days.

3) Model architecture and main features

As classically in Capsis, sources files of main classes for PDG can be found in four directories :

- physiodemogenetics/extension/ioformat : 8 classes to import files (ie reading the inventories) or export files (genetic exports)
- physiodemogenetics/gui : 2 classes to run PDG in the graphical user interface
- physiodemogenetics/model : 29 main classes defining the module
- physiodemogenetics/myscripts : 2 classes maintained to run PDG in mode script

We detail here only the classes of the /model directory

Table 3: List of the 29 classes of the physiodemogenetics/model directory

Color code: **Data class** [contains mainly object or parameters]; **Process class** [contains mainly methods corresponding to key ecological processes]; **Tool data or process class** [contains mostly data or methods implemented as tools for convenience]. Type: classical (in most model of CAPSIS), original (only in PDG), or Common (Common to demo-genetic tree model)

Class	Type	Contents
Functions	Original	Contains mathematical functions called by PDGSpread.java
PDGCell	Classical	Defines the quadratic cell of a PDGplot and its methods; [PDGCell]s hold [PDGTree]s.
PDGDistanceToParameterValuesorter	Original	A comparator which allows to sort integer value (typically the value of target physiological parameters in PDG) as function of their distance to a target value. Used by PDGGrowerSeedling (through PDGFmGrowth)
PDGEvolutionParameters	Original	Embeds the evolution parameters, to be transmitted among classes (typically between script and model classes)
PDGFCRITBBThenTreeIdValueComparator	Original	A comparator which allow to sort trees as function of FCRITBB
PDGFitnessReportingMap	Original	A map connecting each “waiting” seedling to its reference “growing” seedling to report the fitness value after growth
PDGFmCell	Original	Defines the Castanea cell attached to each tree and its methods
PDGFmGrowth	Original	Implements the methods for: <ul style="list-style-type: none"> – Growing a tree during a specified period of time using the Castanea library (through PDGFmCell) – Select growing seedlings for a mode “s” simulation
PDGGlmaxThenTreeIdValueComparator	Original	A comparator which allow to sort trees as function of glmax
PDGGeneticOutput	Original	Implements the methods to export genetic results
PDGGrower	Original	Abstract class defining the methods for the evolution step
PDGGrowerAdult	Original	Defines the methods for the evolution step at adults stage without regeneration
PDGGrowerAdultWithRegeneration	Original	Defines the methods for the evolution step at adults stage with regeneration
PDGGrowerSeedling	Original	Defines the methods for the evolution step at seedlings stage (without regeneration)
PDGInitialParameters	Classical	Contains parameters definition and value for PDG
PDGLoiselleKinshipCoefficient	Common	Contains methods to compute spatial genetic structure within a tree population
PDGMethodProvider	Common	Contains methods or call methods to compute spatial

		genetic structure within a tree population, or to compute stand characteristics
PDGModel	Classical	
PDGModelLight	Classical	
PDGPlot	Classical	Define the physical map containing the PDGStand and its methods; the PDGPlot is made of several PDGcells
PDGRapidEvolution	Original	Define the methods for pre-evolution
PDGSimpleRegeneration	Original	Define the methods for the regeneration step (mating, seed + pollen dispersal)
PDGSpecies	Classical	Define the genetic property attached to the considered population of each species.
PDGSpread	Original	Define the methods for seed + pollen dispersal
PDGStand	Classical	Define the stand (a collection of PDGTrees) and its method
PDGTreeIdThenProbaToSurviveComparator	Original	A comparator which allow to sort trees as function of their probability to survive
PDGTree	Classical	Define the tree (genotypable, spacialized) and its methods
PDGTreeFitnessThenTreeIdValueComparator	Original	A comparator which allow to sort trees as function of their fitness
PDGWriter	Original	Contains method to write result in output files

The **data classes** are linked one to each other, and also with classes of CAPSIS belonging to the kernel or to libraries; in particular, the class PDGTree is central in the philosophy of PDG, as it is connected both to the GENETICS and the CASTANEA libraries (Figure 3). Similarly, the **process classes** of PDG are linked one to each other, as illustrated in Figure 4.

Figure 3: Relationship between PDG **data classes** listed in Table 3 above and other classes from the GENETICS library (in orange) or from the CASTANEA library (in green), or from CAPSIS kernel (in black)

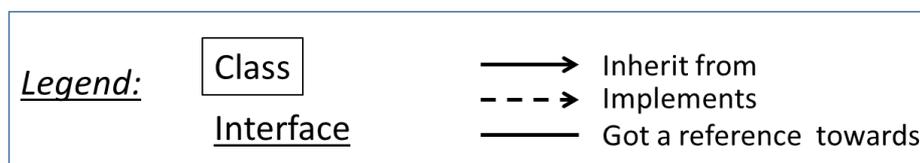
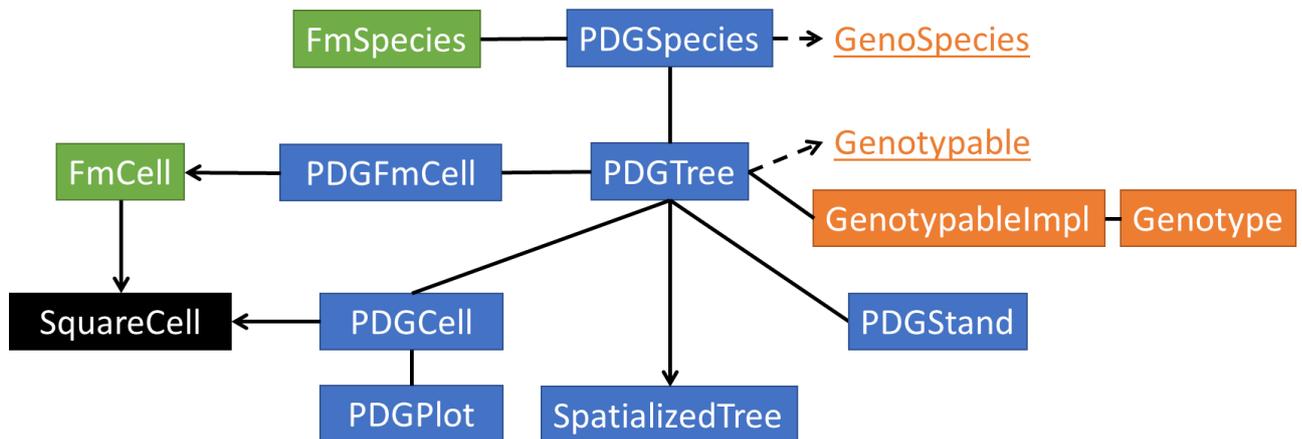
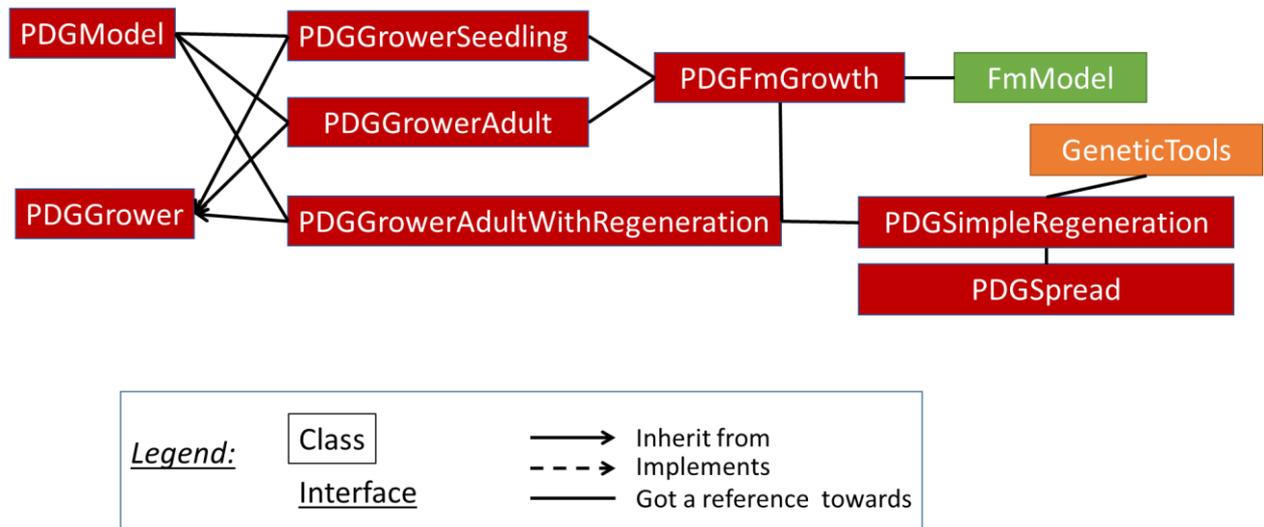


