

Nonlinear Models for Longitudinal Data

Peer-reviewed author version

SERROYEN, Jan; MOLENBERGHS, Geert; VERBEKE, Geert & Davidian, Marie
(2009) Nonlinear Models for Longitudinal Data. In: AMERICAN STATISTICIAN,
63(4). p. 378-388.

DOI: 10.1198/tast.2009.07256

Handle: <http://hdl.handle.net/1942/10278>

Non-linear Models for Longitudinal Data

**Jan Serroyen^{1,2}, Geert Molenberghs^{1,3}, Geert Verbeke^{3,1},
Marie Davidian⁴**

¹ Department of Methodology and Statistics, Maastricht University,
Peter Debyeplein 1, 6229 HA Maastricht, the Netherlands

² I-BioStat, Hasselt University,
Agoralaan 1, 3590 Diepenbeek, Belgium

³ I-BioStat, Katholieke Universiteit Leuven,
Kapucijnenvoer 35, 3000 Leuven, Belgium

⁴ Department of Statistics, North Carolina State University,
Raleigh, North Carolina

Author's Footnote

Jan Serroyen is Assistant Professor of Biostatistics at Methodology and Statistics, University Maastricht (email: jan.serroyen@stat.unimaas.nl). Geert Molenberghs and Geert Verbeke are Professor of Biostatistics at the Interuniversity Institute of Biostatistics and statistical Bioinformatics, Hasselt University and Katholieke Universiteit Leuven, Belgium (email: geert.molenberghs@uhasselt.be, geert.verbeke@med.kuleuven.be). Marie Davidian is William Neal Reynolds Professor of Statistics, Department of Statistics, North Carolina State University, Raleigh, NC. The first three authors gratefully acknowledge the financial support from the IAP research Network P6/03 of the Belgian Government (Belgian Science Policy).

Abstract

While marginal models, random-effects models, and conditional models are routinely considered to be the three main modeling families for continuous and discrete repeated measures with linear and generalized linear mean structures, respectively, it is less common to consider non-linear models, let alone frame them within the above taxonomy. In the latter situation, indeed, when considered at all, the focus is often exclusively on random-effects models. In this paper, we consider all three families, exemplify their great flexibility and relative ease of use, and apply them to a simple but illustrative set of data on tree circumference growth of orange trees.

Keywords: Conditional model; Marginal model; Random-effect model; Serial correlation; Transition model.

1. INTRODUCTION

In practice, data are often collected repeatedly over time on the same subject. Especially for Gaussian data, quite a number of approaches for analyzing longitudinal data have been developed and implemented in standard software packages (Verbeke and Molenberghs 2000). Most of the methodological work has been done in the setting of linear and generalized linear models. While non-linear models have also been extended in various ways, the work is not as taxonomically ‘complete’ as for the linear and generalized linear ones. Moreover, many researchers coerce their problems, and hence models, in the linear and generalized linear frameworks. Sometimes, but not always, this is a sensible route. Owing to methodological as well as computational advances, non-linear models for hierarchical data are definitely a viable option for routine use by anyone with a working knowledge on linear and generalized linear models.

In its broadest sense, the term *non-linear model* refers to any function that depends on parameters of interest in a non-linear fashion. The term also has been used to describe formulations that involve more complicated non-linear dependence on parameters than that in the generalized linear class, where parameters enter linearly at the level of a predictor function (e.g., at the logit scale for binary data), but are then transformed through a well-defined non-linear link function (e.g., the logit link), to describe the mean at the natural scale (e.g., the probability scale). In this paper, we consider the unrestricted type of non-linear model.

With linear models in general and linear mixed models in particular, although the *estimation* of the regression parameters must take into account the correlation among measurements on the same unit, in view of valid precision estimation, their *interpretation* is essentially independent of the variance and correlation structures. With non-linear models, including generalized linear mixed models, different assumptions about the sources of variability and correlation can lead to regression parameters with distinct interpretations (Diggle *et al* 2002, Molenberghs and Verbeke 2005) and often widely varying magnitudes. The data analyst must therefore reflect

ever so carefully about the objectives of the analysis and the sources of variance and correlation in choosing an approach.

Diggle *et al* (2002) and Molenberghs and Verbeke (2005), among others, distinguish between three different model families for longitudinal data: *marginal*, *random-effects*, and *conditional* models, the latter family including so-called *transition* models. While this taxonomy has been extensively used for both linear and generalized linear model settings, it is less routinely encountered in a non-linear context. Arguably, it is common to make use of mixed-effects models in this area, especially in such applications as pharmacokinetic and pharmacodynamic modeling (Davidian and Giltinan 1995, Molenberghs and Verbeke 2005). These are among the most important areas where the reason for adopting a non-linear model is dictated by the science. This usually means that the non-linear model is derived from mechanistic theoretical considerations, typically describing phenomena going on at the level of a unit. This then naturally leads to a random-effects perspective. For example, the basic logistic growth model, to which we will return in subsequent sections, can be derived by making a mechanistic assumption about how the change in growth is related to current growth and then solving a simple differential equation. At the same time, it is entirely legitimate to view the logistic growth model, from an empirical perspective, as a function with an 'S-shape' and hence that can approximate marginal mean functions or individual behavior observed in lots of different areas. While the use of random-effects, or hierarchical, models is definitely sensible in many settings, it is recommendable to also explore the use and relative merits of, the marginal and conditional families.

In Section 2, a simple set of motivating data, the orange tree data, is introduced. The three model families are surveyed and applied to the data in Section 3. Section 4 is devoted to a comparative study in song birds, to which marginal and random-effects models are applied. Concluding remarks are offered in Section 5.

2. ORANGE TREE DATA

Draper and Smith (1998, Exercise 24.N, p. 559) present data from an experiment in which trunk circumference (in mm) was measured for 5 orange trees, on 7 different occasions, over roughly a 4-year period of growth. The observed individual profiles are plotted in Figure 1. The data are presented in Table 1.

A growth study such as this may have been conducted to address different scientific objectives. Under *Objective 1*, interest may focus on characterizing the average growth pattern in the population; here, how mean circumference in the population of trees changes over time. Under *Objective 2*, the goal may be to understand how growth patterns of individual trees vary across the population.

Figure 1 shows that both individual tree profiles and the marginal mean profile exhibit an ‘S shape’ over time, suggesting that appropriate statistical models incorporate a function capable of approximating this pattern. Letting t denote time, the logistic function

$$\frac{\beta_1}{1 + e^{-(t-\beta_2)/\beta_3}} \quad (1)$$

is a natural choice; the interpretation of the parameters β_1 , β_2 , and β_3 in (1) follows from the right-hand panel of Figure 1, with β_1 being the asymptotic circumference, β_2 the time at which half of the asymptotic value is realized, and β_3 a measure for curvature at the time half of the asymptotic value is reached. The logistic function not only is suitable for this purpose on empirical grounds, but it also can be derived by making certain assumptions regarding the mechanisms underlying growth of an individual tree. This route was followed by Draper and Smith (1998), who advocated this model, among others, for representing patterns of growth. In the following sections, we consider random-effects, marginal, and conditional models, starting from (1) and varying to the theme.

