

theoretical guarantees. For all these reasons, we will first investigate the case of synthetic data simulated directly from the GreenLab model. This ensures that there is no uncertainty related to either the plant model or the image analysis algorithm and that the Bayesian parameter estimation method effectively works.

8.1 Simulated data

Data were simulated using the GreenLab model presented in Section 2.4. The control variables were taken to be the same as for the experimental conditions described in Section 7.1. The number of hours per day was set to $n_s = 8\text{h}$, and the daily temperature to $t_n = 21^\circ\text{C}$. The photosynthetically active radiation was taken to be $r_n = 2.52 \cdot 10^{-5} \text{MJ} \cdot \text{cm}^2 \cdot \text{h}^{-1}$. Each leaf was attributed a specific time of appearance: the first two leaves appeared at $n = 1\text{h}$, the third and fourth leaves appeared at $n = 24\text{h}$. The time of appearance for the fifth leaf was set to $n_5 = 40\text{h}$. After that, the $5 + i$ -th leaf, for $i > 1$, appears at time $n_{5+i} = n_5 + i \phi$, where $\phi = 12\text{h}$ is the phyllochron. We focus on the estimation of the parameters $\theta = (e, \mu_1, \mu_2)$, even though more parameters will be estimated in Sections 8.1.8 and 8.2.

8.1.1 Simulation of synthetic data

The dimension of the estimation problem is therefore $d = 3$. All the other parameters were fixed at constant values throughout both the simulation of data and the estimation procedure. These values were:

$$\left\{ \begin{array}{lcl} \mu & = & 3.150 \cdot 10^{+00}, \\ s & = & 5.000 \cdot 10^{+00}, \\ k & = & 7.000 \cdot 10^{-01}, \\ \sigma_1 & = & 4.593 \cdot 10^{-01}, \\ \sigma_2 & = & 3.991 \cdot 10^{-01}, \\ \rho_2 & = & 7.801 \cdot 10^{-02}, \\ q_0 & = & 3.807 \cdot 10^{-05}. \end{array} \right. \quad (8.1)$$

The parameters varying in the population are given true values for their mean and covariance matrix. In the present case:

$$\eta^{\text{true}} = \begin{pmatrix} 1.558 \cdot 10^{-03} \\ 4.531 \cdot 10^{+00} \\ 5.390 \cdot 10^{+00} \end{pmatrix} \quad (8.2)$$

and:

$$\Sigma^{\text{true}} = \begin{pmatrix} 6.069 \cdot 10^{-09} & 0 & 0 \\ 0 & 5.131 \cdot 10^{-02} & 0 \\ 0 & 0 & 7.263 \cdot 10^{-02} \end{pmatrix} \quad (8.3)$$

and $\tau^{\text{true}} = 100$. The choice of the covariance matrix actually corresponds to $\Sigma^{\text{true}} = \text{diag}\{(\eta^{\text{true}}/20)^2\}$ and that of the precision amounts to a multiplicative normal observation noise with standard deviation $\sigma^{\text{true}} = (\tau^{\text{true}})^{-1/2} = 0.1$. Data were simulated for 24 individuals: this was done on a single process and then broadcasted to all others, the underlying motivation being that all processes obviously need to have the same experimental data for the different individuals. For the i -th individual, a set of true individual parameters was first sampled from the true population distribution:

$$\theta_i^{\text{true}} \sim \mathcal{N}(\eta^{\text{true}}, \Sigma^{\text{true}}) \quad (8.4)$$

and then used for the simulation of true leaf areas $(a_{i,n}^v)_{1:\nu_{\max}, 1:n_s:T}$ (hidden states) and their corresponding noised values $(\tilde{a}_{i,n}^v)_{1:\nu_{\max}, 1:n_s:T}$ (observations) where:

$$\tilde{a}_{i,n}^v = a_{i,n}^v (1 + \xi_{v,n}) \text{ with } \xi_{v,n}^v \sim \mathcal{N}\left(0, (\tau^{\text{true}})^{-1}\right). \quad (8.5)$$

