

the water reserve r_n . All these quantities are as usual assumed to be observed with some uncertainty and a multiplicative normal noise is assumed for all of them:

$$y_n = \begin{bmatrix} \tilde{q}_n^{g\ell} \\ \tilde{q}_n^r \\ \tilde{q}_n^s \\ \tilde{q}_n^g \\ \tilde{r}_n \end{bmatrix} = \begin{bmatrix} q_n^{g\ell} (1 + \xi_n^{g\ell}) \\ q_n^r (1 + \xi_n^r) \\ q_n^s (1 + \xi_n^s) \\ q_n^g (1 + \xi_n^g) \\ r_n (1 + \xi_n^r) \end{bmatrix} = g_n(x_n, \theta, \xi_n). \quad (2.57)$$

The dynamics of the soil and plant models are displayed on Figures 2.5 and 2.6 respectively.

2.4 GreenLab for *Arabidopsis thaliana*

Arabidopsis thaliana, most commonly known as thale cress, is a small annual flowering plant from the mustard family (Brassicaceae) originating from Europe, Asia, and northwestern Africa. Its height can reach up to 25 cm. It possesses a single primary root growing vertically downwards, from which originate smaller lateral roots. The leaves form a rosette at the base of the plant and a few leaves also grow on the flowering system. White flowers grow on the stem and the fruits are siliques containing approximately between 20 and 30 seeds. The entire growth cycle of *Arabidopsis thaliana* is usually of 6 weeks.

From the beginning of the 20th century, *Arabidopsis thaliana* began to be used for research purposes. Several factors make it particularly suitable for its use in research: its small size, its rapid lifecycle, its resistance and its efficiency in self-pollination are major attractive features. It is now widely used for the study of plant sciences, including genetics, evolution, population genetics, and plant development. It was the first genome of a plant to be sequenced in 2000 because of its relatively small size: it comprises 157 million base pairs distributed across 5 chromosomes.

More precisely, in the context of this work, our interest for this plant stems from the large amount of data that the Phenoscope, a high-throughput phenotyping platform, can provide for this plant in particular. Time series of images of *Arabidopsis thaliana*, typically over a period of 21 days, can be obtained for many different individuals, and such experiments yield the data necessary for the application of hierarchical modelling to a population of plants. More details will be given in Chapter 7 on the nature of the data obtained from the Phenoscope for *Arabidopsis thaliana*.

The GreenLab model [Yan et al., 2004] is a typical functional-structural model in the sense that it combines the description of plant architectural development and ecophysiological functioning. A version has been developed for the full cycle of *Arabidopsis thaliana* growth in [Christophe et al., 2008]. Basically, a developmental submodel predicts organ appearances while source-sink dynamics is simultaneously described:



Figure 2.7: *Arabidopsis thaliana*.

biomass production is computed via radiation interception by leaf area and the produced biomass is allocated between all growing organs according to individual sink strengths. Individual leaf areas are then deduced from leaf masses. In our study, only the rosette stage of *Arabidopsis thaliana* growth is considered, which particularly simplifies the organogenesis submodel and the number of competing sinks. Furthermore, at this stage, the senescence process has not started yet.

As detailed in Section 7.2.1, leaves first appear in pairs (the 1st and 2nd leaves together, then the 3rd and 4th leaves) before the following ones start appearing rhythmically. It should be noted that, for the sake of clarity, we have numbered leaves including the 2 cotyledons, so that the first true leaf is actually leaf 3 in our numbering. The time span between the appearances of two successive leaves is called the phyllochron [Wilhelm and McMaster, 1995]. It is mostly driven by thermal time, however, the GreenLab model for *Arabidopsis thaliana* will only be used for experiments in controlled environment, and under constant thermal conditions it amounts to considering the calendar time as the main driver of organogenesis. For a better understanding of the source-sink dynamics in this first stage of study, we consider that the leaf appearance times of the first 4 leaves are known, whereas those of the subsequent leaves are such that their difference is always the phyllochron ϕ (h).