Figure 4: Relationship between PDG *process classes* listed in Table 3 and other classes from the GENETICS library (in orange) or from the CASTANEA library (in green).



A main originality of PDG model is that there the historical sequences of steps is not stored during and evolution. This means that at a year y_n , the user cannot access to previous stands at year $<n$. All the information has to be exported in files along the simulations.

4) References

- Ball JT, Woodrow IE, Berry JA (1987) A Model Predicting Stomatal Conductance and its Contribution to the Control of Photosynthesis under Different Environmental Conditions. In: *Progress in Photosynthesis Research* (ed Biggins J), pp. 221–224. Springer Netherlands.
- Bost B, de Vienne D, Moreau L, Dillmann C, others (2001) Genetic and nongenetic bases for the L-shaped distribution of quantitative trait loci effects. *Genetics*, **157**, 1773–1787.
- Dufrêne E, Davi H, François C *et al.* (2005) Modelling carbon and water cycles in a beech forest. *Ecological Modelling*, **185**, 407–436.
- Falconer DS (1960) *Introduction to quantitative genetics*. New York,: Ronald Press Co.
- Oddou-Muratorio S, Davi H (2014) Simulating local adaptation to climate of forest trees with a Physio-Demo-Genetics model. *Evolutionary Applications*, **7**, 453–467.
- Seinave I, Pichot C (2004) *docGeneticsMars04_fc.pdf*. INRA.
http://capsis.cirad.fr/capsis/help_en/genetics?&#documentation

Appendix 1: Parametrization of genotypes, allelic effects and heritability in PDG

(following notations in Falconer 1960 and Bost *et al.* 2001)

In PDG, individual genotype at given set of loci determines the value of two CASTANEA parameters (F_{CRITBB} and g_{1max}) which themselves determine, in interaction with environment, the individual values of the two traits of interest (respectively TBB, the timing of budburst, and WUE, Water Use Efficiency). The inventory file contains all the information related to the relation between genotype and the value of F_{CRITBB} and g_{1max} , as well as the distribution of genotype at initiation (see Appendix 2 for an example of such an inventory file). The objective of this appendix is to 1) introduce notations of for the main quantitative genetic parameters of interest in PDG, 2) relate these parameters to the variables of CAPSIS and 3) describe in practice how F_{CRITBB} and g_{1max} are computed, and related to the list of genotypes given in the inventory file.

1) The quantitative genetic framework

1.1. From genotype to phenotype:

We consider here F_{CRITBB} and g_{1max} as two “parametric traits” to model as function of genotype. Note that F_{CRITBB} and g_{1max} are in fact CASTANEA parameters controlling traits (TBB and WUE, see details below). We assume pure additivity, without dominance nor epistasis. Moreover, we consider that the loci coding for the parametric traits are all bi-allelic (SNP like), with possible alleles A1 and A2. In the following, we develop general notations for the relationship between genotype and phenotype at a parametric traits T_P .

Table A1.1 Definition of the genetic value m_l at a given locus l

Genotype	Frequency	Value (Bost et al)
A1 _l A1 _l	p_l^2	$m_l + \alpha_l$
A1 _l A2 _l	$2p_lq_l$	m_l
A2 _l A2 _l	q_l^2	$m_l - \alpha_l$
mean		$m_l + (p_l - q_l) \alpha_l$

Bost *et al.* (2001) consider a non-centered value, each locus l can contribute to a part m_l of the trait T_P , allelic effect α_l is an additive correction of locus contribution for the heterozygote (note: Bost *et al.* (2001) assume that all m_l are equal, and so do we)

Assuming Hardy-Weinberg, the mean genetic value for locus l is :

$$A_l = p_l^2 (m_l + \alpha_l) + q_l^2 (m_l - \alpha_l) + 2 p_l q_l m_l$$

$$A_l = m_l (p_l^2 + q_l^2 + 2 p_l q_l) + (p_l^2 - q_l^2) \alpha_l$$

$$A_l = m_l + (p_l^2 - q_l^2) \alpha_l = m_l + (p_l + q_l) (p_l - q_l) \alpha_l = m_l + (p_l - q_l) \alpha_l$$

Mean genetic value for n loci :

$$A = \sum_n G_l = n m_l + \sum_n (p_l - q_l) \alpha_l, \text{ with } m_l = \mu_T / 10 \text{ (eq A1.1)}$$

1.2. Allelic effects, genetic variance and heritability

Table A1.2 Definition of allelic effects at locus l

Gamete	Genotypes contributed	Mean offspring value	Mean effect (deviation from the mean of genotypes)
A1	p A ₁ A ₁ + q A ₁ A ₂	p(m _l + α _l) + qm _l = m _l + pα _l	$\underline{A1} = m_l + p\alpha_l - m_l - (p_l - q_l)\alpha_l = q_l\alpha_l$
A2	p A ₁ A ₂ + q A ₂ A ₂	pm _l + q(m _l - α _l) = m _l - qα _l	$\underline{A2} = m_l - q\alpha_l - (p_l - q_l)\alpha_l = -p_l\alpha_l$

The parameter α_l is the mean allelic substitution effect: $\underline{A1} - \underline{A2} = \alpha_l$

NB Bost shows that there is a constraint on allelic substitution effect: $|\alpha_l| < m_l$

NB allelic effects depend of the allelic frequencies.

