TransPopRege Model Documentation

1. Aim & scope of the model

The module TranspopRege under CAPSIS4 platform aims at studying the effects of various ecological and genetical processes on the genetic quality of seedlings in contrasted regeneration context (Figure 1). Indeed, natural regeneration is a key stage in the life cycle of populations within ecosystems because the demographic structure and genetic make-up of the population is mostly determined at this stage, in particular for long lived organisms such as forest trees.

Figure 1: *Demo-genetic processes of tree life-cycle accounted for in TranspopRege*
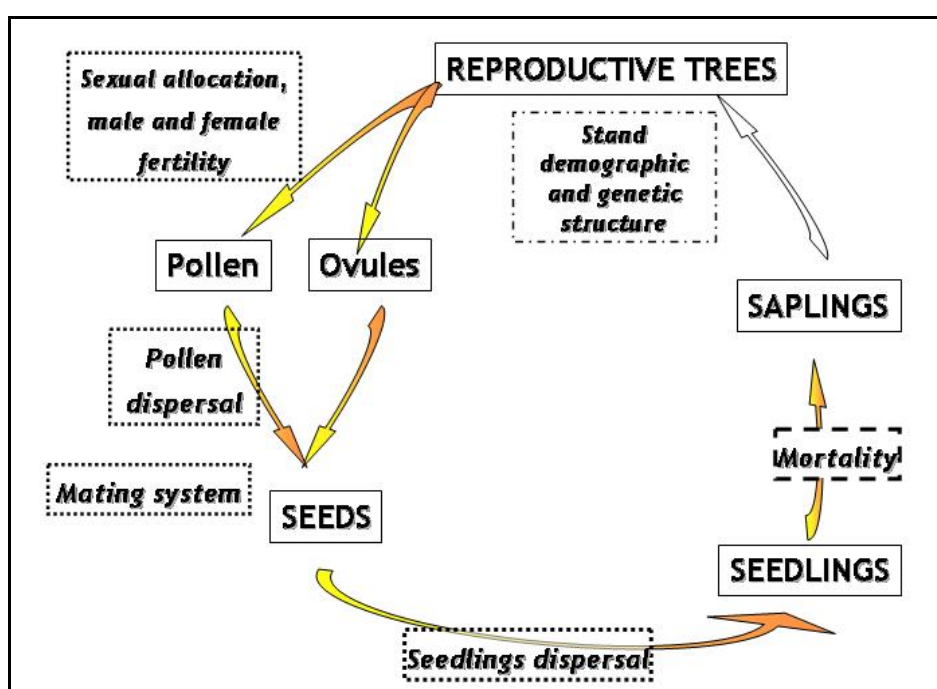
TranspopRege is an individual-based model similar to a gap model. Space is abstracted as a grid of 50x50 m square cells (or patch), with each patch considered as homogeneous in terms of site, species composition, density, age. Recruitment, growth, and mortality are modeled at the individual level. TranspopRege considers only one species, with hermaphroditic individuals. However, as male and female fecundity are described for each individual tree, different sexual determinism can be modeled (i.e. monoecy, dioecy). Each individual tree basically bears dendrometric information (dbh, height, age, location) and genetic information (neutral markers and quantitative traits).

Regeneration, i.e. the number and composition of new seedlings recruited in a particular patch is a function of power and location of both seed sources and pollen source, via dispersal functions across the scene (within and among patch component). Only the mortality of seedlings is accounted for, and modelled through a differential survival probability across patches (as

function of site index and consanguinity). Growth process is not accounted for in TranspopRege.

TranspopRege allows the regeneration process to be simulated as the result of adult tree spatial structure (density, aggregation), mating system (shape and range of the pollen and seed dispersal curves, annual/individual male and female fertility variations) and inbreeding depression (through higher mortality of selfed seedlings). TranspopRege is designed to perform easily sensitivity analyses of the model output (genetic diversity and structure of the seedlings), through a script mode implementation.

References: Oddou-Muratorio et al (2006) Ecological factors shaping the genetic quality of seeds and seedlings in forest trees: a simulation study coupled with sensitivity analyses. IUFRO meeting in Alcala Henares. To be published.

Time scale: TranspopRege operates at an annual time-scale during the regeneration phase of forest stand. TranspopRege is conceived to model high forest sylvicultural regime over few generations
Spatial scale : few management units (~100 ha). The size of the "scene" depends on the number of cells. We used typically 20×20 cells of 50m×50m, that is 100 ha. Regeneration can be simulated on a subsample of these cells.

