Modelling the interindividual variability of organogenesis in sugar beet populations using a hierarchical segmented model

Charlotte Baey, Anne Didier, Lemaire Sébastien, Fabienne Maupas, Paul-Henry Cournède

Modelling the interindividual variability of organogenesis in sugar beet populations using a hierarchical segmented model

Charlotte Baey[a,b,*], Anne Didier[c], Sébastien Lemaire[c], Fabienne Maupas[c], Paul-Henry Cournède[a,b]

[a] *École Centrale Paris, Laboratoire de Mathématiques appliquées aux systèmes, Grande Voie des Vignes, 92290 Châtenay-Malabry, France*
[b] *INRIA Saclay Île-de-France, EPI Digiplante, Parc Orsay Université, 4 rue Jacques Monod, 91893 Orsay Cedex, France*
[c] *Institut Technique de la Betterave, 45 rue de Naples, 75008 Paris*

Abstract

Modelling the interindividual variability in plant populations is a key issue to enhance the predictive capacity of plant growth models at the field scale. In the case of sugar beet, this variability is well illustrated by rate of leaf appearance, or by its inverse the phyllochron. Indeed, if the mean phyllochron remains stable among seasons, there is a strong variability between individuals, which is not taken into account when using models based only on mean population values.

In this paper, we proposed a nonlinear mixed model to assess the variability of the dynamics of leaf appearance in sugar beet crops. As two linear phases can be observed in the development of new leaves, we used a piecewise-linear mixed model. Four parameters were considered: thermal time of initiation, rate of leaf appearance in the first phase, rupture thermal time, and difference in leaf appearance rates between the two phases. The mean population values as well as the interindividual variabilities (IIV) of the parameters were estimated by the model for a standard population of sugar beet, and we showed that the IIV of

*Corresponding author, Tel: (+33) 141 131 559
Email addresses: charlotte.baey@ecp.fr (Charlotte Baey), didier@itbfr.org (Anne Didier), lemaire@itbfr.org (Sébastien Lemaire), maupas@itbfr.org (Fabienne Maupas), paul-henry.cournede@ecp.fr (Paul-Henry Cournède)

the four parameters were significant. Also, the rupture thermal time was found to be non significantly correlated to the other three parameters. We compared our piecewise-linear formulation with other formulations such as sigmoïd or Gompertz models, but they provided higher AIC and BIC.

A method to assess the effects of environmental factors on model parameters was also studied and applied to the comparison of three levels of Nitrogen (control, standard and high dose). Taking into account the IIV, our model showed that plants receiving Nitrogen tended to have a later time of initiation, a higher rate of leaf appearance, and an earlier rupture time, but these differences were not dose-dependent (no differences between standard and high dose of Nitrogen). No differences were found on the leaf appearance rate of the second phase between the three treatments.
Keywords: Nonlinear mixed model, hierarchical model, segmented regression, sugar beet, interindividual variability, plant growth modelling

---

1. Introduction

The need for a better description of plant architectural development has been long acknowledged as a key step towards the understanding of plant functional growth (Fourcaud et al., 2008). For this purpose, a new trend in plant growth modelling is the development of individual-based models combining the description of plant architecture and physiological functioning (Vos et al., 2007). However, the extrapolation to population models is yet at its early stages. The main approach consists in simulating all individuals in the population (Fournier and Andrieu, 1999; Wernecke et al., 2007; Sievänen et al., 2008; Cournède et al., 2009).

However, if there is no doubt on the theoretical interest of these approaches to help understand population functioning, they remain quite limited for concrete applications in agriculture or forestry since it is generally not possible to

describe all plants in a field or in a forest. Stochastic population models that describe the distributions of individuals' characteristics in the population provide a good way to overcome this difficulty. This approach is well-developed in forestry science at least from a descriptive point of view (Dietze et al., 2008; Vieilledent et al., 2010; Courbaud et al., 2012) to study the variability of allometric relationships.

Such interindividual variability has rarely been taken into account in dynamic plant growth models, even though its impact at the agrosystem level is well acknowledged. For example, Brouwer et al. (1993) showed how soil and crop micro-variability can have an impact on yield, as some parts of the field could be more adapted to dryness and could thus compensate poor or less good performances of other parts of the field. Renno and Winkel (1996) also showed how interindividual variability of flowering could prevent from short-term stresses risks.

In sugar beet populations, this variability is well illustrated by the number of leaves, which can be very different from one plant to another, even in the same environmental conditions. In the competition for light occurring between plants, the capacity for some plants to achieve a better and faster ground cover by the leaves will allow them to produce more biomass than their neighbors. Liu et al. (2004) showed how the differences between leaf appearance rates and emergence rates from one individual plant to another can lead to important variations in the final yield. Indeed, as light interception is directly related to biomass production, any factor affecting the speed of leaf area expansion will affect the total leaf surface area and have an impact on the final yield. The leaf appearance rate is thus a crucial parameter of plant development. It is often described through its inverse, the phyllochron, which is then defined as the thermal time (the cumulative sum of daily temperatures above a base

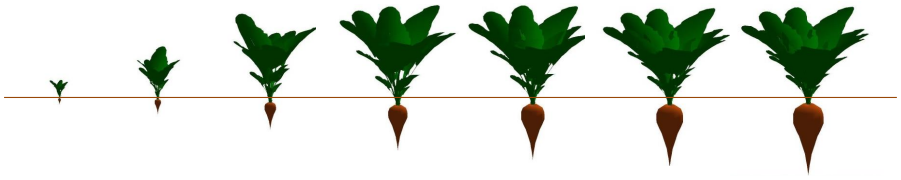

Figure 1: Sugar beet growth simulated with the DigiPant software (Cournède et al., 2006), at growth cycles 15, 17, 20, 26, 30, 34, 39 and 43.

temperature) elapsing between two successive appearances of leaves (Wilhelm and McMaster, 1995).