Table A1.3 Definition of Breeding values at locus l

Genotype	Frequency	Mean effect of the alleles of one parent
A1A1	p ²	$2\underline{A1} = 2q_l\alpha_l$
A1A2	2pq	$\underline{A1} + \underline{A2} = \alpha_l(q_l - p_l)$
A2A2	q ²	$2\underline{A2} = -2p_l\alpha_l$
Mean		$p_l^2(2q_l\alpha_l) + 2p_lq_l\alpha_l(q_l - p_l) - 2q_l^2p_l\alpha_l = 0$
Additive variance V _{A1}		$4p_l^2q_l^2\alpha_l^2 + 2pq(q-p)^2\alpha_l^2 + 4q_l^2p_l^2\alpha_l^2$ $= 4p^2q^2\alpha^2 + 2pq^3\alpha^2 + 2p^3q\alpha^2 - 4p^2q^2\alpha^2 + 4q^2p^2\alpha^2$ $= 2p_lq_l\alpha_l^2$

Additive variance = variance of breeding values (variance of a centered-scaled variable = expectation of its squared value)

For n independent loci, the total additive variance is given as the sum of the contributions at each locus:

$$V_A = \sum V_{A1} \quad (\text{eq A1.2})$$

NB Additive variance is scaled by the allelic effects

$$\alpha_l^2 = V_{A1} / 2p_lq_l$$

Total phenotypic variation and heritability

The total phenotypic variation for each parametric trait is classically defined as:

$$V_P = V_A + V_D + V_E \quad (\text{eq A1.3})$$

where V_D is the dominance variance, and V_E the environmental variance. In Capsis, we assume that V_D = 0

The heritability for each parametric trait is classically defined as :

$$h^2 = V_A / V_P = 1 \quad (\text{eq A1.4})$$

2) Relation with the variables in Capsis GENETICS library and in PDG model

NB: in the following, the names of CAPSIS-GENETICS variables are italicized.

2.1. Individual genotypes:

In PDG, each individual tree has got its own genotype, which is an *IndividualGenotype*. An *IndividualGenotype* is composed of 3 tables (Figure A1.1).

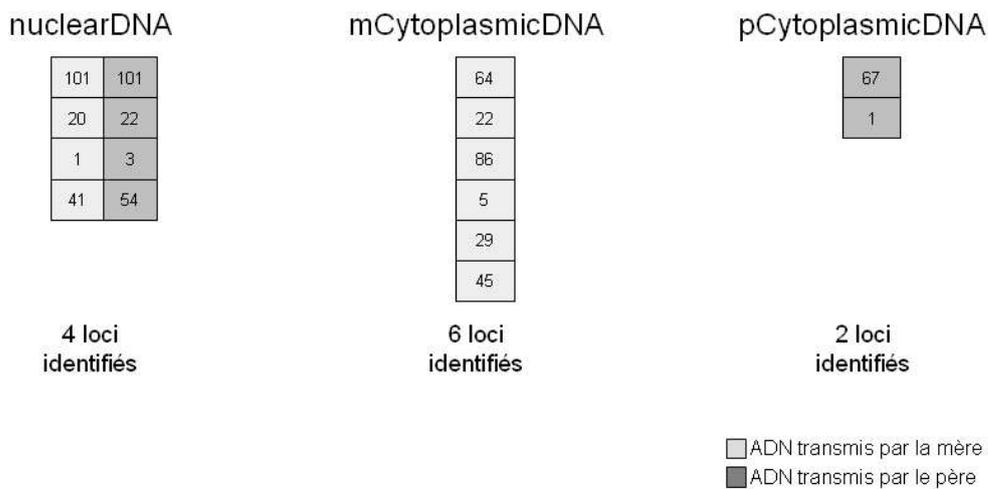
- The first table, named *nuclearDNA*, is two-dimensional and contains the genotype carried by nuclear DNA. This table consists of n lines (n corresponding to the number of loci) and 2 columns, the first for the mother's DNA and the second for the father's derived DNA. Each cell in the table contains an integer encoding the allele of the locus considered. In

PDG, $n=2$ for all the loci coding for parametric traits.

- The second and third arrays (named *mCytoplasmicDNA* and *pCytoplasmicDNA* respectively) are unidimensional and contain respectively (i) the genotype carried by cytoplasmic DNA transmitted by the mother, which is contained, depending on the species, in the mitochondria and / or in the chloroplast, and (ii) the genotype carried by cytoplasmic DNA transmitted by the father (mitochondrial and / or chloroplast depending on the species). These tables consist of n lines (n being the number of loci) and a column, each cell containing an integer encoding the allele at that location.

Accessors : *getGenotype ()*

Figure A1.1: Schematic representation of an Individual Genotype



For trees whose species is genetically considered, two of the three tables may be empty.

All trees of the same species must have the same genotype format, ie (i) the number of loci studied by DNA type (nuclear, maternal and paternal cytoplasmic) is identical for all trees of the same species and (ii) the loci are always in the same order.

For the trees of the initial stand, these genotypes have to be provided in the inventory file (see example in appendix 2). For trees generated during simulations in PDG, the genotype is computed by the GENETICS library, through meiosis and mating models.

2.2. Description of the population-level genetic composition

CAPSIS genetic library rely on two population-level variables describing the genetic composition of the population across generations:

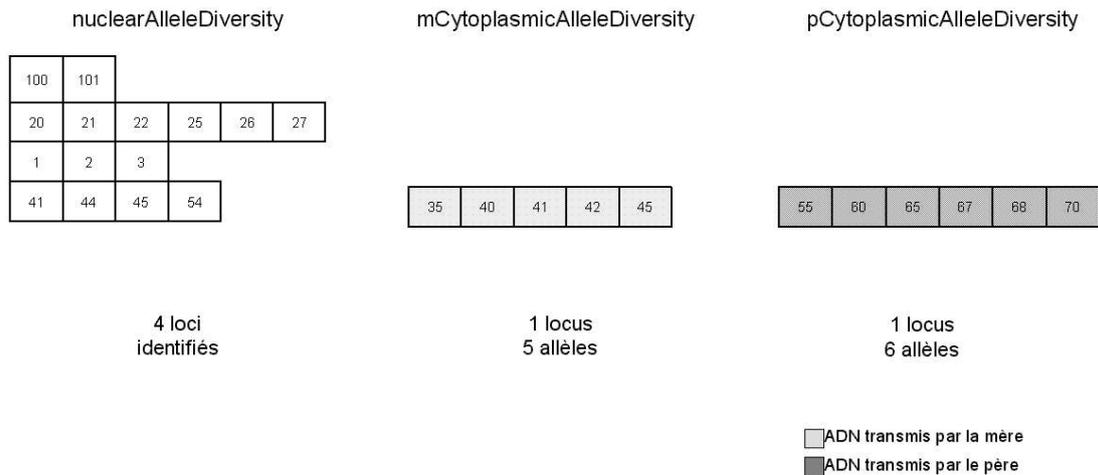
First, the variable *AlleleDiversity* contains three 2-dimensional arrays (Figure A1.2).

- The first table contains all possible alleles per locus of nuclear DNA. It consists of n lines, where n is the number of loci studied. Each row of the table contains a variable number of columns equal to the total number of potential alleles on the corresponding locus. Each cell in the array contains an integer, of type short, encoding the allele. This table is named *nuclearAlleleDiversity*.
- The second and third tables are constructed according to the same principle. They correspond respectively to the cytoplasmic DNA of the mother and the father. These arrays are respectively called *mCytoplasmicAlleleDiversity* and *pCytoplasmicAlleleDiversity*.

When the haplotypes are alleles of a single locus, these arrays have only one line.

Accessors : *getAlleleDiversity ()*, *getNuclearAlleleDiversity ()*,
getMCytoplasmicAlleleDiversity (), *getPCytoplasmicAlleleDiversity ()*.

Figure A1.2: Schematic representation of an AlleleDiversity



Second, the variable *GeneticMap* contains an array, called *recombinationProbas*, gathering probabilities of recombination between successive loci of nuclear DNA. It consists of $n-1$ lines, n being the number of loci studied on the nuclear DNA. Each line i contains a real, of double type, between 0 and 0.5, which represents the probability of recombination between the i th locus and the $(i + 1)$ th. The order of the recombination probabilities in *recombinationProbas* respects the order of the loci as given in the tree genotype.