The areas of all leaves were observed on every day for 21 days which, taking into account that the phyllochron ϕ is fixed and given the rate of appearance of each leaf, amounts to observing 16 leaves over the whole growth and a number of observations of $n_i = 136$ for each individual i , i.e. a total number of observations of $n_{\text{tot}} = \sum_{i=1}^N n_i = 3264$. We denote by $\tau_i^v = (t_{i,1}^v, \dots, t_{i,n_i}^v)$ the timeline for the v -th leaf and by $\tilde{\alpha}_i^v = (\tilde{a}_{i,1}^v, \dots, \tilde{a}_{i,n_i}^v)$ the corresponding observations on leaf area at these time steps. The vector of all concatenated observations thus reads:

$$y_i = (\tilde{\alpha}_i^1, \dots, \tilde{\alpha}_i^{\nu_{\max}}) \in \mathbb{R}^{n_i} \text{ with } n_i = \sum_{v=1}^{\nu_{\max}} n_i^v. \quad (8.6)$$

The calculations were performed on as many processes as individuals, i.e. on $n_p = 24$ processes. For reproducibility issues, the seed of the random number generator was fixed so that the same data was generated all the time. However, for the estimation part and as soon as the experimental data is loaded on the different processes, the seed is reset to a random value on each process.

The flattened vector of experimental data for all individuals is denoted as $Y = (y_1, \dots, y_N) \in \mathbb{R}^{n_{\text{tot}}}$. Examples of simulated data is displayed on Figures 8.1 and 8.2. On Figure 8.1, the leaf areas for each of the first eight leaves are displayed for the 24 individuals and on Figure 8.2 are displayed some examples of typical growth curves for the first eight leaves of several individuals. Despite the rather low population covariance used, there is still a high variability in the leaf areas. This is easily seen from Figure 8.2 when comparing, for instance, the 6th and 8th individuals, the latter having maximum leaf areas twice as high as the former.

8.1.2 Initialization of the prior distributions

Defining the prior distributions of the population parameters is of crucial importance. A first step consists in obtaining realistic values for the individual parameters θ_i . This can be done rather easily by performing a GLS procedure on each individual i . These first sets of parameters can be used to compute reasonable estimates of the population mean vector η and covariance matrix Σ . A first estimate of the individual set is therefore independently computed as:

$$\hat{\theta}_i^{\text{GLS}} = \arg \min_{\theta \in \Theta} (Y_i - h_i(\theta_i))^T \Sigma_i^{-1} (Y_i - h_i(\theta_i)), \quad (8.7)$$

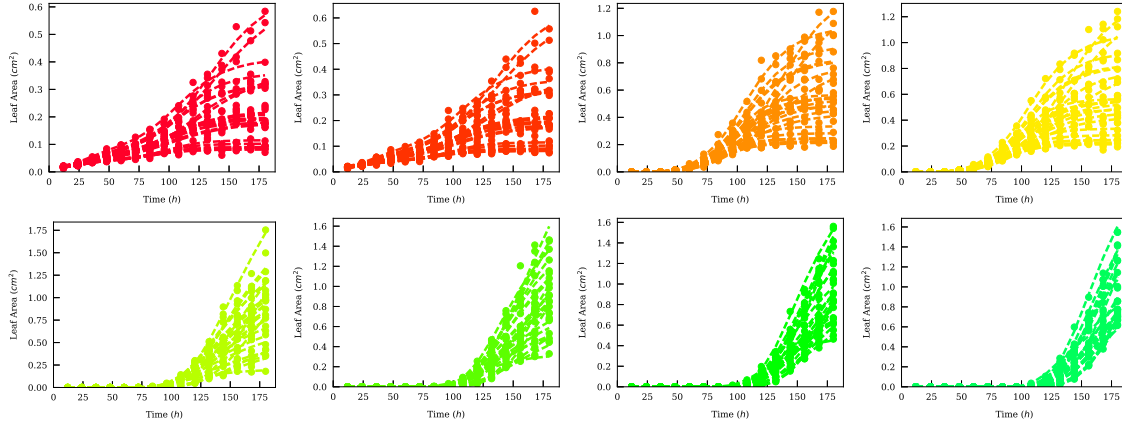


Figure 8.1: Leaf area variability for the first eight leaves. Each dashed line represents the true leaf area (hidden state) of a given individual and filled circles represent the corresponding noised data (observations).