The variability of the phyllochron has been studied for various crops, and several environmental factors have proven to have an influence on this crucial parameter of plant development. In their study of sorghum, Clerget et al. (2008) showed that the phyllochron was positively correlated with soil temperature, and negatively correlated with photoperiod and day length. Similar results were shown by Cao and Moss (1989) for wheat and barley, and a short review of factors having an influence on phyllochron was proposed by Wilhelm and McMaster (1995). In the case of sugar beet, Milford et al. (1985a,b) showed that during a first phase of development, the phyllochron was very stable among seasons and experimental treatments (irrigation, fertilizer, plant density and sowing date). They also showed that the duration of this first phase of development, as well as the phyllochron of the second phase of development, were more subject to change. Lemaire et al. (2008) also observed these two linear phases in the development of new phytomers by the sugar beet plant. A first phase stretched from emergence to approximately the 20th leaf, and then a second phase started quite abruptly with a larger phyllochron (corresponding to a slower rate of leaf appearance rate). Different hypotheses were put forward by Milford et al. (1985a) to explain this slowing down, including changes in base temperature and an increasing competition for assimilates between leaves and

root compartments. Lemaire et al. (2008) showed that this change in leaf appearance rate corresponded to the beginning of the linear phase of root growth, and to canopy closing, when competition for light increases (Lemaire et al., 2009).

However, even though the mean phyllochron is very stable, there is a strong variability between plants within a cultivar (Frank and Bauer, 1995), leading to a high variability of the total number of leaves of each plant in a given field (Figure 2). Likewise, variations in seedling emergence can also have an influence on the final number of leaves, as shown by Durr and Boiffin (1995).

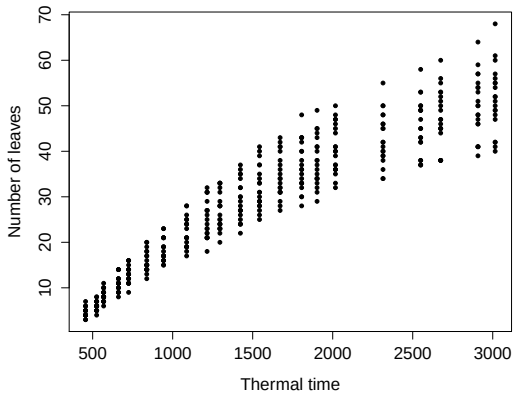

Figure 2: Number of leaves according to thermal time for 20 plants grown in normal density (10.89 pl/m^2) and with a normal level of Nitrogen (136 kg/ha)

However, the effects of the different factors (Nitrogen, density, ...) are usually assessed through the only use of mean population values, without taking into account the interindividual variability, despite its impact at the field scale. Indeed, previous studies of the phyllochron were mainly based on simple linear (or non-linear) models with no random effects. The linear models used were either based on the whole plant population (Xue et al., 2004; Frank and Bauer, 1995; Bauer et al., 1984; Streck et al., 2005; Juskiw et al., 2005), therefore

making the very strong assumption that measurements from the same individual are independent, or based on mean values (Lemaire et al., 2009), circumventing the problem of data correlation, but involving a loss of information. Moreover, with these approaches, it is not possible to estimate the phyllochron variability in the population.

One way of analyzing this variability is through the use of mixed models, in which all individuals' profiles follow the same functional form, but with parameters that vary among individuals. In this paper, we propose to study the dynamics of leaf appearance in sugar beet and develop a model accounting for interindividual variability. The number of leaves as a function of thermal time is described by a piecewise-linear mixed model with four parameters: the thermal time of initiation (corresponding to the seedling emergence), the rupture thermal time (corresponding to the setting up of the second phase) and the two rates of leaf appearance in the two phases underlined by Milford et al. (1985a) and Lemaire et al. (2008). Other models can be tested, with a nonlinear relationship between leaf appearance rate and temperature in each phase (Xue et al., 2004).

Taking the example of Nitrogen, even if the literature is abundant about its influence on sugar beet growth and development, its effect on the four parameters defined above has rarely been evaluated. We can cite for example Lee and Schmehl (1988) who reported no significant effect of Nitrogen alone on the leaf appearance rate, but a significant effect of the interactions between harvest date and Nitrogen, and between harvest date, planting date and Nitrogen on the leaf appearance rate. Stout (1961) on the other hand, reported that a high rate of Nitrogen in sugar beets stimulates the growth of new leaves. However, as stated previously, these studies either assume that all the measurements are independent, or are based on mean population values. Therefore, they are not able to

distinguish variance attributable to interindividual variability from residual error variance, putting all these sources of variability together in a single variance term. This can lead either to over- or under-estimated statistical significance when testing differences between two or more populations.

First, we developed and validated the model on a standard population of sugar beet plants, to estimate the interindividual variability of each parameter. Then, we adapted the model to assess the effects of environmental factors and applied it to the comparison of three levels of Nitrogen: control (no Nitrogen supply), normal dose (136 kg/ha) and high dose (196 kg/ha), to test whether the fertilizer dose had an effect on each parameter, taking into account the interindividual variability.

2. Material and methods

2.1. A nonlinear mixed model

Nonlinear mixed models are of particular interest for the analysis of repeated measures data, in many research fields (Ke and Wang, 2001) such as pharmacokinetics (Comets et al., 2007; Beal and Sheiner, 1982), agriculture (Hall and Bailey, 2001; Li, 2007; Makowski and Lavielle, 2006; Mutz et al., 2004; Nothdurft et al., 2006), ecology (Bolker et al., 2009), epidemiology (Lavielle et al., 2010; Morrell et al., 1995), ... Repeated-measures data can be generated by observing a number of individuals repeatedly under various conditions, assuming that the subjects constitute a random sample of the population of interest. Observations can be made on the same subject at different times (longitudinal data, for example the study of a subject growth), or under two or more different situations (for the comparisons of different treatments, for example). The analysis of repeated measures data requires particular care, to take into account the different sources of variability that could exist in the data, and the correlation

between measurements from the same individual. Classical methods relying on the assumption of independence are no longer valid in this case.

In such models, the functional form of the model linking the response variable to time (in the case of longitudinal data) is the same for all individuals, but some parameters are allowed to vary among individuals (see Lindstrom and Bates (1990), Davidian and Giltinan (1995), Pinheiro and Bates (2000) and the references therein).

2.2. A two-stage formulation

Mixed models can be written as two-stage hierarchical models. In the first stage, the number of leaves of a given plant is modelled according to the thermal time. We obtain a set of parameters for each plant, which are considered as random variables. Their variability is then assessed in the second stage (see Figure 3).