Both *AlleleDiversity* and *GeneticMap* have to be defined in the inventory file (see Appendix 2 below).

2.3. Population-level allele effects and heritability:

The different quantitative genetic parameters detailed in section 1.2 above are stored in a map called *AlleleEffect* whose keys are the parametric traits of interest (in our case F_{CRITBB} and $g1max$). The map *AlleleEffect* contains as many sub-map as the number of parametric traits of interest (2 in our case). Each sub-map is named *ParameterEffect*, and contains three tables, and three parameters:

- The first table (called *nuclearAlleleEffect*) contains on lines the loci of nuclear DNA that influence the corresponding parameter. The rows have a number of columns that are variable and equal to the number of possible alleles on the locus considered, denoted by n , +1. In each row, the first value is equal to the position of the locus in the array of the corresponding possible alleles (*nuclearAlleleDiversity*), that is, its row number. The n following values are the allelic effects (the $\{\alpha_i\}$) of the n possible alleles of the locus. On each line, the sum of the n allelic effects of the alleles is equal to 0. In PDG, we assume that the loci coding for the studied traits are carried by nuclear DNA, so only *nuclearAlleleEffect* is not empty.
- The second and third tables, respectively *mCytoplasmicAlleleEffect* and *pCytoplasmicAlleleEffect*, are constructed on the same principle and contain the loci of the cytoplasmic DNA, respectively maternal and paternal, which influence the parameter considered. In PDG, we assume that the loci coding for the studied traits are carried by nuclear DNA, so *mCytoplasmicAlleleEffect* and *pCytoplasmicAlleleEffect* are empty
- The first parameter (real value of type “double”) corresponds to the theoretical heritability of the parameter; this variable is called *heritability*. The second parameter corresponds to the total environmental variance, and is called *totalEnvironmentalVariance*. The third parameter corresponds to the proportion of the interStep-environmental variance in the total

environmental variance of the parameter, and called *interEnvironmentalVariance*. These two last parameter aim at distinguishing a fixed source of environmental variation from a temporally-variable source of variation. In PDG, we assume most often that $interEnvironmentalVariance = 0$, so that $totalEnvironmentalVariance = fixedEnvironmentalVariance$

Accessors : *getAlleleEffect ()*, *getParameterEffect ()*, *getNuclearAlleleEffect ()*, *getMCytoplasmicAlleleEffect ()*, *getPCytoplasmicAlleleEffect ()*, *getHeritability ()*, *getTotalEnvironmentalVariance ()* et *getInterEnvironmentalVariance ()*.

These variables must be defined in the inventory file (see example below in Appendix 2). Moreover, a main constraint of GENETICS library is that the allelic effects (the $\{\alpha_i\}$) must be integer values. However, they are scaled by the value of the parametric trait, and the user may wish to have a finer precision than that of the parametric trait unit in the variation of allelic effect. Therefore, we introduced in PDG a parameter called *AlleleEffectMultiplCoeff* (typically set to 1000), which allow (i) to define in the inventory file the variable *nuclearAlleleEffect* with the required precision and (ii) to compute the genetic and phenotypic value of each individuals with the right scale of trait variation.

2.4. Individual genetic, environmental and phenotypic values

Moreover, each individual tree has got its own genetic, environmental and phenotypic values. These values are computed in PDG through the GENETICS library, based on the *individualGenotypes* and the *AlleleEffect* variables. They do not have to be provided, even for the initial stand. They are stored in maps of values, whose keys are the parametric traits of interest (in our case F_{CRITBB} and g_{lmax}).

- The map of an individual's genetic values contains the genetic (or genotypic) values of the different quantitative parameters studied. By definition, these values being invariable in time, the Map of the genetic values is also invariable in time. This variable is named *genoValue*.
- The map of an individual's environmental values contains the fixed environmental values (representing the constant environmental effects) of the various parameters studied. The Environmental Values Map has been defined to allow the safeguarding of fixed environmental values as soon as they are calculated, these values being used for the calculation of the phenotypic values. This variable is named *fixedEnvironmentalValue*.
- The map of the phenotypic values of an individual contains the phenotypic values of the different parameters at a given step of the evolution of the stand. The phenotypic value of a parameter is equal to the sum of the genetic value and the total environmental value (fixed part + variable part). Since an individual's phenotypic values are variable from Step to Step (since part of the environmental value varies from Step to Step), for a given individual, a Map of the phenotypic values can be calculated at each Step. This variable is named *phenoValue*.

Table A1.4 Summary of variables related to quantitative genetic parameter of interest and defined in CAPSIS

Variable name in CAPSIS	Notation	Comment
<i>nuclearAlleleEffect</i>	$\{\alpha_l\}$, table A1.2	NuclearAlleleEffect = $\alpha_l \times$ AlleleEffectMultiplCoeff (because AlleleEffect must contain <u>integer values</u>) To be defined in the inventory file
<i>totalEnvironmentalVariance</i>	V_E eq A1.3	To be defined in the inventory file
<i>interEnvironmentalVariance</i>	V_{E-IS}	To be defined in the inventory file
<i>additiveVariance</i>	V_A eq A1.2	Not required in the inventory file
<i>heritability</i>	h^2 eq A1.4	To be defined in the inventory file
<i>individualGenotypes</i>		To be defined for the initial trees in the inventory file
<i>genoValue</i>	A eq A1.1	Computed from <i>individualGenotypes</i> + <i>AlleleEffect</i>
<i>fixedEnvironmentalValue</i>		Computed from <i>AlleleEffect</i>
<i>phenoValue</i>		Computed from <i>genoValue</i> + <i>fixedEnvironmentalValue</i>

3) Parametrization of genotypes, allelic effects and heritability in PDG inventory file

General hypothesis: We generally assume that the influence of environment on the variation at the parametric traits F_{critBB} and g_{lmax} is negligible, so that $V_E = 0$; $V_P = V_A$ and $h^2 = 1$. By contrast, the variation in the traits of interest themselves (TBB and WUE) is shaped both by genetic variation (through F_{critBB} or g_{lmax}) and environmental variation (through temperature or water stress).

However, for control simulations, F_{critBB} and/or g_{lmax} can also be considered:

- either as non-variable ($V_E = 0$; $V_P = V_A = 0$; $h^2 = 0$ by convention)
- or as both environmentally and genetically variable ($0 \leq h^2 < 1$; $V_P = V_E + V_A$)

General objective: In the inventory file, we need to define the initial additive variance V_A , the allelic effects and the genotypes so that they match with the target distribution of each parameter at initiation. Moreover, we want to have an L-shaped distribution of the QTL effects.

Modeling variation at F_{critBB} in *Fagus sylvatica* in the typical case where $h^2_{F_{critBB}} = 1$.

Assuming that $h^2_{F_{critBB}} = 1$, we consider that the value of F_{critBB} is determined by $Nloc_{F_{critBB}}$ bi-allelic and independent loci, with purely additive effects (in this example $Nloc_{F_{critBB}} = 10$).

In the initial population, we impose two constraints on the mean ($\mu_{F_{critBB}}$) and the coefficient of variation ($CV_{F_{critBB}}$) of F_{critBB} :

- $\mu_{F_{critBB}} = \mu_{F_{critBBtarget}}$ (for instance 190 ° C , the mean value of F_{critBB} known for *Fagus sylvatica*)
- $CV_{F_{critBB}} = CV_{F_{critBBtarget}}$ (for instance 10%)

These constraints impose a target value for the standard deviation in F_{critBB} distribution:

$$\sigma_{F_{critBB}} = CV_{F_{critBBtarget}} \times \mu_{F_{critBBtarget}} = 19 \text{ in our example.}$$

Assuming that all loci contribute equally to the target additive variance $V_A = \sigma^2 = 361$, the mean genetic value for locus l is : $VA_l = V_A / Nloc_{F_{critBB}} = 36.1$ (at each of the $Nloc_{F_{critBB}}$ loci). Moreover, the value of α_l corresponding to VA_l depends on the allelic frequencies (see Table A1.5).