We assume a high forest sylvicultural regime, and model only the regeneration phase (Figure 2):

Figure 2 : Time scale in TranspopRege
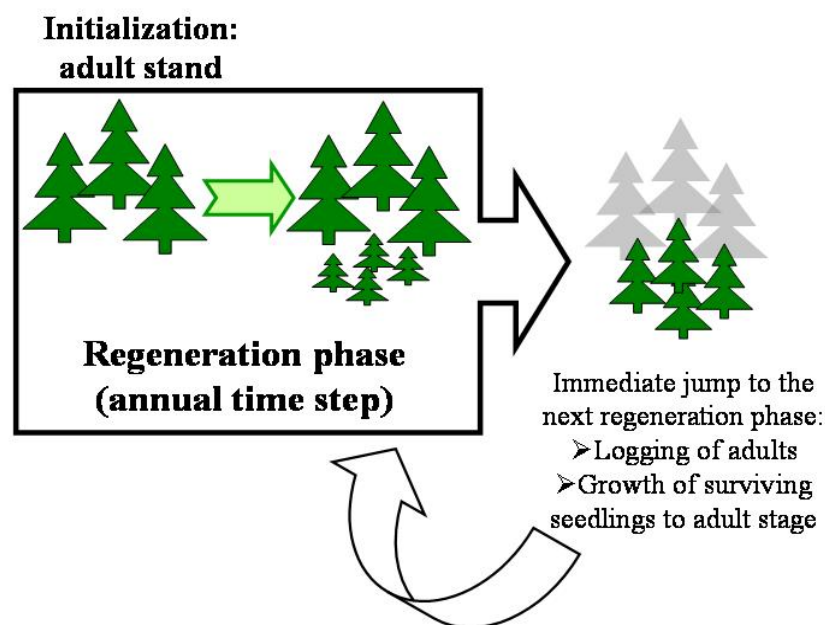
2. Genetic system

The genetic system is described using the "Genetics" library of Capsis4, which is available as a generic package for all the Capsis 4 models considering individual spatial trees. "Genetics" library manages genotypes and genetic processes (meiosis, fecundation). Considering the time scale of most models under Capsis4, mutation is not accounted for. "Genetics" library considers diploid species with sexual reproduction. Genotypes are modelled as several discreet loci, either neutral, or contributing to a trait under selection.

The loci are placed on a simulated genetip map. This includes the neutral loci, and the loci that contribute to the trait under selection. For that map, recombination rates between adjacent loci should given. Recombination rates range between 0 (full linkage) and 0.5 (full independence). A maximum of 256 alleles can be defined for each locus. The positions on the map have to be given (values available in the literature can be used), the default case being full independence between loci.

For all loci, the user has to specify the following values:
- allelic values (additive effects only)
- theoretical heritability for traits under selection (h^2)
- total environmental variance for the trait: $\sigma^2_{env\text{-}total}$
- proportion of variance inter-step* : $\sigma^2_{env\text{-}inter}$

Allelic values are used to compute genetic values under additive model. Other data are used to compute environmental and phenotypic values. Environmental values associated with a trait are assumed to be dissociable in two components: a within-step environmental variance, and an among-step environmental variance. For instance, if the considered trait is growth capacity, the environmental variance associated with this trait is the sum of :
- among-step environmental variance, corresponding to annual variability for instance in rain regime
- within-step environmental variance, corresponding to variability in local environment (such as micro- variability in soil), considered to be constant over time.

Environmental value is computed as the sum of :
A fixed environmental value, drawn in a Gaussian law of parameter $(0; \sigma^2_{env\text{-}intra})$, where $\sigma^2_{env\text{-}intra} = \sigma^2_{env\text{-}total} (1 - \sigma^2_{env\text{-}inter})$.
An among-step environmental variance, in a Gaussian law of parameter $(0; \sigma^2_{inter})$.

The user must supply a value for σ^2_{inter}, but can supply either a value for σ^2_{total} or for h^2. The model will then compute the missing parameters following the classical equation :
$$h^2 = \sigma^2_{gen} / (\sigma^2_{gen} + \sigma^2_{env\text{-}total})$$
where σ^2_{gen}, the genetic variance is computed using all the tree from the initial stand under panmixy hypothesis.

a. Initial allele frequency

There are two ways to specify initial allelic frequencies:
- Either the user specify the list of alleles at each locus and all their allelic frequencies in the inventory file
- Or the user give in the inventory file a list of trees that constitute the initial stand, with theur genotype. The model then computes the "allelic diversity" (list of alleles at each locus) and allelic frequencies.

b. Initial phenotypic values

There are two ways to specify initial allelic frequencies:
- Either the user specifies in the inventory file a list of trees that constitute the initial stand, with their phenotypic values.
- Or the user specifies a number of trees to be randomly generated. The distribution of phenotypic values can then be specified for each trait. Typically the diameter of each tree is drawn in a Gaussian distribution whose parameters are fixed by the user.
-

3. Gene flow

a. Pollen and seed dispersal kernel

As in Oddou-Muratorio et al. , pollen dispersal is modelled using a dispersal kernel $p_p(.;x,y)$ describing the probability for a pollen grain emitted at position $(0,0)$ to participate to the pollen cloud at position (x,y). Following this scheme, we model seed dispersal using a dispersal kernel $p_s(.;x,y)$ describing the probability for a seed emitted at position $(0,0)$ to establish as a seedlings at position (x,y).