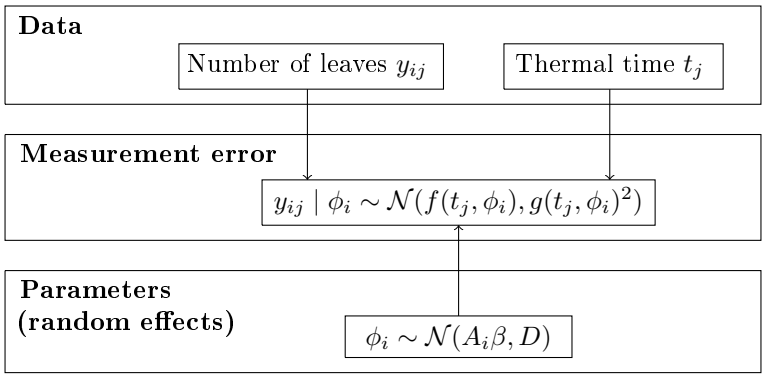

Figure 3: Hierarchical framework of our model. The number of leaves y_{ij} is modelled as a function of the thermal time t_j. The model includes the random effects ϕ_i, with associated mean values $A_i\beta$ and variance D.

First-stage: intra-individual variation

The number of leaves y_{ij} of plant i $(i = 1, \ldots, N)$ at thermal time t_j $(j =$

$1, \ldots, n_i$), is given by the following relationship:

$$y_{ij} = f(t_j, \phi_i) + g(t_j, \phi_i) e_{ij} \tag{1}$$

with ϕ_i a vector of parameters specific to individual i, and e_{ij} a random error term following a normal distribution $\mathcal{N}(0, 1)$. Function g represents the residual error model. In a first approach, we assume that g is constant equal to σ, but this hypothesis can be relaxed to include proportional error term, by letting $g = \sigma f$, or $g = a + bf$.

In our case, f is a two-linear phases function defined as follows:

$$f(t_j, \phi_i) = \phi_{i,1}(t_j - \phi_{i,0})\ \mathbb{1}_{t_j \geq \phi_{i,0}} + \phi_{i,3}(t_j - \phi_{i,2})\ \mathbb{1}_{t_j \geq \phi_{i,2}} \tag{2}$$

with $\phi_{i,0}$ the thermal time of initiation, $\phi_{i,1}$ the leaf appearance rate during the first phase of development (the inverse of the first phyllochron), $\phi_{i,2}$ the rupture thermal time and $\phi_{i,3}$ the difference in leaf appearance rates between the two phases for plant i. We obtain a vector of parameters for each individual, $\phi_i = (\phi_{i,0}, \phi_{i,1}, \phi_{i,2}, \phi_{i,3})$.

With this formulation, we model the change in slopes between the two phases, rather than the two distinct slopes, and we force the two lines to join at the rupture thermal time.

f characterizes the systematic variation and e_{ij} the random variation of measurements from individual i.

Second-stage: interindividual variation

In stage two, the N vectors of parameters obtained in the first stage are considered as random variables and their variability among individuals is modelled

through the following relationship:

$$\phi_i = A_i\beta + b_i, \quad b_i \sim \mathcal{N}(0, D) \tag{3}$$

with A_i a design matrix supposed to be known, β a vector of fixed effects, b_i a vector of random effects associated with individual i, and D a 4×4 covariance matrix. In a first approach, D was assumed to be a general positive-definite matrix.

The interindividual variability (IIV) of each parameter $\phi_{i,k}$ ($k = 0, \ldots, 3$) is represented by the corresponding k^{th} diagonal element of D, noted ω_k^2 and, as the ϕ_i are normally distributed, the mean population value are equal to $A_i\beta$. However, in more general settings, when the relationship between ϕ_i and b_i is not linear, this will generally not be the case.

We consider two formulations of matrix A_i:

- for the analysis of one plant population, no covariates are included in the model, so that $A_i = I_4$ and $\beta = (\beta_1, \beta_2, \beta_3, \beta_4)^t$, leading to the same mean population value $E(\phi_{i,k}) = \beta_k$ for all individuals.

- for the comparison of Nitrogen doses, the mean population value is allowed to vary according to the level of Nitrogen. Therefore, two covariates are introduced in the model: n_i, equal to 1 if plant i received a normal dose of Nitrogen, and 0 otherwise, and h_i, equal to 1 if plant i received a high level of Nitrogen and 0 otherwise. Plants that did not receive any Nitrogen are those for which $n_i = 0$ and $h_i = 0$. We thus have:

$$A_i = \begin{pmatrix} 1 & n_i & h_i & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & n_i & h_i & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & n_i & h_i & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & n_i & h_i \end{pmatrix}$$

and $\beta = (\beta_0, \delta_{n,0}, \delta_{h,0}, \beta_1, \delta_{n,1}, \delta_{h,1}, \beta_2, \delta_{n,2}, \delta_{h,2}, \beta_3, \delta_{n,3}, \delta_{h,3})^t$.

With this formulation, the mean population value of parameter $\phi_{i,k}$ could vary according to the level of Nitrogen through the following relationship: $E(\phi_{i,k}) = \beta_k + n_i\delta_{n,k} + h_i\delta_{h,k}$. This value is equal to β_k for plants without Nitrogen supply, to $\beta_k + \delta_{n,k}$ for plants that received a normal dose, and to $\beta_k + \delta_{h,k}$ for plants that received a high dose of Nitrogen.

Means and variances of the two phyllochrons can be approximated using Tailor Series approximation (Casella and Berger, 2002).

2.3. Inference method

A lot of inference methods have been proposed for the estimation of $\theta = (\beta, D, \sigma)$, most of them based on maximum likelihood estimation, with a likelihood function based on the joint density of the observations given the covariates. However, because of the nonlinearity of f, this density has an integral form which is in general analytically intractable.

A list of the different methods can be found in Davidian and Giltinan (2003): (i) methods based on individual estimates, (ii) methods based on an approximation of the likelihood function, mostly based on first-order and first-order conditional methods (Beal and Sheiner (1982) Wolfinger et al. (1997) Lindstrom and Bates (1990)), (iii) methods based on the "exact" likelihood, which have expanded with the increase of computational power, and (iv) methods based on a Bayesian formulation.

First-order methods are frequently used for nonlinear models, however, the likelihood approximation on which they are based may be poor. This is particularly true when the number of observations per subject is too small, or when the Gaussian assumption no longer holds (Makowski and Lavielle, 2006). An alternative to approximation methods is to use "exact" methods, in which the

likelihood function is maximized directly, using an Expectation-Maximization (EM) algorithm for example (Walker, 1996).