Table A1.5 : Possible values of allelic effects corresponding to a single target VA (361) but with different allelic frequencies.

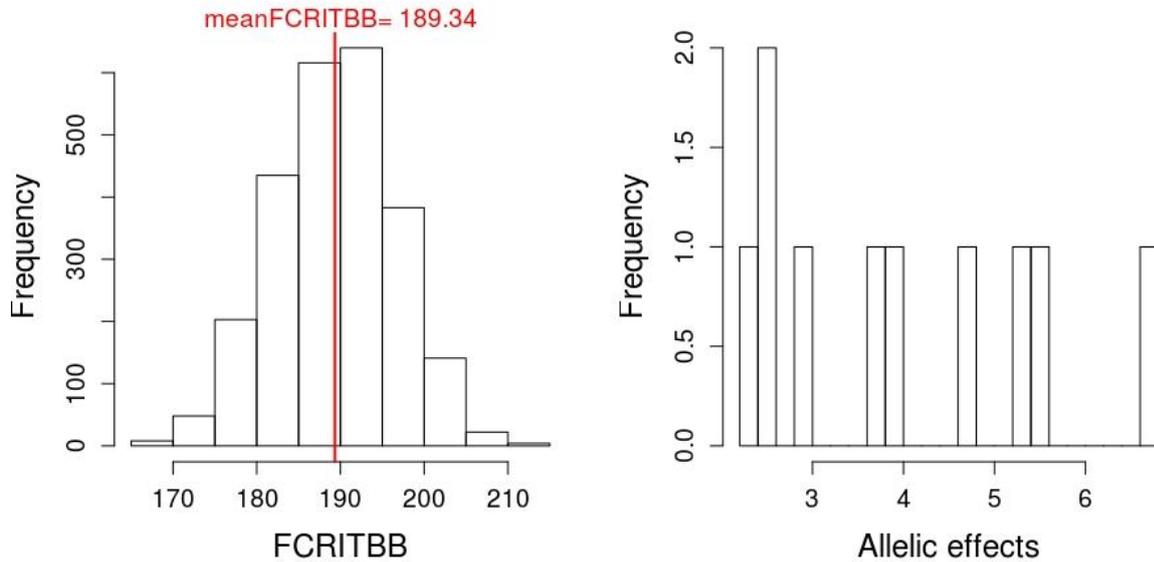
p_1	q_1	α_1^2	α_1	V_{A1}	V_A
0.5	0.5	72.20	8.50	36.1	361
0.6	0.4	75.21	8.67	36.1	361
0.7	0.3	85.95	9.27	36.1	361
0.8	0.2	112.81	10.62	36.1	361
0.9	0.1	200.56	14.16	36.1	361
0.95	0.05	380.00	19.49	36.1	361

To build the inventory file, we simultaneously sample the allelic frequencies (in a Dirichlet law), the individual genotypes (from the frequencies) and the allelic effects α_1 until the resulting V_A converge to its target value. This is done through an iterative procedure implemented in a R script provided in:

data/physiodemogenetics/generateInventoryFile/generateInvFile3.R

To run this code, the user has to provide the two constraining values $\mu_{F_{critBB}target}$ (called target_mean_FCRITBB) and $CV_{F_{critBB}target}$ (called target_CV_FCRITBB), as well as $Nloc_{F_{critBB}}$ and $h^2_{F_{critBB}}$ (called herit_FCRITBB).

Figure A1.3: Example of distribution of F_{critBB} and allelic effects for $\mu_{F_{critBB}target} = 190$ $CV_{F_{critBB}target} = 0.1$, $Nloc_{F_{critBB}} = 10$ and $h^2_{F_{critBB}} = 1$.



Modeling variation at g_{lmax} in *Fagus sylvatica*

We follow the same strategy for g_{lmax} . Assuming that $h^2_{g_{lmax}} = 1$, we consider that the value of g_{lmax} is determined by $Nloc_{g_{lmax}}$ bi-allelic and independent loci, with purely additive effects (here $Nloc_{g_{lmax}} = 10$).

In the initial population, we impose two constraints on the mean and CV of g_{lmax} :

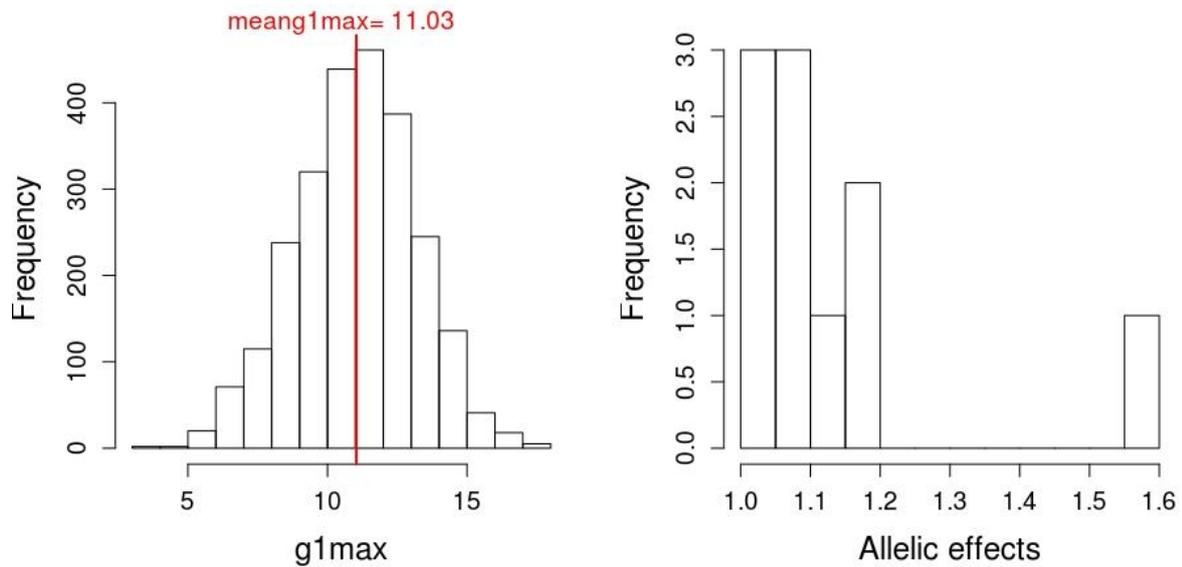
- $\mu_{g_{lmax}} = \mu_{g_{lmax}target}$ (for example 11, the mean value of g_{lmax} known for *Fagus sylvatica*)
- $CV_{g_{lmax}} = CV_{g_{lmax}target}$ (for example 20%)

These constraints impose a target value for the standard deviation in g_{lmax} distribution:

$$\sigma_{g1max} = CV_{g1max-target} \times \mu_{g1max-target} = 2.2$$

Assuming that all loci contribute equally to the target additive variance $V_A = \sigma^2 = 4.84$, the mean genetic value for locus l is : $VA_1 = V_A / Nloc_{g1max} = 0.484$ (at each of the $Nloc_{g1max}$ loci).