3. MODEL FAMILIES

As mentioned earlier, we can distinguish between three model families. We will now give a brief overview of these model families based on Diggle *et al* (2002) and Molenberghs and Verbeke (2005).

Marginal regression methods characterize the marginal expectation of a Gaussian or non-Gaussian response, Y , as a function of explanatory variables, X . By marginal expectation, we mean the average response over the sub-population that shares a common value of X . The methods are designed to permit separate modeling of the regression of Y on X on the one hand, and the association among repeated observations of Y on each individual on the other hand. Marginal models are appropriate when inferences about the population average are the focus. This is to be contrasted with the random-effects and transitional models where the covariate effects and the within-subject association are modeled through a single equation, even though alternative representations are possible (Lee and Nelder 2001). All three approaches lead to the same general class of linear models for Gaussian data. But in the non-Gaussian case, different (non-linear) models can lead to different interpretations for the regression parameters. The choice of model family should therefore be such that it allows for formulating an answer to the research question posed. As we will see in what follows, this does not necessarily imply that there is a one-to-one relationship between a research question and a model family. For example, when the research question is of a marginal nature, the marginal family is obviously a candidate but it is sometimes possible to derive marginal quantities from hierarchical and/or conditional models, especially when there is no theory-prescribed model formulation. Essentially, considerations of statistical and computational efficiency, as well as model parsimony, need to be brought into the picture as well (Lee and Nelder 2004). Having said this, three points should be kept in mind. First, within a model family, due consideration should be given to model assessment, so as to ensure a parsimonious, well fitting model is chosen. Second, often marginal quantities can be obtained from random-effects models as well, for example, through analytical, numerical, or sampling-based marginalization. This means that quantities in one family may result from applying appropriate manipulation to

quantities from another model family. Third, everything else equal, consideration may be given to model-formulation elegance and therefore, often also, computational convenience (Lee and Nelder 2004).

In random-effects models, the response is assumed to be a function of explanatory variables with regression parameters that may vary from one individual to the next. This variability reflects natural heterogeneity stemming from unmeasured (*latent*) factors. Given these so-called subject-specific parameters, the responses are often assumed independent. This does not preclude that more elaborate models are possible if residual dependence is detected. A random-effects model is a reasonable paradigm if the set of parameters from a population of subjects can be thought of as a sample from a distribution.

In conditional models, the sequence of repeated measures is modeled conditional upon (a subset of) the other outcomes. This could be the set of all past measurements or a subset thereof. In transition models, the subset is restricted to the past measurements, often further narrowed to the immediately preceding (few) measurements.

Let us now analyze the orange tree data within each of the three model family settings.

3.1 Random-effect Models

In a random-effects model, we also condition upon a random-effects vector \mathbf{b}_i :

$$E(Y_{ij}|\mathbf{b}_i, \mathbf{x}_{ij}, \mathbf{z}_{ij}) = h(\mathbf{x}_{ij}, \boldsymbol{\beta}, \mathbf{z}_{ij}, \mathbf{b}_i), \quad (2)$$

where \mathbf{x}_{ij} and \mathbf{z}_{ij} are design vectors for the fixed effects and random effects, respectively, and $\boldsymbol{\beta}$ is a vector of fixed-effects regression parameters. Conventionally, but not always, is the distribution of the random effects assumed to be of a normal type. A broad class of alternative formulations can be found in Lee and Nelder (1996, 2001, 2003).

The following non-linear mixed model has been proposed by Pinheiro and Bates (2000) for the orange-tree data:

$$Y_{ij} = \frac{\beta_1 + b_i}{1 + \exp[-(t_{ij} - \beta_2)/\beta_3]} + \varepsilon_{ij}, \quad (3)$$

with $b_i \sim N(0, d_{11})$ and $\varepsilon_{ij} \sim N(0, \sigma^2)$.

A graphical representation of the model, with the interpretation of the associated fixed-effects parameters, is given in Figure 1. Model (3) can be interpreted as a description of what is happening at the individual-tree level, letting the logistic function describe the pattern for tree i . The mixed intercept, $\beta_1 + b_i$, allows the asymptote to vary among the members of the tree population, while the lack of random effects associated with the other two fixed effects implies that these do not. In case one starts from a mechanistic point-of view (Objective 1 in Section 2), then a random-effects model like (5) or (6), to be introduced in what follows, would be a natural choice.

To improve numerical stability, and hence convergence, while fitting the model, both the response values, trunk circumferences, and the covariate values, age, were divided by 100. The model was fitted using the SAS procedure NLMIXED.

A plot of the model fit can be found in Figure 2. The fitted mean profile is a smooth curve with a sigmoidal shape, although the curvature is not pronounced. We will return to the way in which the functional form of the marginal shape is derived. The marginal fit of the model seems acceptable. Parameter estimates and standard errors are given in Table 2. Empirical Bayes predictions, a term used to indicate estimates of the random effects (Verbeke and Molenberghs 2000, Molenberghs and Verbeke 2005) are graphed in Figure 2. The individual-tree model fit seems acceptable, nicely combined with the three parameters' clear meaning.

Note that this model is non-linear in the fixed-effect parameters, but linear in the random effect b_i , simplifying the calculation of the marginal mean over the random-effects distribution. Indeed, the conditional mean:

$$E(Y_{ij}|b_i, t_{ij}) = \frac{\beta_1 + b_i}{1 + \exp[-(t_{ij} - \beta_2)/\beta_3]}, \rightarrow E(Y_{ij}|t_{ij}) = \frac{\beta_1}{1 + \exp[-(t_{ij} - \beta_2)/\beta_3]}. \quad (4)$$

In general, there will be situations where an analytical expression for the marginal mean will either not exist or will impose excessive computational complexity. One can thence revert to expansion-based, numerical-integration-based, or Monte Carlo methods. The expansion-

based methods may be poor in terms of accuracy, whereas the numerical-integration options may be cumbersome computationally. When also measures of precision are required, the issue is compounded further. It is exactly these options that are at the analyst's disposition for maximizing generalized linear and non-linear likelihood functions (Molenberghs and Verbeke 2005).

It is therefore instructive to consider alternative models, such as the following two formulations:

$$Y_{ij} = \frac{\beta_1 + b_{i1}}{1 + \exp[-(t_{ij} - \beta_2 - b_{i2})/\beta_3]} + \varepsilon_{ij}, \quad (5)$$

$$Y_{ij} = \frac{\beta_1 + b_{i1}}{1 + \exp[-(t_{ij}^{1+b_{i2}} - \beta_2)/\beta_3]} + \varepsilon_{ij}. \quad (6)$$

In both cases, it is assumed that the vector of random effects $(b_{i1}, b_{i2})'$ is zero-mean bivariate normally distributed with variance-covariance matrix:

$$\begin{pmatrix} d_{11} & d_{12} \\ d_{12} & d_{22} \end{pmatrix}.$$

In (5), the half-growth is assumed to be random, whereas (6) modulates the effect of time in a tree-specific fashion. Parameter estimates are displayed in Table 2. Comparing each of these models with (3) can be done via a likelihood-ratio test, with reference distribution a 50 : 50 mixture of χ_1^2 and χ_2^2 distributions, given that the null hypothesis lies on the boundary of the parameter space (Self and Liang 1987, Stram and Lee 1994, 1995, Verbeke and Molenberghs 2000, 2003, Molenberghs and Verbeke 2007). For (5), the likelihood ratio test statistic is $G^2 = 1.4$ ($p = 0.3667$). For (6), we find $G^2 = 0.4$ ($p = 0.6729$). Thus, neither of these extensions substantially improves the fit. This is not surprising, because from Figure 2(a) it is clear that the initial random-effects model (3) is well fitting, both in terms of the overall mean, as well as for the individual trees.