We considered here the family of exponential power functions:

$$p(a,b;x,y)=\frac{b}{2\pi a^2 \Gamma(2/b)}\exp\left(-\left(\frac{\sqrt{x^2+y^2}}{a}\right)^b\right),$$

Eqn. 4.1

where Γ is the classically defined gamma function (see e.g. . The parameter b is the shape parameter affecting the tail of the dispersal function and a is a scale parameter homogeneous to a distance . When $b < 1$, the dispersal kernel is fat-tailed , i.e. the long-range decrease is slow (at least slower than an exponential of the distance, for which $b = 1$), increasing the probability of long-distance dispersal events. Conversely, when $b > 1$ (for instance the Gaussian model, for which $b = 2$) the dispersal is thin-tailed, with a rapid decrease of the dispersal function, implying much less long-distance dispersal events than when $b < 1$. When b>2, the dispersal kernel is platykurtic, meaning less long-distance dispersal events than the Gaussian kernel with the same variance. When b<2, the dispersal kernel is leptokurtic (see figure 3) .

The mean distance (δ) travelled by a pollen grain under the kernel $p(a, b;.)$

is given by:

$$\delta = a \frac{\Gamma(3/b)}{\Gamma(2/b)}$$

Eqn. 4.2

In practise, the dispersal parameters used as input parameters in the simulations are b, the shape parameter (respectively b_{pollen} and b_{seed}) and δ, the average distance of dispersal (respectively δ_{pollen} and δ_{seed}). These parameters can be calibrated using available estimates in literature (e.g. .

Figure 3: Example of curves for Gaussian(red line), exponential (orange line), exponential power (green line) with shape parameter b = 2, 1 and 0.5, and scale parameter (a) adjusted so that in all cases δ = 400 m.
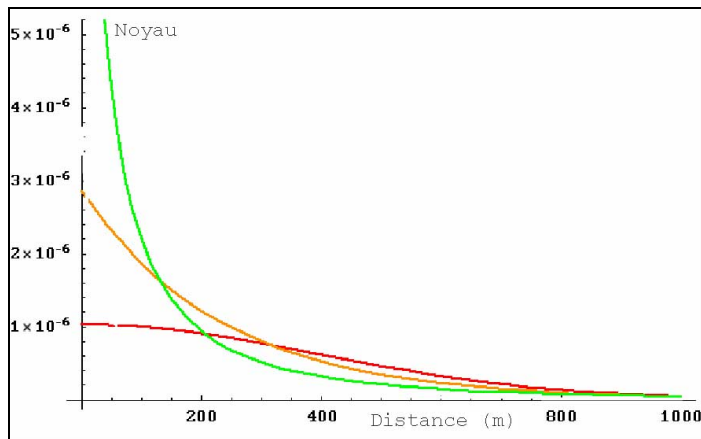
b. Pollen and seed dispersal within and among patches

Pollen dispersal is considered as panmictic within patch, with equal frequencies of each local male in the pollen cloud of local female ($p_{within-patch}$ = $p(a,b, d/2, d/2)$ where d is the length of square cell). Self incompatibility is not accounted for yet. For pollen dispersal from a male tree to a female tree belonging to another patch, the exact spatial positions of both male and female are accounted for.
For seed dispersal, the spatial component of the seed rain is accounted for within patch. For seed dispersal among patches, the average seed rain from a spatial female tree to the center of each square cell is computed, and then seedlings are distributed randomly within patch.

4. Male and female fecundity

a. Production of pollen

We assume that male fecundity depends primarily on tree dimensions (diameter D) and we model the selection gradient of tree diameter on pollen production through an exponential function (e.g. . Selection gradients are defined as the slope of the regression of relative fertility on trait values

measured in single plants. Linear slopes identify directional selection and quadratic components are associated with stabilizing (negative values) or disruptive (positive values) selection. Here, we model only directional selection. The relative fertility of a male k of diameter D_k is given by :

$$FM_k = e^{mFertD_k} \bigg/ \sum_{l \in M} e^{mFertDl}$$

Eqn. 4.3

where M is the collection of fertile male trees, and where mFert reflect the strength of the selection of diameter on male fertility.

b. Male reproductive success

We assume that the relative proportion of a given male-tree k in the pollen cloud of a female-tree j (denoted Π_{jk}) is determined by two independent factors: (1) relative male fertility (see eq (3)) and (2) the distance of mother j to male k and pollen dispersal parameters (see Eqn 4.1). Therefore, π_{jk} is expressed as:

$$\Pi_{jk} = MF_k \times p_p(a_p, b_p, d_{jk})$$

Eqn. 4.4

where D_j is male k diameter, a_p and b_p are the parameters of the pollen dispersal kernel p_p, and d_{jk} the distance of mother j to male k.

c. Production of seeds

We now consider absolute female fertility, i.e. the total number of viable seeds produced by a given mother tree. We model this female fertility as a function of mother tree diameter Dj, and we also account for stochastic inter-annual variability in seed production.