Delyon et al. (1999) proposed a stochastic approximation of the EM algorithm (SAEM), which converges under very general conditions to a local maximum of the function. Kuhn and Lavielle (2004) showed that the convergence to maximum likelihood estimates holds when the algorithm is coupled to an MCMC (Markov-Chain Monte-Carlo) procedure. This method has the advantages of being quicker than a usual EM-algorithm (Kuhn and Lavielle, 2005), and inference on a small number of measurements is feasible. The method is implemented in the free software MONOLIX (The Monolix Team, 2013).

2.4. Model building

A set of initial values is required by the algorithm for the estimation of the vector of parameters $\theta = (\beta, D, \sigma)$. Therefore, we first performed a fixed-effects segmented regression on each plant separately, and used the means and standard deviations of the obtained parameters as initial values. Other sets were also tested to check for the consistence of the results. Due to the high number of parameters in the model when the covariance matrix D is a general positive-definite matrix, we also tested more parsimonious covariance structures:

- a full covariance structure
- no correlation between $\phi_{i,k}$ and $\phi_{i,l}$ for $l \neq k$. For example, for $k = 0$, we have the following covariance structure :

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{pmatrix},$$

where 1 means that the corresponding coefficient will be estimated, and 0 that it is fixed to 0.

- no correlation between $\phi_{i,k}, \phi_{i,l}$ and the two other parameters, for $k \neq l$. For $\phi_{i,0}$ and $\phi_{i,1}$ for example, this lead to the following covariance structure:

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

- a diagonal covariance (i.e. no correlation between the parameters)

Using the best model, we tested whether each parameter should be entered in the model as random or could be treated as a fixed effect, by testing if the interindividual variability ω_k^2 was significantly different from 0. We used likelihood ratio tests (LRT) and model comparisons (using AIC, Akaike's Information Criterion and BIC, Bayesian Information Criterion). Similarly, the three different residual error models (additive, proportional and combined, as described in Section 2.2) were also compared using the same criteria.

For the comparison of the three Nitrogen doses, Wald tests were used to test for the covariates effects (The Monolix Team, 2013).

2.5. Model evaluation

The normalized prediction distribution errors (npde) were analyzed to test for the validity of the model (Comets et al., 2008). They are computed in Monolix from the empirical cumulative distribution function of y_{ij}. Under the null hypothesis that the model suits the data satisfyingly, these npde follow a Gaussian distribution with mean 0 and variance 1. Adequacy to the normal law was assessed by a Kolmogorov-Smirnov test. The prediction distribution was

also plotted to illustrate the model's predictions, and to check for possible over- or under-estimation of the overall variability.

2.6. Field experiments

Field experiment took place in 2011 at La Selve, France, N49°34'22", E3°59'24", on a sandy loam soil. A commercial variety, Python, was sown on March 21st with 45 cm between rows and 18 cm between seed-plots. Three different doses of Nitrogen were applied to the field: control (no Nitrogen supply, although soil was not tested for residual Nitrogen), normal level (136 kg/ha) and high level (196 kg/ha). In each of these three conditions, the final density reached 11.08 pl/m^2, 10.89 pl/m^2 and 9.14 pl/m^2 respectively.

Daily mean values of air temperature (°C) were obtained from French meteorological advisory services (Météo France) located 7 km away from the experimental site. Thermal time was computed using a base temperature of 0°C (Lemaire et al., 2008).

Leaf appearance was measured weekly and non-destructively on a group of 60 randomly selected plants (20 plants for each level of Nitrogen). Colored plastic rings were put around the petioles of the 1st, 6th, 16th, 21st, and 26th phytomers for an easier differentiation. We considered that a leaf has appeared when its length reached 10mm.

3. Results

The model was first applied to the population of 20 plants that received a normal dose of Nitrogen (136 kg/ha). Among the different covariance structures tested, the best corresponded to the case where $\phi_{i,2}$, i.e. the rupture thermal time, was uncorrelated to the other parameters. Regarding the residual error models comparison, the additive and combined models gave almost identical AIC and BIC values, that were smaller than with the proportional model. The

Description	Parameter	Estimate	Standard error	LRT for IIV
Thermal time of initiation	β_0	241	11	-
Leaf appearance rate (first phase)	β_1	0.0257	0.001	-
Rupture thermal time	β_2	1580	41	-
Difference in leaf appearance rates	β_3	-0.0136	0.0009	-
IIV of $\phi_{i,0}$	ω_0	45.5	9.1	$p < 0.0001$
IIV of $\phi_{i,1}$	ω_1	0.0043	0.0007	$p < 0.0001$
IIV of $\phi_{i,2}$	ω_2	172	31	$p < 0.0001$
IIV of $\phi_{i,3}$	ω_3	0.0038	0.0006	$p < 0.0001$
Correlation between $\phi_{i,0}$ and $\phi_{i,1}$	$\text{corr}(\phi_{i,0}, \phi_{i,1})$	0.55	0.18	-
Correlation between $\phi_{i,0}$ and $\phi_{i,3}$	$\text{corr}(\phi_{i,0}, \phi_{i,3})$	-0.52	0.19	-
Correlation between $\phi_{i,1}$ and $\phi_{i,3}$	$\text{corr}(\phi_{i,1}, \phi_{i,3})$	-0.84	0.07	-
Residual error	a	0.811	0.07	-
Residual error	b	0.0032	0.0023	

Table 1: Parameter estimation for a standard sugar beet population (2011 data, normal dose of Nitrogen). For a given parameter, its mean population value is given by β_k and its standard deviation in the population is given by ω_k

combined model was chosen since the Gaussian hypothesis for the residuals was verified with this model. It is worth noting however that the two models gave very similar results in terms of parameter estimation. Results from this model are provided in Table 1.

The interindividual variability (IIV) of the four parameters was significantly different from 0 (results from likelihood ratio tests), indicating that these parameters should be treated as random. The IIV of the rupture thermal time was higher than that of the thermal time of initiation (172°Cd for the rupture thermal time vs. 45°Cd for the thermal time of initiation). In the same way, the IIV of the second phyllochron was higher than the IIV of the first phyllochron. Using Taylor Series Expansion, the mean and standard deviation of the two phyllochrons were estimated at 39°Cd (IIV = 6.5°Cd) and 83°Cd (IIV = 16.0°Cd) for phases 1 and 2 respectively. Taking the mean population values, the predicted number of leaves at the rupture thermal time was estimated around 35.