Figure A1.4: Example of distribution of $g1max$ and allelic effects for $\mu_{G1max-target} = 11$
 $CV_{G1max-target} = 0.2$, $Nloc_{G1max} = 10$ and $h^2_{G1max} = 1$.



NB : Allelic effects do not change dynamically across generations in simulation with PDG, while individual genotypes, and thus allelic frequency, dynamically change.

Generating null inventory with no variation or environmental variation in F_{critBB} and/or $g1max$

Two extreme cases can also be considered:

- 1) No variation (for instance at F_{critBB}): this is done by setting $herit_h^2_{FcritBB} = 0$ and $CV_{FcritBB-target} = 0$; we should expect $Nloc_{FcritBB} = 0$ in this case, but if the user wishes to consider a set of additional $Nloc_{FcritBB}$ SNP loci, genotypes at these loci can be generated as for other neutral SNP loci. All allelic effects are null. The value of F_{critBB} is set to $\mu_{FcritBB-target}$ for all individuals

- 2) Full environmental variation (for instance at F_{critBB}): this is done by setting $h^2_{FcritBB} = 0$ and $CV_{FcritBBtarget} > 0$ (for instance 0.1) ; we should expect $Nloc_{FcritBB} = 0$ in this case, but if the user wishes to consider a set of additional $Nloc_{FcritBB}$ SNP loci, genotypes at these loci are generated as for other neutral SNP loci. All allelic effects are null. The value of F_{critBB} is drawn in a Gaussian law of variance: $(CV_{FcritBBtarget} \times \mu_{FcritBBtarget})^2$

All other intermediate cases with $0 < h^2_{FcritBB} < 1$ can also be considered, though they were not tested.

Appendix 2 : inventory, climate and parameter files

1) Inventory file

An example of inventory file containing adults for a simulation in even-aged context is given in data/physiodemogenetics/invent100trees.inv. This file is commented below

1.1. General parameters

The inventory file begins with a list of “#General parameters”, divided in 4 categories

#SIMULATION parameters

parameter	Value	Comment	Nature
year	0	Initial year from which simulations begin	Integer
selectionOnSeedlings	false	A flag indicating if mortality of seedlings is driven purely by density-dependance (false) or by selection (true)	Boolean
simulAfterPreviousSimulations	false	A flag indicating if the inventory file is the generation 0 (false) or if the inventory file is the result of previous simulations with PDG from generation 0 to generation n	Boolean
initGeneration	5	if simulAfterPreviousSimulations = true, a value corresponding to the generation of the initial stand	Integer
preEvolutionTime	0	nb of generations to bring pop to migration-drift equilibrium ¹	Integer
seedProductionIsEquals	false	A flag indicating whether individuals produce different number of seeds (asd function of their reserves) or if they all produce the same number of seeds (artificially set to the mean seed production)	Boolean
simulationType	1	0= gradient-shaped plot ; 1 = flat plot for test	Integer

¹Here, we use inventory files where genotypes have been generated while accounting for migration-drift-selection equilibrium

#PLOT parameters

parameter	Value	Comment	Nature
origin	(0 , 0)	coordinates of the first cell of the plot (top left)	Integer
nLin	5	number of cells on the y axis	Integer
nCol	5	number of cells on the x axis	Integer
cellWidth	50	square cell dimensions	Double, m
minimalAltitude	1200	minimal altitude of the plot	Double, m
slope_deg	0	45= gradient-shaped plot ; 0= flat plot for test	degree
longitude	-5.14	Plot longitude	Double
latitude	44.1	Plot longitude	Double

#DEMOGRAPHIC parameters

parameter	Value	Comment	Nature
deltaSeed	18.3	The mean distance of seed dispersal (m)	Double, m
bSeed	1	The shape of the of seed dispersal kernel (dimensionless)	Double
pollenKernel	true	A flag indicating if pollen dispersal is modeled following a diispersal kernel (true) or if mating is assumed to be panmictic within a given radius	Boolean
deltaPollen	100	If pollenKernel=true, the mean distance of pollen dispersal (m)	Double, m
bPollen	0.3	If pollenKernel=true, the shape of the of pollen dispersal kernel (dimensionless)	Double
maxPollinationDistance	50	If pollenKernel=false, the radius with which mating occurs at random	Double, m
selfing	true	A Boolean specifying if selfing occurs (true) or no	Boolean
selfingRate	0.05	The selfing rate (%)	Double
probaLDseedFlow	0.99	Probability of long-distance seed flow, not yet used	Double
probaLDPollenFlow	0.95	Probability of long-distance seed flow- not yet used	Double
adultAge	40	the age at which trees become adult	Integer
recruitmentAge	10	the age at which seedlings are recruited	Integer
dbh_recruitedSeedlings_mean	5	the mean diameter of seedlings at recruitment	Double, cm
dbh_recruitedSeedlings_sd	1	the standard deviation of seedlings diameter at recruitment	Double, cm
heigth_recruitedSeedlings_mean	5	the mean heigth of seedlings at recruitment	Double, m
heigth_recruitedSeedlings_sd	1	the standard deviation of seedlings heigth at recruitment	Double, m
dbh_SeedlingsToAdult_mean	13.8	the mean diameter of adults at seedlingsToAdult transition	Double, cm
dbh_SeedlingsToAdult_sd	0.9	the standard deviation of adults diamter at seedlingsToAdult transition	Double, cm
height_SeedlingsToAdult_mean	9	the mean heigth of adults at seedlingsToAdult transition	Double, m
height_SeedlingsToAdult_sd	0.34	the standard deviation of adults heigth at seedlingsToAdult transition	Double, m
rateGermination	0.485	Rate of seed germination	Double
rateEmptySeed	0.33	Rate of empty seeds	Double
rateSurvival	0.25	Rate of seed survival	Double
maxCanopyRadius	5	The maximum radius of an adult tree canopy;used only for reevolution, then dynamically updated after tree growth	Double, m
maxOverlappingPercentage	0.01	The maximum overlap (between 0 and 1) between two neighbor trees canopies	Double
nbAdultPerCell	4	Maximum number of adult per square cell	Integer
maxGrowingNumberofSeedlings	5	Maximum number of seedlings per square cell for which growth is simulated	Integer

#GENETIC parameters

parameter	Value	Comment	Nature
numberOfGeneticParameters	2	the number of parameters of CASTANEA that are genetically variable (the number of lines in the allele effect record below); should be equal to 2 (FCRITBB and g1)	Integer
numberOfTraits	5	the number of traits that will be exported ; includes Bbday and WUE, as well as ETR, GPP and seedProduction	Integer
nbLocus_FCRITBB	10	number of SNP locus controlling FCRITBB	Integer
nbLocus_g1max	10	number of SNP locus controlling g1max	Integer
nbLocusMicrosat	0	number of microsatellites	Integer
nbLocusNeutralSNP	10	number of neutral SNP	Integer
distanceClassBoundsForSGS	{10; 25; 50; 75;100}	a set of distances classes that will be used to compute spatial genetic structure	vector of double
meanInitialGValue_FCRITBB	189.062559	$\mu_{FCritBB}$ in the initial stand, see Appendix 1 paragraph 3	Double
sdInitialGValue_FCRITBB	19.11	$\sigma_{FCritBB}$ in the initial stand, see Appendix 1 paragraph 3	Double
meanInitialGValue_g1max	11.0482081	μ_{g1max} see Appendix 1 paragraph 3	Double
sdInitialGValue_g1max	2.19	σ_{g1max} in the initial stand, see Appendix 1 paragraph 3	Double
meanTargetValue_FCRITBB	190	Target value of $\mu_{FCritBB}$ used to draw trees of the initial stand	Double
targetedCV_FCRITBB	0.1	Target coefficient of variation of FCRITBB used to draw trees of the initial stand	Double
meanTargetValue_g1max	11	Target value of μ_{g1max} used to draw trees of the initial stand	Double
targetedCV_g1max	0.2	Target coefficient of variation of g1max used to draw trees of the initial stand	Double
alleleEffectMultiplCoeffFCRITBB	1000	AlleleEffectMultiplCoeff, see Appendix 1 paragraph 2.2	Integer
alleleEffectMultiplCoeffg1max	1000	AlleleEffectMultiplCoeff, see Appendix 1 paragraph 2.2	Integer