We used data on various trees species to adjust the following general predictive model of the expected number of seeds produced by a female tree j of diameter Dj :

$$FF_j = \alpha \times (\Pi D_j - 100)^{fFert} - (A \times \alpha + B)$$

Eqn. 4.5

Where α the interannual effect on seed production where fFert is similar to a selection gradient of tree diameter (see equ 4.3). In practise, we used for instance 5 years of measures of seed production on silver fir to calibrate this model, and obtained the values of parameters A,B, and fFert, and fives values of α (one per year) (Christian Pichot, unpublished data).

This predictive model was used as a basis in TranspopRege to compute the realised number of seeds produced by a female tree j of diameter Dj, with two additional sources of stochasticity, nested one in another in the following way:

1. we introduced inter-annual stochasticity by considering the parameter α as a random variable following a Gaussian law of parameters (0, $\sigma_{interannual}$)

At each step, the value of parameter α is thus fixed at the same value α_{step} for all trees, and we compute:

$$\exp FF_j = \alpha_{step} \times (\Pi D_j - 100)^{fFert} - (A \times \alpha_{step} + B) \qquad \text{Eqn. 4.5}$$

2. We introduced inter-individual stochasticity by considering the realised number of seeds produced by a female tree j as a random variable following a Gaussian law of parameters ($\exp FF_j$, $\sigma_{inter\text{-}indiv}$), .

For each reproductive tree, we thus compute its realised female fertility by drawing a random number in the law described above.

d. Female reproductive success

We assume that the number of seedlings dispersed by an individual female *j* on patch *i* is determined by two independent factors: (1) female fertility, which is modelled as above and (2) the distance of mother *j* to patch *i* (d_{ij}) and seedlings dispersal parameters. The number of seedlings from female *j* received on a patch is given by:

$$N_{ij} = f(D_j, \sigma_{inter-annual}, \sigma_{inter-individual}) \times p_S(a_S, b_S, d_{i_j}) \qquad \text{Eqn. 4.6}$$

5. Mortality

a. Description of mortality functions

The average survival of seedlings on patch *i* is drawn in a Gaussian law of parameters ($\mu_{cellFert}$ $\sigma_{cellFer}$;). These parameters are constant all along the simulation.

b. Genetic selection

Inbreeding depression: for selfed individuals, the death probability is increased by a factor <u>SELF,</u> constant all along the simulation.
If no allelic values are specified with any loci (no traits under selection), the main changes in genetic composition is due to genetic drift, which is the direct result of demographic stochasticity combined to variation in male and female reproductive success. .

6. Management interventions
A broad range of management regimes can be formulated and evaluated in TranspopRege, through the parameterization of the rhythm, intensity, and type of thinning operation.

a. Thinning and tending operations

Basically, every kind of thinning operation can be modeled in Capsis, through in particular group specification. Using this "Group" tool, the user can specify at each step age class, diameter class or spatial class to be removed.

Any type of sylvicultural regime can thus be specified by combining thinning regime across time. Typically, we used TranspopRege to simulate high forest sylvicultural regime, with a regeneration phase (typically 20 years) and an "adult" phase (see Figure 4.2). The length of the regeneration period (**regeLength**) is a parameter of the model. During the regeneration phase, the time step can be fixed by the user (from annual time step, to multi-annual time step, for instance to mimic masting).

b. Logging and harvest operations

As any thinning operation, the date of the final harvest is specified by the user.

7. Initial reproductive stand characteristics:

The distribution of adult trees of the initial stand is generated as a clustred point process of a given **Density**, following a Neymann-Scott distribution (2 parameters: **number** and **diameter** of the aggregates, respectively **agrNumber** and **agrSize**). The diameter of each tree is drawn in a Gaussian distribution (2 parameters: **meanDiam** and **varDiam**). The genotypes of individual tree can be simulated at the user choice (nuclear or cytoplasmic genome, independent or linked loci, distribution laws for alleles).

8. Description of post-processing activities (e.g. calculation of statistics on genetic diversity or –differentiation; spatial analysis)

At each step, genetic and spatial data can be exported at the format required by Genepop and Spagedi software's

Computation of basic population genetic parameters were implemented in the script mode, in order to obtain as output a single file for all simulations, with one line/simulation, and on each line the values of the output variables of interest (genetic parameters). This file can then be used for sensitivity analyses using *R* functions (ANOVA, Morris method…)

9. References