For (5) and (6), marginalization is not straightforward and has to be done numerically or sampling based. Related to this, the fixed effects in models (5) and (6) no longer enjoy the simple meaning of describing 'the average tree,' since, for example,

$$E_{b_{i1}, b_{i2}} \left\{ \frac{\beta_1 + b_{i1}}{1 + \exp[-(t_{ij} - \beta_2 - b_{i2})/\beta_3]} \right\} \neq \frac{\beta_1}{1 + \exp[-(t_{ij} - \beta_2)/\beta_3]}, \quad (7)$$

with the same holding for (6), of course. Rather, and this is a somewhat subtle point, the fixed-effects parameters describe *a hypothetical tree with the average values for the random parameters*. Generally, only when the random-effects enter linearly, then both interpretations coincide. Evidently, the (negative) result of (7) is the norm, with model (3) being the exception.

3.2 Marginal Models

A marginal non-linear model for an outcome Y_{ij} at measurement occasion j for subject i , conditional on a vector of covariates \mathbf{x}_i , would take the form:

$$E(Y_{ij}|\mathbf{x}_{ij}) = h(\mathbf{x}_{ij}, \boldsymbol{\beta}), \quad (8)$$

where $\boldsymbol{\beta}$ is the vector of regression parameters and h is the non-linear link function. One might start from such a formulation to address Objective 1. Expression (8) does not specify the full joint distribution, and the association structure needs to be specified, linearly or non-linearly. We can then consider full-likelihood approaches. However, these can become computationally complex, even for outcome sequences of moderate length, except when the outcomes are Gaussian. Classically, this has motivated the formulation of non-likelihood alternatives, such as generalized estimating equations (GEE, Liang and Zeger 1986). One ought to be aware that a method such as GEE is sensible only when the correlation structure is not of direct interest, because it is treated merely as a nuisance characteristic in GEE. More generally, GEE does not allow for probabilistic statements, an issue not occurring for fully specified model. Therefore, the problem is rather rare with Gaussian data.

Note that, at this point, there is a fundamental difference between (non-linear) models for continuous, normally distributed data and non-Gaussian data. For the former, next to the correlation structure also the variance needs to be modeled, whereas for the latter one typically starts from an exponential family formulation, where the variance is a deterministic function of the mean, the so-called variance link.

We consider marginal model (4), with serially correlated errors, where the serial correlation

is assumed to be of an exponential type. This means that the correlation between pairs of measurements on the same tree decreases linearly on the log-scale with increasing time lag u between measurements, i.e., a functional form $\exp(-\phi u)$ is assumed. The resulting model is given by

$$Y_{ij} = \frac{\beta_1}{1 + \exp[-(t_{ij} - \beta_2)/\beta_3]} + \varepsilon_{(1)ij} + \varepsilon_{(2)ij}, \quad (9)$$

with $\varepsilon_{(1)ij} \sim N(0, \sigma^2)$ and $\varepsilon_{(2)i} \sim N(0, \tau^2 H_i)$. Note that, even though the same parameters have been used in (9) as in (3), they do have a different meaning. The same comment will apply to transition model (14) as well.

Obviously $\varepsilon_{(2)i} = (\varepsilon_{(2)i1}, \dots, \varepsilon_{(2)i7})'$. The elements of H_i take the form $h_{i,jk} = \exp(-\phi u_{jk}) = \exp(-\phi |t_{ij} - t_{ik}|)$. It is possible to extend the model in such a way that it combines features of two, or even all, model families. For example, when (9) is supplemented with random-effects, as in (3), a combined model would arise, i.e., a random-effects model, with the conditional independence assumption replaced by one of residual association, in addition to the one captured by random effects. Note that both the serial as well as the random-effect components allow for subject-specific inferences. These are displayed, in Figure 2, for the orange tree data.

This marginal model was implemented in the software package R. Note that the R function GNLS is available to this effect, as well. Our implementation also allows for the flexibility typically required in a longitudinal context. The marginal log-likelihood function for the multivariate normal distribution was constructed and then maximized using a general-purpose numerical optimizer based on a quasi-Newton method, allowing to directly model the serial correlation. The exponential serial correlation function was selected, because it leads to convergence and fits well (see below). The orange tree data set is relatively small, forcing us to keep the modeled correlation function reasonably simple. The parameter estimates of the marginal model can be found in Table 2, while Figure 2 shows the observed and fitted profiles for the marginal model. When comparing marginal with random-effects parameter estimates, we observe that the parameter estimates are very similar. This is not surprising, because the

parameters in the random-effects model (3) have a marginal interpretation. Generally, when marginal descriptions are desired from random-effects models, one may, once again, have to resort to numerical methods.

Nevertheless, in spite of the similarity among the parameter estimates, Table 2 displays a much wider gap between the random-effects-model-based and marginal standard errors, especially for the half-growth and shape parameters. This combines with a rather large estimate for the rate of exponential decay, ϕ . To shed some light on this issue, note that (9) implies a constant variance over time, $\sigma^2 + \tau^2$. Indeed, many commonly used models for serial correlation would imply a constant variance and a correlation function that, among other things, depends solely on the time lag between measurements. However, Figure 1 provides evidence, in harmony with Table 1, for an increasing spread over time, and hence such assumptions would be too restrictive. Random-effects model (3) allows for this because its marginal model has mean (9) and covariance matrix $V_i = d\mathbf{N}_i\mathbf{N}_i' + \sigma^2 I_{n_i}$, with obvious notation, and the components of \mathbf{N}_i given by $N_{ij} = \{1 + \exp[-(t_{ij} - \beta_2)/\beta_3]\}^{-1}$. Evidently, while tractable, this covariance matrix is structured but not constant and, importantly, different from the covariance assumptions made earlier for (9). In this case, the random-effects model does allow for a time-varying variance function. The corresponding marginal covariance matrix, derived from random-effects model (3), takes the form:

$$\widehat{V}_{r.e.} = 0.1 \cdot \mathbf{N}\mathbf{N}' + 0.006 \cdot I_7. \quad (10)$$

Details of the matrices are not shown here, but can be found in a web based appendix to this manuscript. In contrast, the variance-covariance matrix resulting from marginal model (9) is:

$$\widehat{V}_{\text{marg}} = 0.042 \cdot \exp(-24.72|\mathbf{t} - \mathbf{s}|) + 0.005 \cdot I_7.$$

Here, $\mathbf{t} \equiv \mathbf{t}_i = (t_{i1}, \dots, t_{i7})'$, the vector consisting of the seven measurement occasions and \mathbf{s} is another set of 7 measurement occasions. Comparing both of these estimated variance-covariance matrices with the observed-data counterpart shows that the random-effects version, while not perfect in terms of fit, is considerably closer than the marginal-model-based one.