The prediction distribution is represented on Figure 4, and shows a good

handling of the population variability by the model. Results of the normality test for npde confirmed the adequacy of our model to observations (p=0.06, see Figure

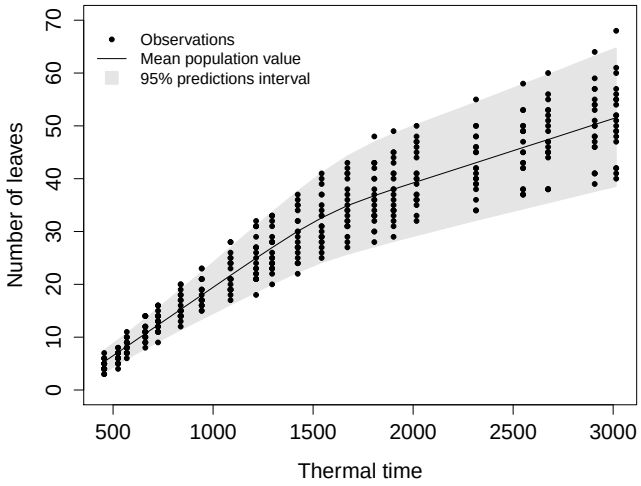

Figure 4: Prediction distribution on the population of 20 plants receiving a normal dose of Nitrogen. The solid line represents the median of predicted leaves.

Convergence of the algorithm was assessed using different sets of initial values. Other nonlinear functional forms were also tested, as suggested by Xue et al. (2004), leading to higher AIC and BIC (see Table 3 in Appendix).

Results of the comparison of the effects of the Nitrogen doses are presented in Table 2. Prediction distribution is given in Figure 6.

The thermal time of initiation was significantly earlier for plants that did not receive Nitrogen ($p<0.001$ for both doses), but there were no differences between the two doses of Nitrogen ($p=0.51$). However, the addition of Nitrogen had a beneficial effect on the leaf appearance rate of the first phase ($p=0.002$ for normal dose vs. no Nitrogen and $p<0.001$ for high dose vs. no Nitrogen), although the administered dose did not seem to have an influence on this first

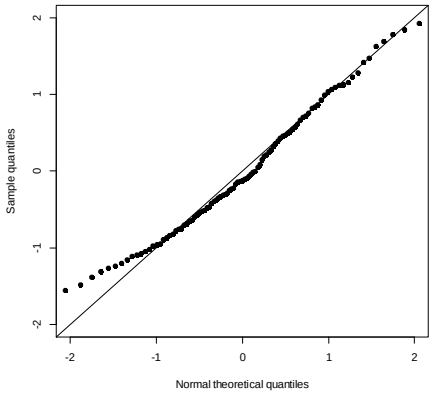

Figure 5: qqplot for the residuals of the combined residual error model.

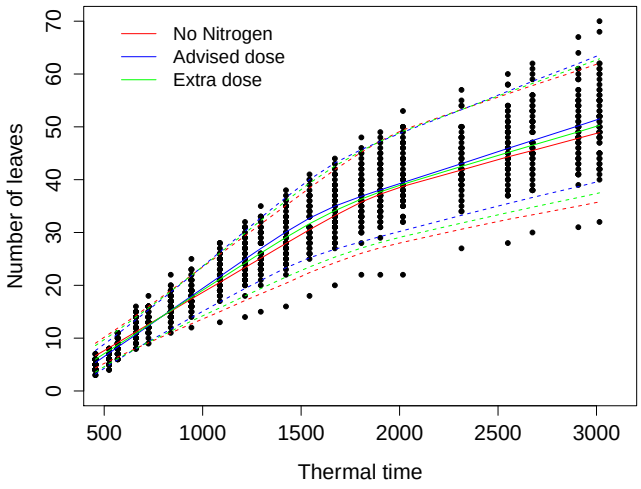

Figure 6: Prediction distribution according to the dose of Nitrogen received. The solid line represents the median of predicted leaves, and the dashed lines represent the 95% prediction interval.

Description	**Parameter**	**Estimate**	**Standard error**	**p-value***
Thermal time of initiation				
No Nitrogen	β_0	147	11	-
Normal dose	$\beta_0 + \delta_{n,0}$	242	11	< 0.001
High dose	$\beta_0 + \delta_{h,0}$	252	11	< 0.001
Leaf appearance rate of the first phase				
No Nitrogen	β_1	0.0219	0.0008	-
Normal dose	$\beta_1 + \delta_{n,1}$	0.0257	0.0008	0.002
High dose	$\beta_1 + \delta_{h,1}$	0.0263	0.0008	< 0.001
Rupture thermal time				
No Nitrogen	β_2	1860	43	-
Normal dose	$\beta_2 + \delta_{n,2}$	1580	41	< 0.001
High dose	$\beta_2 + \delta_{h,2}$	1640	41	< 0.001
Difference in leaf appearance rates				
No Nitrogen	β_3	-0.012	0.0008	-
Normal dose	$\beta_3 + \delta_{n,3}$	-0.0137	0.0008	0.15
High dose	$\beta_3 + \delta_{h,3}$	-0.0136	0.0008	0.18
Interindividual variability				
IIV of $\phi_{i,0}$	ω_0	41.6	5.2	< 0.001
IIV of $\phi_{i,1}$	ω_1	0.00373	0.00035	< 0.001
IIV of $\phi_{i,2}$	ω_2	170	18	< 0.001
IIV of $\phi_{i,3}$	ω_3	0.00357	0.00036	< 0.001
Correlations				
	$\text{corr}(\phi_{i,0}, \phi_{i,1})$	0.54	0.11	-
	$\text{corr}(\phi_{i,0}, \phi_{i,3})$	-0.46	0.12	-
	$\text{corr}(\phi_{i,1}, \phi_{i,3})$	-0.77	0.06	-
Residual error	σ	0.98	0.022	-

Table 2: Parameter estimation for the comparison of Nitrogen doses. * in the last column, the p-value corresponds to Wald Tests for the covariates, and to Likelihood ratio tests for the IIV

rate ($p=0.62$). The rupture thermal time occurred earlier for plants that received Nitrogen ($p<0.001$ for both doses). But, again, this difference was not significant between the two doses ($p=0.31$). Using the mean population values, the predicted numbers of leaves at the rupture time for the three levels of Nitrogen were estimated around 37 for control, 34 for the normal dose, and 36 for the high dose. Interestingly, no significant differences were found between the three plant populations regarding the difference in leaf appearance rates. However, given that the rate of appearance during the first phase was higher for fertilized plants, it meant that this rate remained higher during the second phase of development.