#CASTANEA parameters

parameter	Value	Comment	Nature
CO2mode	CO2_PAST_EVOLUTION	Indicates how CO ₂ concentration varies among years. Possible values of CO2mode : CO2_PAST_EVOLUTION (fixed value), CO2_RCP2_6 (variable following climatic scenario RCP2.6), CO2_RCP4_5, CO2_RCP6, CO2_RCP8_5	Prescribed string value
Ca	380	if CO2mode = CO2_PAST_EVOLUTION, the fixed value for CO ₂ concentration	Integer, ppm
frostEffectCoefficient	0.1	The % of bud destroyed by frost	Double
iFROST	1	The way frost effect are accounted for ² 1= frost effects on bud + LAI are accounted for but no reflusing	Integer
LAI mode	LAI_NSC_AND_DBH	Possible ways for LAI to vary : " LAI_FIXED " = no variation in LAI; " LAI_NSC " = LAI growth depends on reserves; " LAI_NSC_AND_DBH " = LAI growth depends on reserves and DBH; " LAI_NSC_DBH_AND_RDI " = LAI growth depends on reserves , DBH and RDI	Prescribed string value
fixedTronvivi	false	A flag indicating if the proportion of alive cell is fixed across tree lifetime ("true") or not ("false")	Boolean
phenoMode	PHENO_CASTANEA	Possible phenology models : "PHENO_CASTANEA" is the one described in this documentation; other mode are imported from Pheofit or PHELIB library : "PHENO_FITLIB_UNIFORC"; "PHENO_FITLIB_UNICHILL"; "PHENO_FITLIB_GAUZERE"; "PHENO_PHELIB";	Prescribed integer value
stomataStress	STOMATA_STRESS_GRANIER	Possible ways stomatal stress is simulated; with " STOMATA_STRESS_GRANIER ", stomataStress depends on water potential; with " STOMATA_STRESS_RAMBAL ", depends on soil water content "	Prescribed string value
vcmaxStress	false	A flag indicating whether water stress affect Vcmax	Boolean
eps	1.00E-05	a value for epsilon	Double
T_oldinit	0	first temperature of the previous day at the end of the diurnal period	Double
iKarst	0	0 = no water reserve in deep soil; 1= reserve in deep karstic soil	Prescribed integer value
rootsCavitation	false	A flag indicating whether roots can die through cavitation	Boolean
temperatureEffectOnPhotosynthesis	TEMPERATURE_EFFECT_BERNACCHI	Possible ways for temperature to affect photosynthesis Possible values = "TEMPERATURE_EFFECT_BERNACCHI"; "TEMPERATURE_EFFECT_ARRHENIUS"	Prescribed string value

- three values corresponding to *heritability*, *totalEnvironmentalVariance* and *interEnvironmentalVariance* (see Appendix 1 section 2.3)

```
#Phi
# speciesPhi DefaultPhi
1 {} 0.
```

The third sub-section is mandatory and defines the individual consanguinity of trees of the initial stand.

1.3. Plot and cells definition

The next section gives the list of cells with their properties

```
## CELLS
#CID clign ccol solHeight stoneContent wfc wilt propMacro bulk SOLCLAYtop SOLCLAYsol SOLFINtop SOLFINsol SOLSANDtop SOLSANDsol deepSoilDepth stoneContentDeep prac pracDeep
1 0 0 1059.0984742586 0.2 0.39 0.25 0 1.07 0.5045 0.4354 0.91 0.8164 0.0895 0.18 0 0 0.6 0
2 0 1 961.043675711529 0.2 0.39 0.25 0 1.07 0.5045 0.4354 0.91 0.8164 0.0895 0.18 0 0 0.6 0
3 0 2 1029.07127262217 0.2 0.39 0.25 0 1.07 0.5045 0.4354 0.91 0.8164 0.0895 0.18 0 0 0.6 0
4 0 3 523.161460070906 0.2 0.39 0.25 0 1.07 0.5045 0.4354 0.91 0.8164 0.0895 0.18 0 0 0.6 0
5 0 4 987.650518150082 0.2 0.39 0.25 0 1.07 0.5045 0.4354 0.91 0.8164 0.0895 0.18 0 0 0.6 0
...
96 19 0 953.884438455166 0.2 0.39 0.25 0 1.07 0.5045 0.4354 0.91 0.8164 0.0895 0.18 0 0 0.6 0
97 19 1 1009.41635080784 0.2 0.39 0.25 0 1.07 0.5045 0.4354 0.91 0.8164 0.0895 0.18 0 0 0.6 0
98 19 2 632.453189281614 0.2 0.39 0.25 0 1.07 0.5045 0.4354 0.91 0.8164 0.0895 0.18 0 0 0.6 0
99 19 3 867.939968674302 0.2 0.39 0.25 0 1.07 0.5045 0.4354 0.91 0.8164 0.0895 0.18 0 0 0.6 0
100 19 4 1054.08065815679 0.2 0.39 0.25 0 1.07 0.5045 0.4354 0.91 0.8164 0.0895 0.18 0 0 0.6 0
```

There is one line per cell, and each line (cell) contains:

- An integer corresponding to the cell ID (cID)
- Two integer corresponding to the coordinates on the y-axis (clign) and on the x-axis (ccol). Note that the size of each cell in m is given in the general parameter above (in our case, each cell measures 50 m, and thus correspond to 0.25 ha). The total plot is thus 25 ha in area. Moreover, note that together with the parameters minimalAltitude and slopeInDegree (given in the general section above), these parameters define the altitude of each cell (see Figure A2.1 below for a plan of the simulated plot)
- Fifteen double values describing, soil properties and listed in the table below

Parameter	Unit	Definition
solHeight	mm	soil height
stoneContent	Proportion	Proportion of stones
wfc		
wilt		
propMacro		
bulk		
SOLCLAYtop	Proportion	Prop of Clay between 0 and 30 cm
SOLCLAYsol	Proportion	Prop of Clay between 30 and 100 cm
SOLFINtop	Proportion	Prop of fine particles bet. 0 and 30 cm
SOLFINsol	Proportion	Prop of fine particles bet. 30 and 100 cm
SOLSANDtop	Proportion	Prop of sand between 0 and 30 cm
SOLSANDsol	Proportion	Prop of sand between 30 and 100 cm
deepSoilDepth	cm	Depth of deep soil (for karstic area)
stoneContentDeep	Proportion	Proportion of stones in deep soil
prac		
pracDeep		

The soil Water Holding Capacity (WHC, in mm) is variable among cells and computed as:

$$WHC = (1 - \text{stoneContent}) \times \text{solHeight} \times (\text{wfc} - \text{wilt}) \times Da$$

where $(\text{wfc} - \text{wilt}) = 0.14$ and $Da = 1.07$

In the inventory, the variability in WHC is artificially simulated by varying solHeight only (see Figure A2.2)