So far, we have focused on the variance function. The dependence can be studied via the corresponding correlation matrices. Inspection of these (not shown) reveals that the discrepancies between the observed correlation and both sets of estimates is less dramatic than for the variance function of the marginal model. Also here, while not perfect, the fit of the random-effects model is better than that of its marginal counterpart. However, the overall fit is clearly acceptable as further evidenced, for example, from the non-significance of extended random-effects models (5) and (6).

Overall, the conclusion would be that the marginal model, unlike in the random-effects case, needs extension in primarily the variance structure and perhaps also somewhat in the correlation structure. It is therefore useful to allow for a general decomposition of the form $V_i = A_i^{1/2} \cdot R_i \cdot A_i^{1/2}$, where A_i is a diagonal variance matrix for tree i and R_i a correlation structure. For example, we could assume a linearly or quadratically increasing variance function and an first-order auto-regressive correlation matrix. To make the latter consistent with the correlation structure induced by marginal model (9), it is advisable to account for the varying time lags between adjacent measurements. To implement this, we will retain the mean structure of model (9) but change the correlation and variance structures. For the correlation matrix $R = R_i$, assume the elements are $r_{i,jk} = \exp(-\phi u_{i,jk})$ where $u_{i,jk} = |t_{ij} - t_{ik}|$. Further, two variance functions A_i are considered:

$$a_{i,jj,2} = \exp(\theta_0 + \theta_1 t_{ij}), \quad (11)$$

$$a_{i,jj,3} = \exp(\theta_0 + \theta_1 t_{ij} + \theta_2 t_{ij}^2). \quad (12)$$

The first one assumes a linear mean function on the logarithmic scale, whereas the second one allows for a quadratic effect, too. Parameter estimates are displayed in Table 2. The variances, corresponding to (11), are much closer to the observed variances. The variance function for (12) is closer still to the observed variances than is the case for (11). Also here, the fit has improved but is not perfect. Extending the model further is challenging, given the dataset consists of only 5 trees, each of which provides 7 measurements, i.e., the number of

measurements is larger than the number of independent replicates.

Comparing nested models (12) and (9), the likelihood ratio test statistic equals $G^2 = 15.20$ ($p < 0.0001$). In comparison with the random-effects models, the fixed effects are less precisely estimated across all marginal models. This underscores it can be of value to consider a random-effects model, even when scientific interest predominantly is of a marginal nature.

3.3 Conditional and Transition Models

A conditional non-linear model would in addition allow \overline{Y}_{ij} , the set of all outcomes save Y_{ij} , as an argument of h :

$$E(Y_{ij}|Y_{ik,k \neq j}, \mathbf{x}_{ij}) = h(\mathbf{x}_{ij}, \beta, \overline{Y}_{ij}, \alpha), \quad (13)$$

where α is a vector grouping the auto-regressive effects as well as the variance-component parameters.

From a mechanistic perspective on growth, in harmony with Objective 1, conditional models are hard to justify; they really are solely empirical representations (Objective 2) that can induce a marginal model and take account of correlation.

Generally specifying a set of compatible models of the type (13) is not straightforward. Indeed, even for two outcomes, it is far from automatic that specifying both $f(y_{i1}|y_{i2})$ and $f(y_{i2}|y_{i1})$ would correspond to one and only one bivariate distribution. Often, no such distribution might exist. This issue is well-known in, for example, spatial statistics (Cressie 1991), where the Hammersley-Clifford result governs the conditions under which conditionally specified distributions correspond to a joint distribution. For the log-linear model, elegant and tractable conditions result, but this is generally not the case. Such issues have been studied by Arnold, Castillo, and Sarabia (1992) as well.

For longitudinal data, the time-ordered nature of the measurements comes to rescue, because restricting the conditioning in (13) to past measurements automatically ensures a coherent specification, thanks to the law of total probability.

Inspired by the random-effects modeling exercise already undertaken, one possible conditional model for the orange tree data is a transition model where the random effect in (3) is replaced by the previous measurement $Y_{i,j-1}$. This model can be expressed as

$$Y_{ij} = \frac{\beta_1 + \gamma Y_{i,j-1}}{1 + \exp[-(t_{ij} - \beta_2)/\beta_3]} + \varepsilon_{ij}. \quad (14)$$

Additional calculations are required for the marginal mean, because it does not follow directly from the conditionally specified model (14). However, it is relatively straightforward to compute it in a recursive fashion. Indeed, the obvious recursive mean relationship is:

$$E(Y_{ij}) = \frac{\beta_1 + \gamma E(Y_{i,j-1})}{1 + \exp[-(t_{ij} - \beta_2)/\beta_3]}. \quad (15)$$

In Figure 2, we show a plot of the model fit. The fitted mean profile seems to provide an acceptable fit. This curve is not as smooth as the random-effects model, but nevertheless follows the data closely. Individual predictions are presented in Figure 2. Table 2 provides an overview of the parameter estimates. The γ parameter estimate is positive, indicating an increment over time, i.e., *growth*. As expected, parameter estimates for β_1 , β_2 , and β_3 are clearly different from the estimates for the random-effects model. The conditional parameter interpretation renders answering the substantive question rather difficult (Diggle *et al* 2002, p. 142–144), when such a question is framed in marginal terms.

Evidently, (14) is not the only option to formulate a transition model. To emphasize this, let us add $Y_{i,j-1}$ to the denominator of the model expression:

$$Y_{ij} = \frac{\beta_1}{1 + \exp\{-(t_{ij} - \beta_2)/\beta_3 + \gamma Y_{i,j-1}\}} + \varepsilon_{ij}. \quad (16)$$

In this case, calculation of a recursive relationship is less obvious, owing to the non-linear fashion in which $Y_{i,j-1}$ enters (16). One can proceed by calculating a Taylor-series expansion. Opting for a second-order expansion, the following approximate recursive relations are sufficient to compute the mean and variance profiles:

$$\begin{aligned} E(Y_{ij}) &\simeq \beta_1 \left[\pi_{ij} + \gamma \pi_{ij} (1 - \pi_{ij}) E(Y_{i,j-1}) + \frac{\gamma^2}{2} \pi_{ij} (1 - \pi_{ij}) (1 - 2\pi_{ij}) E(Y_{i,j-1}^2) \right], \\ E(Y_{ij}^2) &\simeq \beta_1^2 \left[\pi_{ij}^2 + 2\gamma \pi_{ij}^2 (1 - \pi_{ij}) E(Y_{i,j-1}) + \gamma^2 \pi_{ij} (1 - \pi_{ij}) (2 - 3\pi_{ij}) E(Y_{i,j-1}^2) \right], \end{aligned}$$

where $\pi_{ij} = \{1 + \exp [-(t_{ij} - \beta_2)/\beta_3]\}^{-1}$. Checking the asymptotic correlation matrix of the parameter estimates reveals signs of multicollinearity; strong correlations were observed between β_2 and β_3 (0.99), between β_3 and γ (0.96), and between β_2 and γ (0.92). Thus, issues of model stability can be raised. The parameter estimate for the asymptote (Table 2) is much more in line with the estimate from the random-effects model, which is in contrast to the first transition model, while the opposite is true for the half-growth and shape parameter. This underscores that it is very difficult to make overall statements about the behavior of a particular modeling family relative to another one in the non-linear case (Davidian and Giltinan 1995).