As usual, one considers that plant growth starts at germination. At this time, the biomass is supposed to be that of the seed q_0 . To take into account the photoperiod and the differences in temperature between day and night, the time step is taken to be the hour. The plant is assumed to grow only during the day, which lasts typically $n_s = 8\text{h}$ in the experimental conditions of our study. Once growth has started, the biomass produced at time step n is given by the usual Beer–Lambert law:

$$q_n = r_n \mu s \left[1 - \exp \left(-\frac{k}{s e} \sum_{v \in \llbracket 1, \nu_n \rrbracket} q_n^v \right) \right] \quad (2.58)$$

where r_n is the photosynthetically active radiation ($\text{MJ} \cdot \text{cm}^{-2}$), μ is the radiation use efficiency ($\text{g} \cdot \text{MJ}^{-1}$), s is related to the projected area of the plant (cm^2), k is the Beer–Lambert law coefficient of light extinction (dimensionless), e is the leaf mass per area ($\text{g} \cdot \text{cm}^{-2}$), ν_n is the number of leaves of the plant at time step n

and q_n^v is the biomass of the v -th leaf (g).

The pool of produced biomass is then allocated to the different organs of the plant. In the present case, that is to say the rosette stage, only the leaves are considered. It actually amounts to consider that a constant proportion α of the biomass produced at time step n is allocated to the root system, with thus a real radiation use efficiency $\mu^{\text{real}} = \mu/(1 - \alpha)$. If this assumption is known to be oversimplifying on the whole growth cycle of the plant, it remains reasonable during the rosette stage in the absence of water stress. The biomasses allocated to the different leaves are proportional to their respective demands, or sink strengths, which are functions of their expansion stage, i.e. the thermal time since appearance. In previous GreenLab models, Beta distributions were used for the sink functions. This was not judged to be the best option here since the expansion period of the leaves is not known and they did not yield optimal results. Instead, log-normal distributions were used as they allow for a similar growth dynamics with an ever ongoing growth. As was suggested by the analysis of the areas of the different leaves, two different functions were used for the first 4 (performed) leaves on the one hand and the leaves with rank higher than 5 on the other, the demand of the v -th leaf hence being:

$$d_n^v = \rho_1 \frac{f_{\log \mathcal{N}(\mu_1, \sigma_1)}(\tau_n - \tau^v)}{\|f_{\log \mathcal{N}(\mu_1, \sigma_1)}\|_\infty} \quad (2.59)$$

if $v \leq 4$ and:

$$d_n^v = \rho_2 \frac{f_{\log \mathcal{N}(\mu_2, \sigma_2)}(\tau_n - \tau^v)}{\|f_{\log \mathcal{N}(\mu_2, \sigma_2)}\|_\infty} \quad (2.60)$$

if $v > 4$. Dividing by the uniform norms ensures a proper normalization to avoid variations of the functions maximum with their parameters, though a coefficient ρ_2 allows for a different intensity of the two different kinds of leaves (ρ_1 is actually always equal to 1 and defined only for the sake of notations). An index $k(v)$ indicates whether a leaf belongs to the first 4 leaves or not, i.e. $k(v) = 1 + \mathbb{1}(v > 4)$. Here, $f_{\log \mathcal{N}(\mu, \sigma)}$ is the pdf of a log-normal distribution this time parameterized for convenience by its mean μ and standard deviation σ :

$$f_{\log \mathcal{N}(\mu, \sigma)}(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\log(x) - \mu)^2}{2\sigma^2}\right), \quad (2.61)$$

τ_n is the thermal time at time step n and τ_v is the accumulated thermal time of the v -th leaf since its emergence (both in $^\circ\text{C} \cdot \text{h}$). (μ_1, σ_1^2) and (μ_2, σ_2^2) are the parameters of the log-normal distributions for the performed leaves and those with rank higher than 5 respectively. The biomass allocated to a leaf is then the relative demand of the available produced biomass:

$$\delta q_n^v = \frac{d_n^v}{\sum_{w \in \llbracket 1, \nu_n \rrbracket} d_n^w} q_n, \quad (2.62)$$

which allows to update the cumulated biomass of each leaf:

$$q_n^v = q_{n-1}^v + \delta q_n^v, \quad (2.63)$$

and compute the related leaf areas:

$$a_n^v = e^{-1} q_n^v. \quad (2.64)$$

The GreenLab transition equations can be summarized as:

$$x_{n+1} = \begin{bmatrix} q_{n+1} \\ (d_{n+1}^v)_{v \in \llbracket 1, \nu_n \rrbracket} \\ (\delta q_{n+1}^v)_{v \in \llbracket 1, \nu_n \rrbracket} \\ (q_{n+1}^v)_{v \in \llbracket 1, \nu_n \rrbracket} \\ (a_{n+1}^v)_{v \in \llbracket 1, \nu_n \rrbracket} \end{bmatrix} = \begin{bmatrix} r_n \mu s [1 - \exp(-\frac{k}{s} \sum q_n^v)] \\ \left(\rho_{k(v)} f_{\log \mathcal{N}(\mu_{k(v)}, \sigma_{k(v)})}(\tau_n - \tau^v) \| f_{\log \mathcal{N}(\mu_{k(v)}, \sigma_{k(v)})} \|_{\infty}^{-1} \right)_{v \in \llbracket 1, \nu_n \rrbracket} \\ \left(\frac{d_{n+1}^v}{\sum d_{n+1}^w} q_{n+1} \right)_{v \in \llbracket 1, \nu_n \rrbracket} \\ (q_n^v + \delta q_{n+1}^v)_{v \in \llbracket 1, \nu_n \rrbracket} \\ (e^{-1} q_{n+1}^v)_{v \in \llbracket 1, \nu_n \rrbracket} \end{bmatrix} = f_n(x_n, u_n, \theta)$$

where $x_n = (q_n, (d_n^v)_{1:\nu_n}, (\delta q_n^v)_{1:\nu_n}, (q_n^v)_{1:\nu_n}, (a_n^v)_{1:\nu_n})$ represents the hidden state, $u_n = (r_n, \tau_n)$ the external variables and $\theta = (\phi, \mu, s, e, k, \mu_1, \sigma_1, \mu_2, \sigma_2, \rho_2, q_0)$ the parameters. There is no process noise in this model.

The GreenLab observation equations are:

$$y_n = \begin{bmatrix} \tilde{a}_n^1 \\ \vdots \\ \tilde{a}_n^{\nu_n} \end{bmatrix} = \begin{bmatrix} a_n^1 \times (1 + \xi_{1,n}) \\ \vdots \\ a_n^{\nu_n} \times (1 + \xi_{\nu_n,n}) \end{bmatrix} = g_n(x_n, \theta, \xi_n) \quad (2.65)$$

where $\xi_n \sim \mathcal{N}(0, \sigma^2 I_{\nu_n, \nu_n})$ represents the observation noises. The observations associated to the whole growth cycle of a plant is a sequence of vectors of leaf areas. Since the number of leaves is not the same at each time, a step of post-processing is used to transform this set of observations into a matrix of size $T \times \nu_{\max}$, where ν_{\max} is the maximum number of leaves, i.e. the number of leaves at time step T . The values of leaf areas that are related to:

- either leaves not emerged yet (during the first time steps, no leaf area can be attributed to a leaf that has not emerged yet),
- or unobserved values (in particular, as will be detailed in Chapter 7, the first two leaves become quickly recovered by other leaves and undetectable via analysis of zenithal images),

will be considered and marked as \circ . An example of experimental data for this model would look like:

$$y \doteq (y_{nv})_{n \in \llbracket 1, T \rrbracket, v \in \llbracket 1, \nu_{\max} \rrbracket} = \begin{pmatrix} 0.030 & 0.035 & \circ & \dots & \circ \\ 0.038 & 0.041 & 0.020 & \dots & \circ \\ \dots & \dots & \dots & \dots & \dots \\ \circ & \circ & 0.512 & \dots & 0.853 \end{pmatrix}. \quad (2.66)$$