Figure A2. 1: Simulated plot in the example inventory file. The grey grid corresponds to the 100 cells, with their cID indicated. In blue are given the coordinate of the center of the cell on the x-axis (bottom), on the y-axis (left side) and on the z-axis (right side).

y (center cell)						z (center cell)
975	1	2	3	4	5	1675
925	6	7	8	9	10	1625
875	11	12	13	14	15	1575
825	16	17	18	19	20	1525
775	21	22	23	24	25	1475
725	26	27	28	29	30	1425
675	31	32	33	34	35	1375
625	36	37	38	39	40	1325
575	41	42	43	44	45	1275
525	46	47	48	49	50	1225
475	51	52	53	54	55	1175
425	56	57	58	59	60	1125
375	61	62	63	64	65	1075
325	66	67	68	69	70	1025
275	71	72	73	74	75	975
225	76	77	78	79	80	925
175	81	82	83	84	85	875
125	86	87	88	89	90	825
75	91	92	93	94	95	775
25	96	97	98	99	100	725
x (center cell)	25	75	125	175	225	

- pID: the ID of the father of the tree (here -1 indicate that this ID is unknown)

1.5. R script to generate inventory file

The R script generateInventoryFile_v3.R is available in the directory /data/physiodemogenetics/ to generate inventory files with the appropriate format.

Basically, the script includes three parts:

STEP1 = loading of parameters; all the parameters value are defined here

STEP 2 = computations. Allelic effects for each trait are computed (optimization loop constrained by the target mean and standard variation of each trait) and individual's genotypes are drawn.

STEP 3 = writing of the inventory file in the appropriate format

2) Climate file

An example of climate file containing adults for a simulation in even-aged context is given in data/physiodemogenetics/climate/Example100indiv.inv. This file is commented below.

#year	month	day	RG	HR	V2	RR	TX	TN	Tmoy
1960	1	1	4.479	78.2	2.8	0.2	7.5	0.8	4.15
1960	1	2	1.913	79	2.5	3.1	5.3	2.7	4
1960	1	3	2.468	89.9	2	0.1	8.1	2.6	5.35
1960	1	4	7.623	70.9	4.1	0	7.7	3.4	5.55
1960	1	5	7.247	83	2.3	0	6.6	0	3.3
1960	1	6	6.987	74.2	2.7	0	6.1	3.7	4.9
1960	1	7	7.922	76.4	3.1	0	3.2	-2.3	0.45
1960	1	8	7.012	72.1	2.7	0	4.9	-2.4	1.25
1960	1	9	8.893	83.6	4.3	0	0.4	-1.8	-0.7
...									
2011	12	18	8.28	89	3.92	0.2	-0.5	-2.7	-1.6
2011	12	19	7.9	100	3.15	3	-0.2	-3.9	-2.05
2011	12	20	3.75	99	2.59	3.2	1.7	-5	-1.65
2011	12	21	5.24	100	2.87	0	7	0.8	3.9
2011	12	22	7.4	100	3.99	0	10.1	5.2	7.65
2011	12	23	7.12	100	1.47	0.1	13	2.1	7.55
2011	12	24	8.96	86	5.53	0.1	2	1	1.5
2011	12	25	6.78	78	1.54	0	5.6	-2.1	1.75
2011	12	26	7.95	63	0.7	0	9.1	-5.8	1.65
2011	12	27	7.9	100	0.7	0	10.3	-3.9	3.2
2011	12	28	8.79	97	0.63	0	10.1	-2.8	3.65
2011	12	29	7.7	95	4.69	0	3.3	0.5	1.9
2011	12	30	7.69	100	4.69	2.4	6.2	-0.3	2.95
2011	12	31	6.86	80	3.71	0.4	9.3	1.8	5.55

This file contains 53 years *12 months *28-31 days = 8993 lines. Each line contains:

- Year,month,day : the date
- RG: the global radiation
- HR: the relative humidity
- V2: Wind speed
- RR:
- TX
- TN
- Tmoy

3) Species parameters

Table to be provided to give the list of parameters in the file CastaneaSpecies3.txt.

Appendix 3: Output files

1) yearlyResults.csv

The Table below gives the code and signification of all the variables of the file yearlyResults.csv

Variable	Signification	Unit	type
ID	ID of the tree		
cell	ID of the square cell of the plot		
x	X		
y	Y		
altitude	Altitude		
gen	Generation		
pop	Population		
climaYear	Climatic year		
year	Yearly precipitation		
species	Species code		
LAI	Leaf Area Index	m/m ²	output variables
GPP	Gross Primary Production	gC/m ² /y	output variables
ETRveg	Evapotranspiration of canopy	mm/y	output variables
ETRsol	Evapotranspiration of soil	mm/y	output variables
TR	tree transpiration	mm/y	output variables
StressLevel	Water Stress integral	Mpa	output variables
gel	Number of late frost days		output variables
DBBV	wood growth	gC/m ² /y	output variables
DBF	leaf growth	gC/m ² /y	output variables
DBRF	fine roots growth	gC/m ² /y	output variables
Rtot	Autotrophic respiration	gC/m ² /y	output variables
Rcanopy	Canopy respiration	gC/m ² /y	output variables
BSSmin	Minimal NSC content	gC/m ²	output variables
endWood	Day of stopping wood growth	day	output variables
Delta13C			output variables
BF	Foliar Biomass	gC/m ²	output variables
reserves	NSC content	gC/m ²	output variables
roots	Fine roots Biomass	gC/m ²	output variables
BiomassOfTrunk	Trunk biomass	gC/m ²	output variables
AliveWoodBiomass	Alive wood biomass		output variables
BBday	Budburst day	julian day	output variables
endLeaf	leaf fall day	julian day	output variables
Tmoy	Average Daily Temperature	°C	input climate variables
Tmax	Maximal daily temperature	°C	input climate variables
Tmin	minimal daily temperature	°C	input climate variables
Rg	Global radiation	MJ/y	input climate variables
PRI	Yearly precipitation	mm/y	input climate variables

age	age		output variables
rw	ring width increment	mm/y	output variables
dbh	diameter à breast height	cm	output variables
height	tree height	m	output variables
seed	seed production		output variables
RU	Soil Water Capacity	mm	output variables
drainage	Soil Water Drainage	mm/y	output variables
ca	CO2 concentration	ppm	input atmospheric variables
CrownProjection	Tree Crwon projection	m2	output variables
PLeafmin	Minimal Leaf Water Potential	Mpa	output variables
DBHinit	Initial DBH	cm	input cell variables
TSUMBB	Sum of temperature required for budburst	°C	input cell variables
g1	slope of the Ball & al relationship		input cell variables
nc	slope of relationship between Vcmax and Nitrogen per leaf Area		input cell variables
nf	Nitrogen content		input cell variables
LMA	Leaf Mass Per Area		input cell variables
cofrac	Ratio between fine roots and leaf		input cell variables
GBVmin	initial allocation coefficient to wood		input cell variables
woodStop			input cell variables
CRBV	construction cost of wood		input cell variables
potsoiltowood	Soil water potentila at which wood growth stops	Mpa	input cell variables

Other results files to be improved and commented

Appendix4 : Some detailed examples

Dynamic and demography of the populations

nbAdultPerCell

The maximum final density of the stand after a “s” step is D trees/ha (where the parameter D is typically the density of adult trees)