To summarize, the addition of Nitrogen delayed plant emergence, but this negative effect was counterbalanced by a higher rate of leaf appearance for fertilized plants. The rupture time also tended to be earlier for fertilized plants. However, no differences were found between the two doses of Nitrogen.

4. Discussion

The hierarchical segmented model presented here allowed for the modelling of interindividual variability, and thus a better statistical description of plant populations. Unlike classical linear models with no random effects, we were able to use the whole plant population, taking into account the correlated structure of the data. We showed that the four parameters, thermal time of initiation, rupture thermal time and leaf appearance rates had a significant interindividual variability and should therefore be treated as random parameters. Moreover, we were able to estimate this variability in a plant population, and to compare different plant populations taking into account the inherent variability.

For the study of phyllochron in the reference population (normal dose of Nitrogen, 20 plants), we found that the IIV of rupture time was higher than those

of initial time. This could be explained by the fact that differences between plants that were observed during the early stages of development became more pronounced as thermal time increased. Plants that emerged earlier and had a higher rate of leaf appearance in the first phase would tend to produce more leaves and to grow higher, and would cast shade on slower plants, which, in return, would produce less leaves as they received less light. More generally, this change of phase seems strongly influenced by plant ecophysiology and varying environmental conditions as illustrated for example by Lemaire et al. (2009), showing the strong influence of density conditions on this rupture thermal time while the phyllochrons in the two different phases were not affected. Another interesting result of our model is the non correlation between the rupture thermal time and the other parameters. This could mean that this rupture time is associated with a rather fixed number of leaves emitted by the plant, and that, whatever happens during the first growth phase does not influence the thermal time at which the second phase begins.

As for the Nitrogen doses comparison, the fact that initiation occurred later for fertilized plants can be explain by the potential damaging effects of Nitrogen on seed germination, especially when concentrated in a too large quantity near the seeds (Draycott and Christensen, 2003). Later, as the fertilized plants had a higher rate of leaf appearance, they tended to produce more leaves, and the rupture thermal time occurred also earlier than for non-fertilized plants. The standard dose of Nitrogen for sugar beet was probably sufficient to ensure that the crop did not suffer from any Nitrogen stress, as the differences were not significant between the two Nitrogen doses, and increasing the standard dose had a very low impact on crop performance. It is also interesting to note that, in addition to the fact that the rupture time was not significantly correlated with the other parameters, it could be associated with a rather fixed number of

leaves emitted by the sugar beet (around 36 leaves).

The difference of leaf appearance rate, observed during the second phase of development, was similar for the three treatments, which could mean that all plants reduced their rate of leaf production in the same way, whatever the treatment they received. Similar results were found by Lemaire et al. (2009), on the comparison of three plant densities. They found that the rhythm ratio (the ratio between leaf appearance rates of the two phases) was the same among the densities, but that the rupture point varied. Indeed, plants are in competition for light before canopy closing, but once this point has been reached, they adopt the same behavior. The same thing has also been observed in other plants like turnips, rape, kale or swede (Fletcher et al.). Sibma (1977) also showed that an increase in yield is possible by bringing forward the time of canopy closing, in a period where the potential production is higher (longer days, and more sunshine, for example). It is indeed very important to reach full leaf cover at the more appropriate time, when solar radiation is optimal (Durr and Boiffin, 1995).

The first interest of our approach is to better understand the key process of organogenesis, its variations among individuals and different environmental conditions. Moreover, a good description of organogenesis is also crucial for individual-based crop models. Being able to take into account interindividual variability is an important improvement compared to the classical strategy of simply using the mean population value as underlined by de Reffye et al. (2009): for nonlinear systems, the mean yield is not a direct function of the mean population parameters (Brouwer et al., 1993; Liu et al., 2004), and an estimation of yield variability is also a variable of interest (for example if we are interested in the sizes of organs, as for sugar beet root for example).

Once the variability of model inputs are well quantified, the main issue is

then to compute the propagation of these sources of probabilistic uncertainty in dynamic systems of plant growth. This issue has already been studied in the context of crop models by Monod et al. (2006), with the use of uncertainty propagation methods. Given a probability distribution for the input factors of a model, uncertainty analysis can be used to evaluate the uncertainty distribution of the output variables, through the use of Monte Carlo methods. This approach produces probability distributions for the outputs instead of a single value that could be misleading or difficult to interpret.

The method presented here can also be extended to other plants. Current studies are performed on winter oilseed rape, which also exhibits two linear phases of leaf development with an abrupt change in the rhythm of leaf appearance at some stage, but, contrary to the sugar beet, it corresponds to an acceleration (Jullien et al., 2011). Likewise, a simple hierarchical linear model could be used for crops with a constant phyllochron all along the growth, like maize or sunflower.

Acknowledgement

The authors are grateful to M. Lavielle (Monolix team) for helping us take full advantage of Monolix software.

Bibliography

A. Bauer, A. B. Frank, and A. L. Black. Estimation of spring wheat grain dry matter assimilation from air temperature. *Agrnomy journal*, 5(77):743–752, 1984.

S. L. Beal and L. B. Sheiner. Estimating population kinetics. *Critical reviews in biomedical engineering*, 8(3):195–222, 1982.

B. M. Bolker, M. E. Brooks, C. J. Clark, S. W. Geange, J. R. Poulsen, M. H. H. Stevens, and J.-S. S. White. Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology and Evolution*, 24(3): 127–135, 2009.

J. Brouwer, L. K. Fussell, and L. Herrmann. Soil and crop growth micro-variability in the west african semi-arid tropics: a possible risk-reducing factor for subsistence farmers. *Agriculture, Ecosystems and Environment*, 45(3-4): 229 – 238, 1993.