Note that fitting transition models is computationally rather simple. Thanks to the law of total probability, a sequence of n_i correlated outcomes is replaced by n_i univariate outcomes, given the history. When models would be linear, simple linear regression or generalized linear models technology can be used. In our case the non-linearity persists. It is then still advisable to use a tool such as the SAS procedure NLMIXED. Of course, given there is no need to include random effects, reaching convergence will typically be considerably simplified and computation time, when compared to random-effects and marginal models, often drastically shortened.

Generally speaking, convergence is a tricky, multi-faceted problem, of which the choice of software tool is but one aspect. There is also similarity between the non-linear and generalized linear cases. For example, Sheiner and Beal (1980) presented a method for non-linear mixed models, which corresponds to Breslow and Clayton's (1993) MQL for GLMM. The same is true for Lindstrom and Bates (1990) relative to PQL, and Vonesh (1996) to Laplace approximation. Furthermore, there is also the hierarchical likelihood approach of Noh and Lee (2008) for the non-linear case, which builds upon Lee and Nelder (1996, 2001, 2003, 2004).

4. A COMPARATIVE STUDY IN SONG BIRDS

We consider a study in song birds (Serroyen *et al* 2005). Earlier, Van der Linden *et al* (2002) and Van Meir *et al* (2004) established a novel in-vivo magnetic resonance imaging

(MRI) approach to discern the functional characteristics of specific neuronal populations in a strongly connected brain circuitry, the so-called song control system in the song bird brain. The high vocal center (HVC), one of the major nuclei in this circuit, contains interneurons and two distinct types of neurons projecting respectively to the so-called nucleus robustus arcopallii (RA). This is graphically represented in Figure 3. Of interest is the effect of testosterone on the dynamics of Mn^{2+} accumulation in RA of female starling in individual birds injected with manganese in their HVC.

Precisely, ten first-year female starlings were caught in the wild during the winter before February and housed in two indoor cages on a stable 10–14 h light-dark light cycle, selected to maintain birds in a durable state of photo-sensitivity. All birds were studied by MRI for the first time between March 15 and April 30, 2001. One or two days after the first MRI measurement, the five treated birds were implanted subcutaneously in the neck region with a capsule of crystalline testosterone. The capsule was left empty for the five control birds. Birds were studied by MRI again five to six weeks after the treatment.

We consider two non-linear mixed model and a triplet of marginal models. The first non-linear mixed model we consider, in line with Serroyen *et al* (2005), takes the form:

$$SI_{ij}(\text{RA}) = \frac{(\phi_0 + \phi_1 G_i + f_i) t_{ij}^{\eta_0 + \eta_1 G_i + n_i}}{(\tau_0 + \tau_1 G_i + k_i)^{\eta_0 + \eta_1 G_i + n_i} + t_{ij}^{\eta_0 + \eta_1 G_i + n_i}} + \gamma_0 + \gamma_1 G_i + \varepsilon_{ij}. \quad (17)$$

where $SI_{ij}(\text{RA})$ is the measurement of MRI signal intensity for a region of interest (RA) at occasion j for bird i , G_i is an indicator for group membership (1 for testosterone treated birds and 0 otherwise), and t_{ij} is the measurement time, referring to the before *versus* after treatment epoch. The fixed-effects portion of the maximal signal intensity, SI_{\max} , is denoted by ϕ_{0i} for an untreated bird and $\phi_{0i} + \phi_{1i}$ for a treated one. Likewise for time required to reach 50% of this maximum (t_{50}): τ_{0i} and $\tau_{0i} + \tau_{1i}$, respectively. Finally, this holds true as well for the shape of the curve, which is governed by the parameters η_{0i} and $\eta_{0i} + \eta_{1i}$. The vector (f_i, n_i, k_i) is a bird-specific vector of random effects, assumed to follow a trivariate normal distribution with mean $\mathbf{0}$ and variance D . Finally, $\varepsilon_{ij} \sim N(0, \sigma^2)$.

The second random-effects model is (5), while for the marginal model, we choose models (9), with extensions (11) and (12). In all four cases, treatment effect is allowed for on the shape parameter $\beta_3 = \beta_{30} + \beta_{31}G_i$. Parameter estimates (standard errors) are reported in Table 3. Let us now discuss these.

In the first non-linear mixed model, using likelihood ratio tests, a number of non-significant fixed-effects and variance components were removed. The final model reported features one significant treatment effect parameter η_1 . For the second mixed model and for the marginal models, there is evidence that the third, most complicated marginal model, is necessary to adequately describe the within-bird correlation as well as the non-constant variance functions. It is interesting to see that in the simplest marginal model (9), the treatment has a significant effect, $p = 0.0354$, whereas it does neither in the second mixed model nor in the other two marginal models, with $p = 0.8733$, $p = 0.0922$, and $p = 0.0924$, respectively.

It is striking that the evidence for treatment effect is radically different between both random-effects models. An obvious choice is to extend the model with three random effects, one per β parameter. However, this still produces non-significant treatment effects, for each of these parameters separately, as well as overall (details not shown). Within the marginal family, it is important to adequately address the variance-covariance structure, because otherwise spurious effects will come to surface. This leaves us with the question as to why the first random-effects model is able to find a treatment effect that the marginal model does not. First note, as is clear from Figure 4, that the model fits well. The second random-effects model fits well too, though slightly less so than the first one (graph not shown). The same figure reveals that there is quite a bit of between-bird variability, both within the untreated group (birds 1–5) as the treated groups (birds 6–10). This heterogeneity is carefully separated out in the random-effects model. The marginal model, on the contrary, does not separate the between- and within-bird variability, hence a reduced power, in this case, to detect a treatment effect. Finally, note that one could consider in principle a transition model as well. This is omitted here for brevity.

5. CONCLUDING REMARKS

Exactly like in the linear and generalized linear cases, one can divide non-linear models for repeated measures and otherwise hierarchical data into marginal, random-effects, and conditional models. Somehow, this taxonomy has received less consideration, because focus has been, to a large extent spear-headed by pharmacokinetic/pharmacodynamic research, on random-effects models.