Since the GreenLab model will be used in particular in the context of population models, some additional notations are introduced. If y_i denoted the observations related to an individual i , then the actual experimental data for the k -th leaf of this individual is:

$$y_i(k) = \{x \in [y_{nk}]_{n \in \llbracket 1, T \rrbracket} \mid x \neq \circ\}, \quad (2.67)$$

and the vectorized experimental data:

$$y_i = [y_i(1), y_i(2), \dots, y_i(\nu_{\max})] \doteq (y_{ij})_{j \in \llbracket 1, n_i \rrbracket} \in \mathbb{R}^{n_i} \quad (2.68)$$

where it is recalled that n_i represents the total number of observations for the i -th individual thus corresponding to observations at different times for each leaf. In the rest of this thesis, each leaf is identified by its rank with a specific colour as specified in Table 2.1. The dynamics of the demands and areas for each leaf are depicted on Figure 2.9 with this colour code. Only the evolution of the plant during the $n_s = 8h$ of the day is represented, and the n -th day corresponds to the n_s n -th hour on the graphs.

rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
colour	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●

Table 2.1: Colours attributed to each leaf of a specific rank for easier identification.

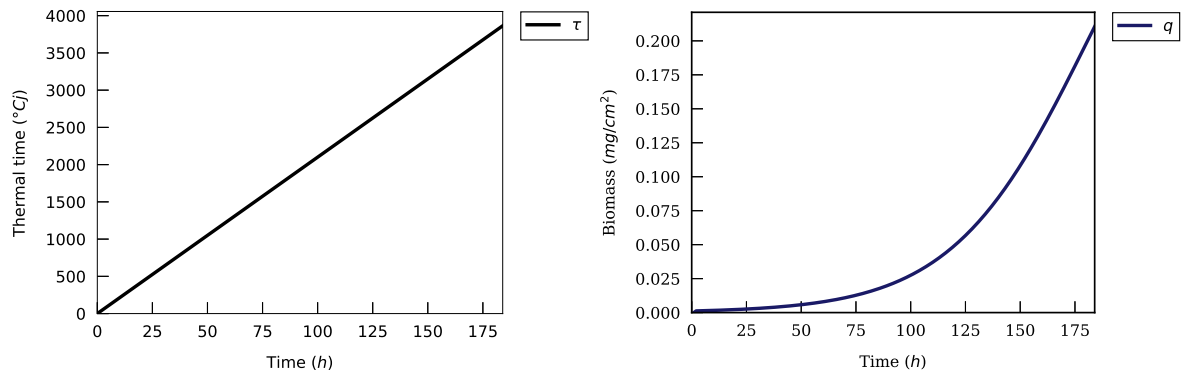


Figure 2.8: Left: example of thermal time dynamics for constant temperature. Right: production of biomass, no process noise is considered in this model for the production of biomass, contrarily to the model for sugar beet.

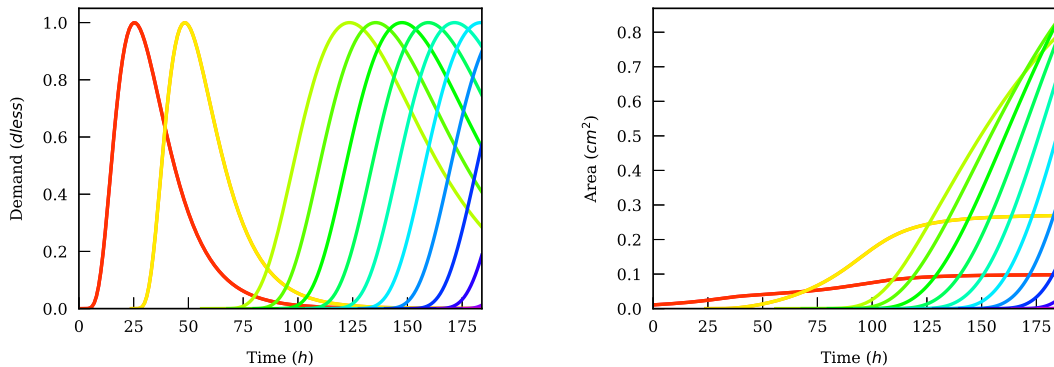


Figure 2.9: Left: demand curves d_v (dimensionless) for the different leaves, they have all been normalized with a maximum value of 1 for clarity so that the coefficients ρ_1 and ρ_2 do not play any role on this graph. Right: leaf area a_v (cm²) for the different leaves.