W. Cao and D. N. Moss. Temperature and Daylength Interaction on Phyllochron in Wheat and Barley. *Crop Science*, 29(4):1046, 1989.

G. Casella and R. L. Berger. *Statistical inference*. Pacific Grove, CA: Thomson Learning, 2nd edition, 2002.

B. Clerget, M. Dingkuhn, E. Gozé, H. F. W. Rattunde, and B. Ney. Variability of phyllochron, plastochron and rate of increase in height in photoperiod-sensitive sorghum varieties. *Annals of Botany*, 101(4):579–94, 2008.

E. Comets, C. Verstuyft, M. Lavielle, P. Jaillon, L. Becquemont, and F. Mentré. Modelling the influence of MDR1 polymorphism on digoxin pharmacokinetic parameters. *European journal of clinical pharmacology*, 63(5):437–49, May 2007.

E. Comets, K. Brendel, and F. Mentré. Computing normalised prediction distribution errors to evaluate nonlinear mixed-effect models: the npde add-on package for R. *Computer Methods and Programs in Biomedicine*, 90(2):154–166, 2008.

B. Courbaud, G. Vieilledent, and G. Kunstler. Intra-specific variability and

the competition-colonisation trade-off: coexistence, abundance and stability patterns. *Theoretical Ecology*, 5(1):61–71, 2012.

P.-H. Cournède, M.-Z. Kang, A. Mathieu, J.-F. Barczi, H.-P. Yan, B.-G. Hu, and P. de Reffye. Structural factorization of plants to compute their functional and architectural growth. *Simulation*, 82(7):427–438, 2006.

P.-H. Cournède, T. Guyard, B. Bayol, S. Griffon, F. de Coligny, P. Borianne, M. Jaeger, and P. de Reffye. A forest growth simulator based on functional-structural modelling of individual trees. In B. Li, M. Jaeger, and Y. Guo, editors, *3rd international symposium on Plant Growth and Applications (PMA09), Beijing, China*. IEEE Computer Society (Los Alamitos, California), November 9-12 2009.

M. Davidian and D. M. Giltinan. *Nonlinear Models for Repeated Measurement Data*. Chapman and Hall/CRC, 1st edition, 1995.

M. Davidian and D. M. Giltinan. Nonlinear models for repeated measurement data: An overview and update. *Journal of Agricultural, Biological, and Environmental Statistics*, 8(4):387–419, 2003.

P. de Reffye, S. Lemaire, N. Srivastava, F. Maupas, and P.-H. Cournède. Modeling inter-individual variability in sugar beet populations. In B. Li, M. Jaeger, and Y. Guo, editors, *3rd international symposium on Plant Growth and Applications (PMA09), Beijing, China*. IEEE Computer Society (Los Alamitos, California), November 9-12 2009.

B. Delyon, M. Lavielle, and E. Moulines. Convergence of a stochastic approximation version of the EM algorithm. *The Annals of Statistics*, 27(1):94–128, 1999.

M. C. Dietze, M. S. Wolosin, and J. S. Clark. Capturing diversity and interspecific variability in allometries: A hierarchical approach. *Forest Ecology and Management*, 256(11):1939 – 1948, 2008.

A. P. Draycott and D. R. Christensen. *Nutrients for sugar beet production: Soil-plant relationships*. CABI Publishing, 2003.

C. Durr and J. Boiffin. Sugarbeet seedling growth from germination to first leaf stage. *Journal of Agricultural Science*, 124:427–435, 1995.

A. L. Fletcher, E. Chakwizira, S. Maley, and M. George. Canopy development and radiation use efficiency of four forage brassica crops. In I. Yunusa, editor, *Proceedings of the 16th Australian Agronomy Conference*, Armidale, NSW.

T. Fourcaud, X. P. Zhang, A. Stokes, H. Lambers, and C. Körner. Plant growth modelling and applications: The increasing importance of plant architecture in growth models. *Annals of Botany*, 101(8):1053–1063, 2008.

C. Fournier and B. Andrieu. ADEL-maize: An L-system based model for the integration of growth processes from the organ to the canopy. Application to regulation of morphogenesis by light availability. *Agronomy*, 19(3-4):313–327, 1999.

A. Frank and A. Bauer. Phyllochron differences in wheat, barley and forage grasses. *Crop Science*, 35(1):19–23, 1995.

D. Hall and R. Bailey. Modeling and prediction of forest growth variables based on multilevel nonlinear mixed models. *Forest Science*, 47(3):311–321, 2001. ISSN 0015-749X.

A. Jullien, A. Mathieu, J.-M. Allirand, A. Pinet, P. de Reffye, P.-H. Cournède, and B. Ney. Characterisation of the interactions between architecture and

source:sink relationships in Winter Oilseed Rape (Brassica Napus L.) using the GreenLab model. *Annals of Botany*, 107(5):765–779, 2011.

P. Juskiw, J. Helm, and J. Nyachiro. Measuring phyllochrons in barley to use for seeding rate recommendations. In *18th North American Barley Researchers Workshop*, July 2005.

C. Ke and Y. Wang. Semiparametric nonlinear mixed-effects models and their applications. *Journal of the American Statistical Association*, 96(456):1272–1298, 2001. ISSN 0162-1459.

E. Kuhn and M. Lavielle. Coupling a stochastic approximation version of EM with an MCMC procedure. *ESAIM: Probability and Statistics*, 8:115–131, 2004.

E. Kuhn and M. Lavielle. Maximum likelihood estimation in nonlinear mixed effects models. *Computational Statistics & Data Analysis*, 49(4):1020–1038, 2005.

M. Lavielle, A. Samson, A. K. Fermin, and F. Mentré. Maximum Likelihood Estimation of Long Term HIV Dynamic Models and Antiviral Response. *Biometrics*, 2010.

G. S. Lee and W. Schmehl. Effect of planting date and nitrogen fertility on appearance and senescence of sugarbeet leaves. *Journal of Sugar Beet Research.*, 25(1):28–41, 1988.

S. Lemaire, F. Maupas, P.-H. Cournède, and P. de Reffye. A morphogenetic crop model for sugar-beet (*beta vulgaris l.*). In *International Symposium on Crop Modeling and Decision Support: ISCMDS 2008, April 19-22, 2008, Nanjing, China*, 2008.