We have shown here, in a simple but instructive tree growth data example, that one can insightfully construct models within every one of the three families. For the aforementioned ubiquity of random-effects models, fitting is slightly less of a problem, thanks to the availability of dedicated tools, such as the SAS procedure NLMIXED and the *nlme* library in R and S-Plus. Nevertheless, reaching convergence can be challenging and some patience on the analyst's side may be called upon. Conditional models are usually of a transition type in longitudinal designs, in the sense that measurements are independent of each other, apart from a dependence on a usually small number of previous measurements. This implies that one can fit such models using software for non-linear models, fitted to cross-sectional data.

Marginal models are a little more challenging, because they require explicitly addressing the covariance structure. In fully specified models, specific parametric functions for the variance and correlation structure need to be considered, such as spatial Gaussian and spatial exponential structures. When the association is not of direct interest, generalized estimating equations type ideas can be used (Liang and Zeger 1986, Vonesh *et al* 2002). This would be possible for continuous and non-continuous data alike. Furthermore, nothing precludes the consideration of non-linear mean functions, neither.

In terms of complexity, random-effects models arguably take a position in between conditional and marginal models. This suggests that one may want to consider random-effects or conditional formulations, *even when the research question is in marginal terms*. In such a case, the marginal question would require the model to be marginalized. We have argued that this,

too, requires additional work. It is conceivable though, that marginalizing a conditional or random-effects model might still be computationally more economical than turning straight to a marginal model. In addition, it can be beneficial from a precision standpoint. This is clear in the song birds study, where the first random-effects model leads to significant treatment effects, whereas the other models do not. Evidently, it depends on the situation whether a marginalized model is sufficient to answer the research question at hand. For many non-linear functions, such a marginalization may be in terms of (graphical) summaries and not directly in terms of one or a few model parameters. Thus, in conclusion, the researcher must judge carefully the advantages and disadvantages of the various model families in terms of computational requirements on the one hand, and inferential goals on the other hand. Given the vast class of non-linear models, and their defying general rules, principles, and computational schemes, unlike with linear models, general guidelines are hard to provide. Decisions will often be *ad hoc*, ideally supplemented with experience.

Note that, by presenting the taxonomy and providing illustrations in terms of both the orange tree data and the song birds study, the complexity of fully non-linear model is seen not to excessively go beyond that of (generalized) linear models for hierarchical data. Non-linear hierarchical models are very flexible and have been implemented in a variety of standard software tools. Arguably, therefore, they ought to be considered by the practicing statistician more than they currently do.

Practically, the methods were implemented using the SAS procedure NLMIXED and a user-written function in R. In principle, it is possible to write user-defined programs in any sufficiently generic statistical programming package with decent matrix manipulation facilities. The SAS and R code is available from the first author's web site and upon simple request.

SAS, R, and GAUSS programs for the models fitted in this paper can be found in the web based supplementary materials.

References

- Arnold, B.C., Castillo, E. and Sarabia, J.M. (1992), *Conditionally Specified Distributions*, New York: Springer.
- Breslow, N.E. and Clayton, D.G. (1993), "Approximate inference in generalized linear mixed models," *Journal of the American Statistical Association*, 88, 9–25.
- Cressie, N.A.C. (1991), *Statistics for Spatial Data*, New York: John Wiley & Sons.
- Davidian, M. and Giltinan, D.M. (1995), *Nonlinear Models for Repeated Measurement Data*, London: Chapman & Hall.
- Diggle, P.J., Heagerty, P.J., Liang, K.-Y., and Zeger, S.L. (2002), *Analysis of Longitudinal Data (2nd ed.)*, Oxford: Clarendon Press.
- Draper, N.R., and Smith, H. (1998), *Applied Regression Analysis (3rd ed.)*, New York: John Wiley & Sons.
- Liang, K.-Y. and Zeger, S.L. (1986), "Longitudinal data analysis using generalized linear models," *Biometrika* 73, 13–22.
- Lee, Y. and Nelder, J.A. (1996), "Hierarchical generalized linear models (with discussion)," *Journal of the Royal Statistical Society, Series B* 58, 619–678.
- Lee, Y. and Nelder, J.A. (2001), "Hierarchical generalized linear models: a synthesis of generalized linear models, random-effect models and structured dispersions," *Biometrika* 88, 987–1006.
- Lee, Y., and Nelder, J.A. (2003), "Extended-REML estimators," *Journal of Applied Statistics* 30, 845–856.
- Lee, Y., and Nelder, J.A. (2004), "Conditional and marginal models: another view (with discussion)," *Statistical Science* 19, 219–238.

- Lindstrom M.J and Bates, D.M. (1990), "Nonlinear mixed effects models for repeated measures data," *Biometrics* 46, 673–687.
- Molenberghs, G. and Verbeke, G. (2007), "Likelihood ratio, score, and Wald tests in a constrained parameter space," *The American Statistician* 61, 1–6.
- Molenberghs, G. and Verbeke, G. (2005), *Models for Discrete Longitudinal Data*, New York: Springer.
- Noh, M. and Lee, Y. (2008), "Hierarchical-likelihood approach for nonlinear mixed-effects models. *Computational Statistics Data Analysis* 52, 3517–3527.
- Pinheiro, J.C. and Bates, D.M. (2000), *Mixed-Effects Models in S and S-Plus*, New York: Springer.
- Self, S.G. and Liang, K.Y. (1987), "Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions," *Journal of the American Statistical Association* 82, 605–610.
- Serroyen, J., Molenberghs, G., Verhoye, M., Van Meir, V., and Van der Linden, A. (2005), "Dynamic manganese enhanced (DME) MRI signal intensity processing based on nonlinear mixed modeling to study changes in neuronal activity," *Journal of the Agricultural, Biological, and Environmental Statistics*, 10, 170–183.
- Sheiner, L.B. and Beal, S.L. (1980), "Evaluation of methods for estimating population pharmacokinetic parameters. I. Michaelis-Menten model: Routine clinical pharmacokinetics data," *Journal of Pharmacokinetics and Biopharmaceutics* 8, 559–571.
- Stram, D.O. and Lee, J.W. (1994), "Variance components testing in the longitudinal mixed effects model," *Biometrics* 50, 1171–1177.
- Stram, D.A. and Lee, J.W. (1995), "Correction to: Variance components testing in the longitudinal mixed effects model," *Biometrics* 51, 1196.

- Van der Linden, A., Verhoye, A., Van Meir, V., Tindemans, I., Eens, M., Absil, P., and Balthazart, J. (2002), "In vivo manganese-enhanced magnetic resonance imaging reveals connections and functional properties of the songbird vocal control system," *Neuroscience*, 112, 467–474.
- Van Meir, V., Verhoye, M., Absil, P., Eens, M., Balthazart, J., and Van der Linden, A. (2004), "Differential effects of testosterone on neuronal populations and their connections in a sensorimotor brain nucleus controlling song production in songbirds: a manganese enhanced-magnetic resonance imaging study," *Neuroimage*, 21, 914–923.
- Verbeke, G. and Molenberghs, G. (2000), *Linear Mixed Models for Longitudinal Data*, New York: Springer.
- Verbeke, G. and Molenberghs, G. (2003), "The use of score tests for inference on variance components," *Biometrics* 59, 254–262.
- Vonesh, E.F. (1996), "A note on the use of Laplace's approximation for non-linear mixed-effects models," *Biometrika* 83, 447–452.
- Vonesh, E.F., Wang, H., Nie, L., Majumdar, D. (2002), "Conditional second-order generalized estimating equations for generalized linear and nonlinear mixed-effects models," *Journal of the American Statistical Association* 97, 271–283.

