Transnational University Limburg is a unique collaboration of two universities in two countries: the University of Hasselt and Maastricht University.

UHASSELT
KNOWLEDGE IN ACTION

www.uhasselt.be
Universiteit Hasselt
Campus Hasselt:
Martelarenlaan 42 | 3500 Hasselt
Campus Diepenbeek:
Agoralaan Gebouw D | 3590 Diepenbeek

UHASSELT
KNOWLEDGE IN ACTION

Maastricht University

Faculty of Sciences
School for Information Technology
Master of Statistics and Data Science

Master's thesis

Statistical and Numerical Challenges in Fitting Non-Linear Mixed-effects Models

Christine Katabarwa

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics and Data Science, specialization Biostatistics

SUPERVISOR :
Prof. dr. Geert MOLENBERGHS
Prof. dr. Iuliu Sorin POP

2024
2025

Maastricht University

Faculty of Sciences
School for Information Technology
Master of Statistics and Data Science

Master's thesis

Statistical and Numerical Challenges in Fitting Non-Linear Mixed-effects Models

Christine Katabarwa

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics and Data Science, specialization Biostatistics

SUPERVISOR :
Prof. dr. Geert MOLENBERGHS
Prof. dr. Iuliu Sorin POP

Acknowledgments

First and foremost, I would like to express my deepest gratitude to my supervisors, Prof. Dr. Geert Molenberghs and Prof. Dr. Sorin Pop, for their invaluable guidance, continuous support, and patient mentorship throughout this research journey. Their expertise and insights were fundamental to the development and completion of this thesis.

I am also grateful to the staff of Hasselt University, especially those who taught and inspired me during the Master's in Biostatistics program. The skills and knowledge I gained here have been pivotal.

On a personal note, I would like to thank my family and friends for their unwavering belief in me, especially during the most challenging moments of this academic pursuit.

Finally, I am grateful to God for His grace, wisdom, and strength that sustained me throughout this journey.

Abstract

Non-linear mixed-effects models offer a flexible means for analyzing longitudinal data by allowing fixed and random effects to enter a model nonlinearly. Despite their flexibility and interpretability, these models present substantial statistical and numerical challenges. The key challenge is the lack of explicit antiderivatives required for the marginal likelihood in parameter estimation, which necessitates the use of numerical integration techniques. This thesis explores these challenges by fitting various non-linear mixed-models to longitudinal data from the songbird brain regions HVC, RA, and a so-called $Area\_X$.

Model fitting was conducted using SAS PROC NLMIXED, systematically evaluating estimation procedures across different integral approximations (Gaussian and Laplace) and varying the number and adaptiveness of quadrature points. The integration techniques were combined with various optimization techniques which happen to be modifications of the Newton-Raphson method.

The findings show that different combinations of integration and optimization techniques have an effect on the parameter estimates when good starting values are chosen. These results offer practical guidance for researchers modeling complex biological processes and contribute to a better understanding of both the statistical and biological aspects of how signal intensity changes in songbirds.

Keywords: Non-linear mixed-effects models, integration methods, optimization techniques, parameter estimation.

1 Introduction

Scientific research is often focused on understanding how things change over time. We frequently deal with data gathered from the same units at various times, whether we are studying learning and memory, tracking the growth of a population, or tracking the effects of a new medication. Rarely do these repeated measurements, or longitudinal data, exhibit straightforward, linear patterns. Rather, real-world processes are frequently non-linear, and each unit may exhibit a distinct pattern of change [1], [2].

Non-linear mixed models (NLMM) have become more popular because of their flexibility in handling such complex data collected from various fields, such as economics, pharmacokinetics and ecology. These models extend linear mixed-effects frameworks by allowing parameters to enter the model nonlinearly, while accounting for both population-level effects (fixed effects) and individual-specific deviations (random effects) that cannot be adequately described by linear models [1], [2].

However, with this flexibility comes a set of unique challenges. Fitting an NLMM to data is much more complicated than fitting a linear model. The main difficulty lies in estimating the model parameters: because the model is non-linear in the random effects. Estimation of NLMM parameters is typically performed using maximum likelihood estimation (MLE) which works by selecting the parameters whose estimated values maximize the likelihood function. Maximization of this likelihood requires integrating over the random effects for which unlike in the linear case, non-linear mixed models usually have no explicit antiderivates. The likelihood function that we need to maximize involves integrals that are usually impossible to evaluate exactly [1]. To overcome this, statisticians have developed a range of approximation methods. Early approaches, like first-order (FO) and conditional first-order (CFO) linearizations, use Taylor expansions to simplify the problem, but can introduce bias, especially when the data are sparse or the nonlinearity is strong [3], [4]. More accurate alternatives, such as the Laplace approximation and Gaussian quadrature, have been developed to better approximate these difficult integrals, though they require more computational resources and careful tuning [1], [5]. These computational challenges are compounded by the need for robust optimization algorithms, such as Newton-Raphson or quasi-Newton methods, to maximize the approximated likelihood, which can be sensitive to starting values and prone to convergence issues in complex models [2].