S. Lemaire, F. Maupas, P.-H. Cournède, J.-M. Allirand, P. de Reffye, and B. Ney. Analysis of the density effects on the source-sink dynamics in sugar-beet growth. In B. Li, M. Jaeger, and Y. Guo, editors, *3rd international symposium on Plant Growth and Applications(PMA09), Beijing, China*. IEEE Computer Society (Los Alamitos, California), November 9-12 2009. to appear.

X. Li. Hierarchical segmented regression models with application to a wood density study. Master's thesis, Simon Fraser University, 2007.

M. J. Lindstrom and D. M. Bates. Nonlinear Mixed Effects Models. *Biometrics*, 46:673–687, 1990.

W. Liu, M. Tollenaar, G. Stewart, and W. Deen. Response of corn grain yield to spatial and temporal variability in emergence. *Crop Science*, 854(44):847–854, 2004.

D. Makowski and M. Lavielle. Using SAEM to estimate parameters of models of response to applied fertilizer. *Journal of Agricultural, Biological, and Environmental Statistics*, 11(1):45–60, 2006.

G. F. J. Milford, T. O. Pocock, and J. Riley. An analysis of leaf growth in sugar beet. I. Leaf appearance and expansion in relation to temperature under controlled conditions. *Annals of Applied Biology*, 106(1):163–172, 1985a.

G. F. J. Milford, T. O. Pocock, and J. Riley. An analysis of leaf growth in sugar beet. II. Leaf appearance in field crops. *Annals of Applied Biology*, pages 173–185, 1985b.

H. Monod, C. Naud, and D. Makowski. Uncertainty and sensitivity analysis for crop models. In D. Wallach, D. Makowski, and J. W. Jones, editors, *Working with Dynamic Crop Models*, pages 55–96. Elsevier, 2006.

C. Morrell, J. Pearson, H. Carter, and L. Brant. Estimating Unknown Transition Times Using a Piecewise Nonlinear Mixed-Effects Model in Men with Prostate Cancer. *Journal of the American Statistical Association*, 90(429), 1995.

R. Mutz, E. Guilley, U. Sauter, and G. Nepveu. Modelling juvenile-mature wood transition in Scots pine (Pinus sylvestris L.) using nonlinear mixed-effects models. *Annals of Forest Science*, 61(8):831–841, 2004.

A. Nothdurft, E. Kublin, and J. Lappi. A non-linear hierarchical mixed model to describe tree height growth. *European Journal of Forest Research*, 125(3): 281–289, Mar. 2006. ISSN 1612-4669.

J. C. Pinheiro and D. M. Bates. *Mixed-Effects Models in S and S-PLUS*. Springer, 2000.

J.-F. Renno and T. Winkel. Phenology and reproductive effort of cultivated and wild forms of Pennisetum glaucum under experimental conditions in the Sahel : implications for the maintenance of polymorphism in the species. *Canadian Journal of Botany*, 74:959–964, 1996.

L. Sibma. Maximization of arable crop yields in the netherlands. *Netherlands Journal of Agricultural Science*, 25:278–287, 1977.

R. Sievänen, J. Perttunen, E. Nikinmaa, and P. Kaitaniemi. Toward extension of a single tree functional-structural model of scots pine to stand level: effect of the canopy of randomly distributed, identical trees on development of tree structure. *Functional Plant Biology*, 35(10):964–975, 2008.

M. Stout. A new look at some nitrogen relationships affecting the quality of sugar beets. *J. Am. Soc. Sugar Beet Technol*, 11(5):388–398, 1961.

N. A. Streck, R. A. Bellé, E. K. da Rocha, and M. Schuh. Estimating leaf ap-

pearance rate and phyllochron in safflower (*Carthamus tinctorius L.*). *Ciència Rural*, 35:1448–1450, 2005.

The Monolix Team. User Guide to Monolix, 2013.

G. Vieilledent, B. Courbaud, G. Kunstler, J.-F. Dhôte, and J. S. Clark. Individual variability in tree allometry determines light resource allocation in forest ecosystems: a hierarchical Bayesian approach. *Oecologia*, 163(3):759–73, 2010.

J. Vos, L. F. M. Marcelis, and J. B. Evers. Functional-structural plant modelling in crop production. In J. Vos, L. F. M. Marcelis, P. H. B. de Visser, P. C. Struik, and J. B. Evers, editors, *Functional-structural plant modelling in crop production*, chapter 1. Springer, 2007.

S. Walker. An EM Algorithm for Nonlinear Random Effects Models. *Biometrics*, 52(3):934–944, 1996.

P. Wernecke, J. Müller, T. Dornbusch, A. Wernecke, and W. Diepenbrock. The virtual crop-modelling system VICA specified for barley. In J. Vos, L. F. M. Marcelis, P. H. B. de Visser, P. C. Struik, and J. B. Evers, editors, *Functional-Structural Plant Modelling in Crop Production*, pages 58–69. Citeseer, 2007.

W. W. Wilhelm and G. S. McMaster. Importance of the phyllochron in studying in development and growth in grasses. *Crop Science*, 35(1):1–3, 1995.

R. D. Wolfinger, X. Lin, and S. A. S. Drice. Two Taylor-series approximation methods for nonlinear mixed models. *Computational Statistics and Data Analysis*, 25:465–490, 1997.

Q. Xue, A. Weiss, and P. S. Baenziger. Predicting leaf appearance in field-grown winter wheat: evaluating linear and non-linear models. *Ecological Modelling*, 175(3):261–270, 2004.

Model	LRT (p-value)	AIC	BIC
Piecewise linear with full covariance matrix	-	1380	1396
Piecewise linear with no correlations between $\phi_{i,2}$ and $\phi_{i,k}, k \neq 2$	0.32	1377	1390
Piecewise linear with diagonal covariance matrix	< 0.0001	1399	1409
Gompertz model	-	1604	1610
Sigmoid	-	2140	2155

Table 3: Comparison of different versions of the model: the initial piecewise-linear with full covariance matrix, the same formulation but with a diagonal covariance matrix, and Gompertz or sigmoid formulation with diagonal covariance matrix.

Appendix - Model comparison

In this appendix we give the results of the comparison between four versions of the model: the original piecewise-linear version with a full covariance matrix, a piecewise-linear model with a diagonal covariance matrix, a quadratic formulation with a diagonal covariance matrix and an exponential formulation with a diagonal formulation.