Table 1: *Orange Tree Data. Measurements in mm of trunk circumference.*

Day	Trunk circumference				
	Tree 1	Tree 2	Tree 3	Tree 4	Tree 5
118	30	33	30	32	30
484	58	69	51	62	49
664	87	111	75	112	81
1004	115	156	108	167	125
1231	120	172	115	179	142
1372	142	203	139	209	174
1582	145	203	140	214	177

Table 2: *Orange Tree Data. Parameter estimates (standard errors) for various models.*

Effect	Par.	R.E. (3)	R.E. (5)	R.E. (6)
Asymptote	β_1	1.921(0.157)	1.913(0.167)	1.940(0.164)
Half-growth	β_2	7.279(0.353)	7.170(0.381)	7.417(0.587)
Shape parameter	β_3	3.481(0.271)	3.465(0.258)	3.583(0.416)
R.E. variance	d_{11}	0.100(0.065)	0.120(0.079)	0.080(0.002)
R.E. variance	d_{22}		0.147(0.303)	0.0002(0.0022)
R.E. covariance	d_{12}		0.102(0.118)	0.004(0.021)
Residual variance	σ^2	0.006(0.002)	0.006(0.002)	0.006(0.002)
Effect	Par.	marg. (9)	marg. (11)	marg. (12)
Asymptote	β_1	1.910(0.144)	1.924(0.183)	2.131(0.327)
Half-growth	β_2	7.246(0.618)	7.250(0.797)	9.701(2.811)
Shape parameter	β_3	3.461(0.497)	3.499(0.422)	4.785(1.087)
Serial variance	τ^2	0.042(0.022)		
Rate of exp. decay	ϕ	24.72(17.65)	6.278(2.789)	12.011(10.515)
Residual variance	σ^2	0.005(0.002)		
Constant var. term	θ_0		-5.090(0.757)	-9.945(0.883)
Linear var. term	θ_1		0.168(0.064)	1.413(0.245)
Quadr. var. term	θ_2			-0.060(0.012)
Effect	Par.	trans. (14)	trans. (16)	
Asymptote	β_1	0.421(0.082)	2.079(0.089)	
Half-growth	β_2	-0.419(1.778)	-8.386(3.729)	
Shape parameter	β_3	1.782(0.953)	-6.668(2.402)	
Renewal parameter	γ	0.841(0.059)	0.034(0.006)	
Residual variance	σ^2	0.014(0.003)	0.012(0.003)	

Table 3: Song Bird Data. Parameter estimates (standard errors) for two non-linear mixed models and three marginal models, fitted to $SI_{ij}(\text{RA})$ at the second period.

Effect	Par.	R.E. (17)	Effect	Par.	R.E. (5)
	ϕ_0	0.453(0.048)	Asymptote	β_1	0.3864(0.042)
	η_0	2.183(0.080)	Half-growth	β_2	2.596(0.115)
	η_1	0.429(0.106)	Shape par.	β_{30}	0.824(0.032)
	τ_0	2.848(0.176)	Treatm. eff.	β_{31}	0.0066(0.0404)
Var(f_i)	d_{11}	0.023(0.010)	R.E. var.	d_{11}	0.018(0.008)
Var(k_i)	d_{22}	0.288(0.134)	R.E. var.	d_{22}	0.125(0.058)
			R.E. covar.	d_{12}	0.0061(0.0153)
Var(ε_{ij})	σ^2	1.9E-4(1.7E-5)	Resid. var.	σ^2	3.5E-3(3.2E-5)
Effect	Par.	marg. (9)	marg. (11)	marg. (12)	
Asymptote	β_1	0.453(0.035)	0.437(0.045)	0.445(0.042)	
Half-growth	β_2	2.056(0.229)	2.172(0.256)	2.173(0.243)	
Shape par.	β_{30}	2.826(0.195)	2.749(0.236)	2.785(0.237)	
Treatm. eff.	β_{31}	0.611(0.291)	0.566(0.336)	0.549(0.326)	
Serial var.	τ^2	0.088(0.017)			
Rate exp. decay	ϕ	17.23(7.181)	8.310(3.101)	8.642(3.243)	
Resid. var.	σ^2	0.008(0.002)			
Const. var. term	θ_0		-5.746(0.437)	-6.169(0.470)	
Lin. var. term	θ_1		0.193(0.055)	0.545(0.192)	
Quadr. var. term	θ_2			-0.051(0.027)	

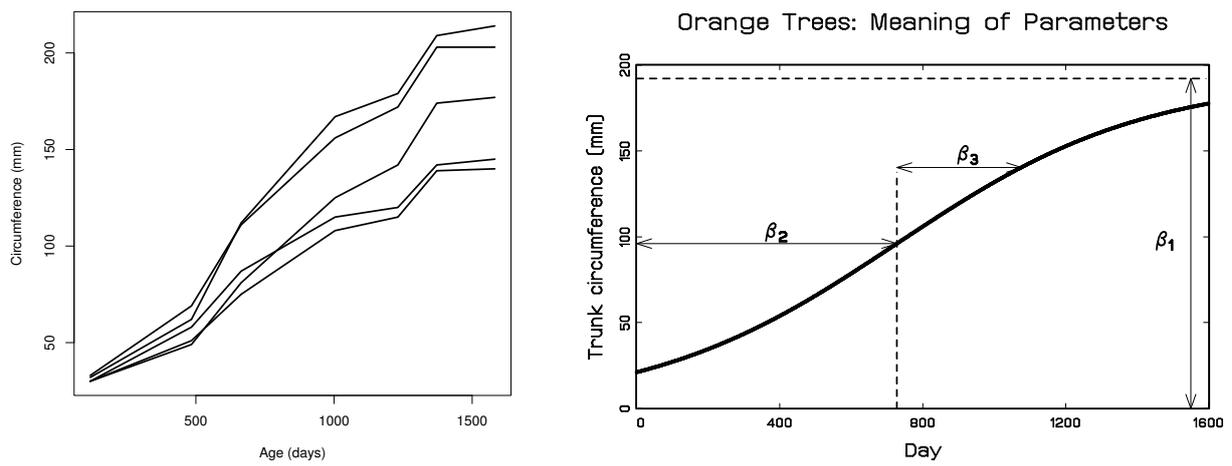


Figure 1: Orange Tree Data. Left hand panel: Growth curves of trunk circumference for each of the five trees. Right hand panel: Interpretation of the fixed-effects parameters in the random-effects model.