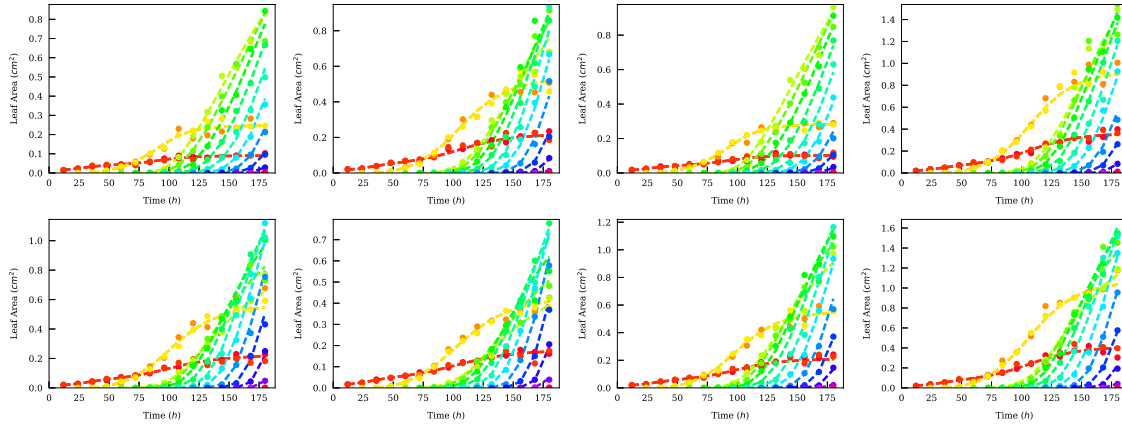


Figure 8.2: Leaf area for the first eight leaves for 8 different individuals. The hidden states are represented by dashed lines and observations by filled circles.

	e	μ_1	μ_2
$\mathbb{E}_i(\delta_i)$	$3.661 \cdot 10^{-03}$	$2.333 \cdot 10^{-03}$	$8.618 \cdot 10^{-04}$
$\max_i(\delta_i)$	$9.969 \cdot 10^{-03}$	$7.328 \cdot 10^{-03}$	$2.189 \cdot 10^{-03}$

Table 8.1: Average and maximum value of the relative errors for all individual parameters $(\theta_i)_{i \in \llbracket 1, N \rrbracket}$.

where $\Sigma_i = \text{blockdiag}_{v \in \llbracket 1, \nu_{\max} \rrbracket} \{\mathbb{V}(\tilde{\alpha}_i^v) I_{n_i^v}\}$ is the heteroskedastic matrix of the variances for each leaf, and a Gauss–Newton optimization procedure with a maximum of 1,000 iterations was used. In practice, a couple of tens iterations suffice to reach a local minimum. To assess their relevance, one can simulate the hidden states from the GreenLab model using the GLS estimates $h(\hat{\theta}_i^{\text{GLS}})$ and these results can further be compared to the hidden states that would have been obtained without a GLS algorithm; such graphs are displayed on Figure 8.3. Since the true values of the parameters are known, it is possible to compare θ_i^{true} to $\hat{\theta}_i^{\text{GLS}}$. Relative errors were calculated for each parameter and each individual. The mean and maximum values of $(\delta_i^e)_{i \in \llbracket 1, N \rrbracket}$, $(\delta_i^{\mu_1})_{i \in \llbracket 1, N \rrbracket}$ and $(\delta_i^{\mu_2})_{i \in \llbracket 1, N \rrbracket}$ were calculated to illustrate the goodness of the GLS estimates.