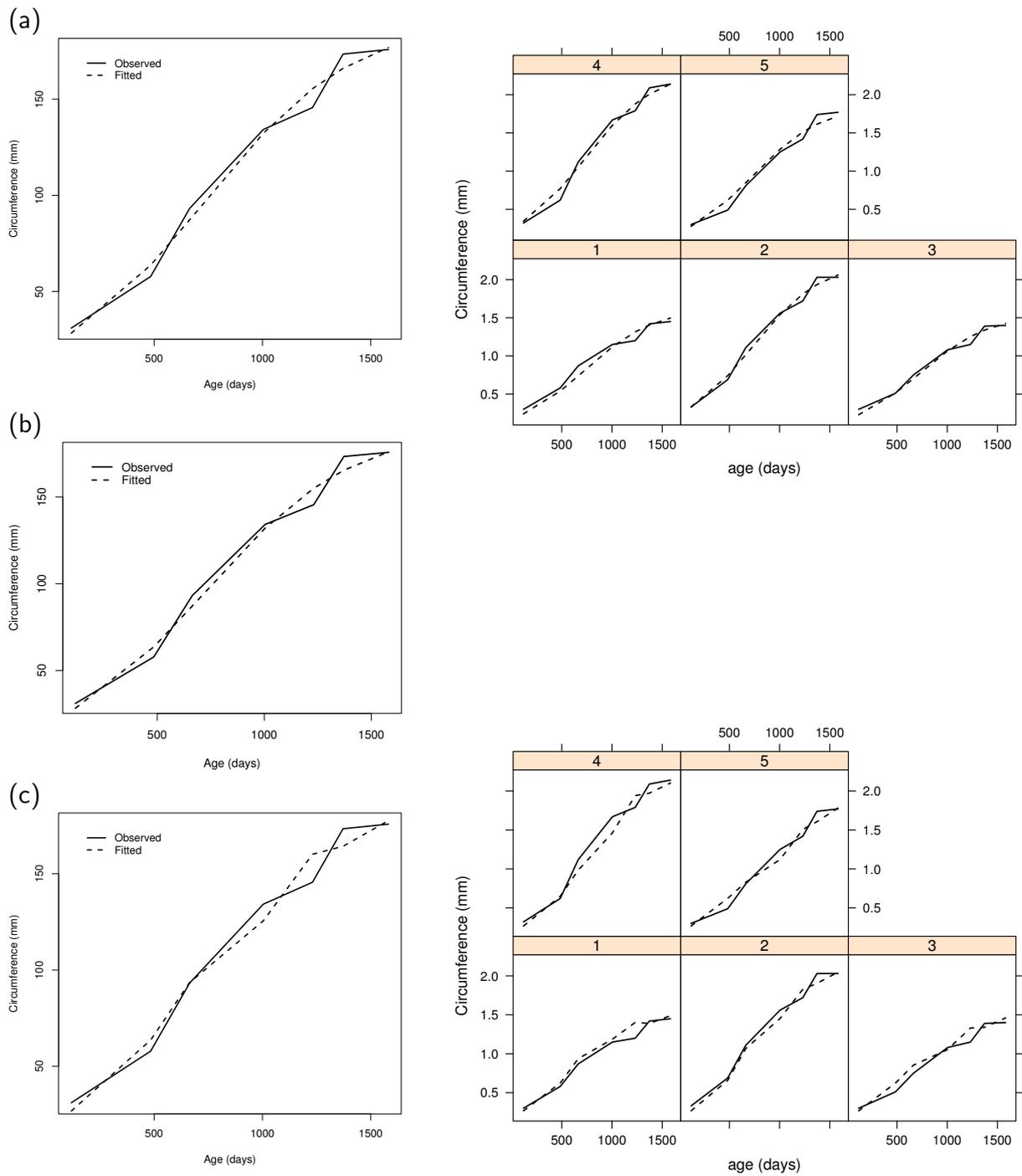


Figure 2: Orange Tree Data. Left hand panels: Plots of observed and fitted mean profiles. Right hand sides: Observed (solid line) and fitted (dashed line) individual profiles. (a) Random-effects model (3); (b) Marginal model (9); (c) Transition model (14).

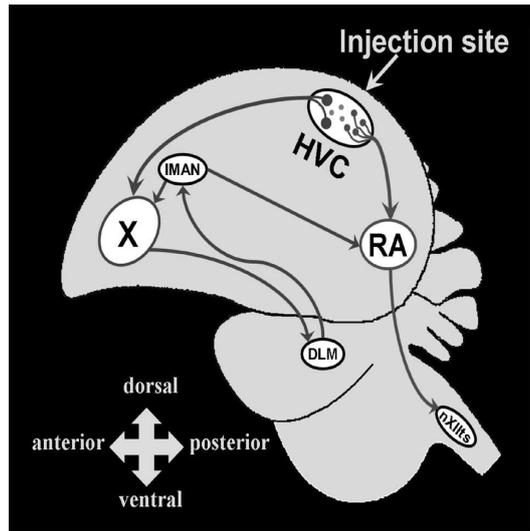


Figure 3: *Song Bird Data.* Schematic representation of song control nuclei in the songbird brain

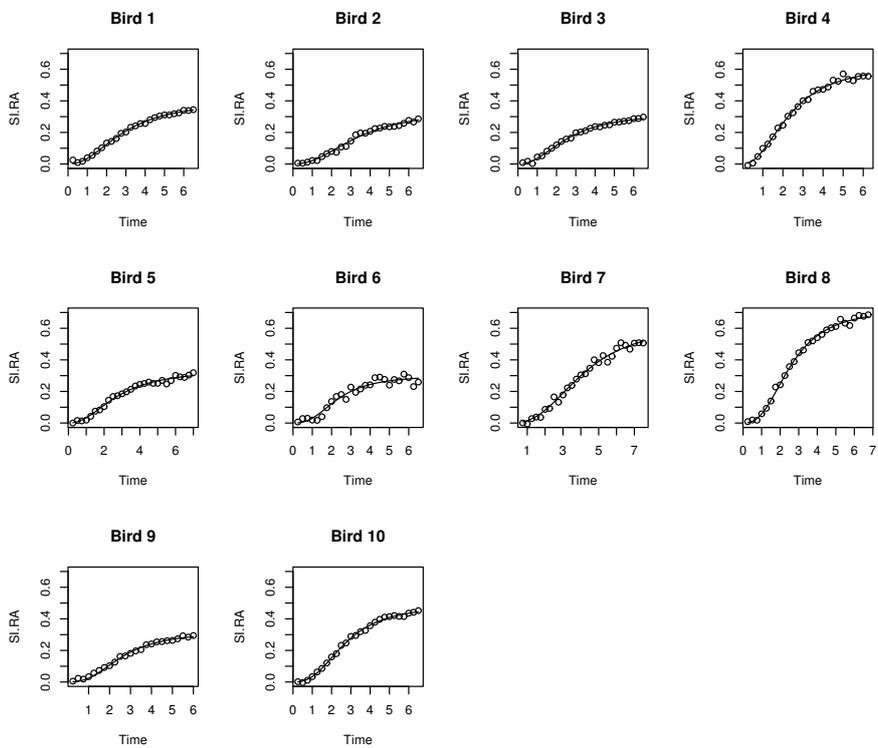


Figure 4: *Song Bird Data.* Fitted curves for $SI_{ij}(RA)$ at the second period, for each individual bird separately.