Modern statistical software, such as SAS PROC NLMIXED, R's `nlme` package, and Stata's `menl` command, provide tools for fitting NLMMs using these approximation techniques [6], [7]. These procedures allow researchers to choose among different integral approximation methods and optimization algorithms, and to assess the impact of these choices on parameter estimates and model fit. For example, the number and placement of quadrature points

in Gaussian quadrature, or the starting values used in optimization, can affect both the accuracy of parameter estimates and the time it takes to fit the model [1].

The goal of this thesis is to explore the statistical and computational challenges that arise when fitting non-linear mixed-effects models, with a particular focus on the impact of different integral approximation methods and optimization strategies. By systematically evaluating these approaches, this work aims to provide practical guidance for researchers analyzing complex longitudinal data.

2 Motivating study

This thesis is inspired by a study wherein 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 hour light-dark light cycle , selected to maintain birds in a durable state of photosensitivity. All birds were studied with magnetic resonance imaging (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 with a capsule of crystalline testosterone subcutaneously in the neck region. The capsule was left empty for the five control birds. Five to six weeks after treatment, the birds were studied again by MRI and measurements of their signal intensity were taken in three areas that is, the high vocal center (HVC), one of the major nuclei in this circuit which contains interneurons and two distinct types of neurons projecting respectively to the so-called nucleus robustus archistriatalis (RA) or to area_X [8]. This is graphically represented in Figure 1.

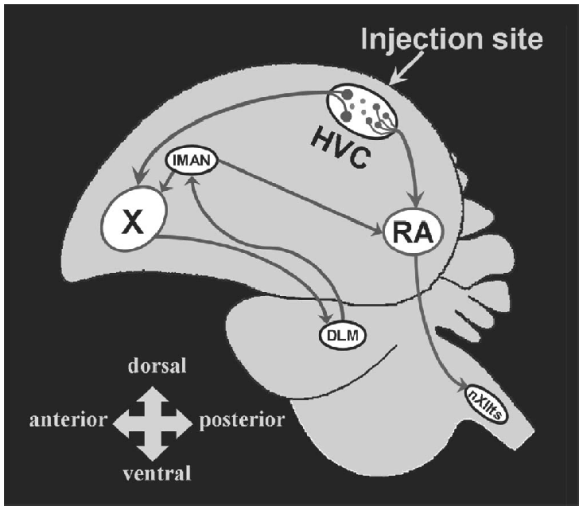

Figure 1: Schematic representation of song control nuclei in the songbird brain [8]

.

The data consists of ten birds each with about 30 repeated measurements of signal intensity. The outcomes of interest are therefore the signal intensity(SI) of RA (SI_RA), $area\_X$ (SI_area_X) and HVC (SI_HVC). The times at which the measurements were taken are also included in the dataset. Individual profiles of the birds for the three outcomes SI_HVC, SI_RA and SI_Area_X were plotted (2, 3, 4), and they all exhibit non-linear behavior thereby calling for the employment of non-linear models. Furthermore, there is clearly between-bird variability indicated by the differences in rise and shapes of the different curves for the different birds, this calls for the use of random effects to account for these differences.

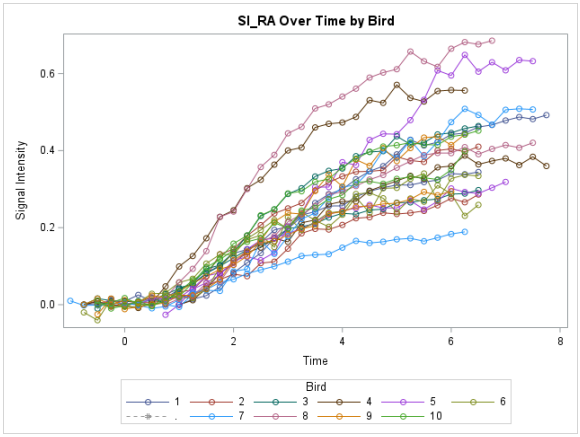

Figure 2: Individual-bird profiles for the signal intensity in the RA region

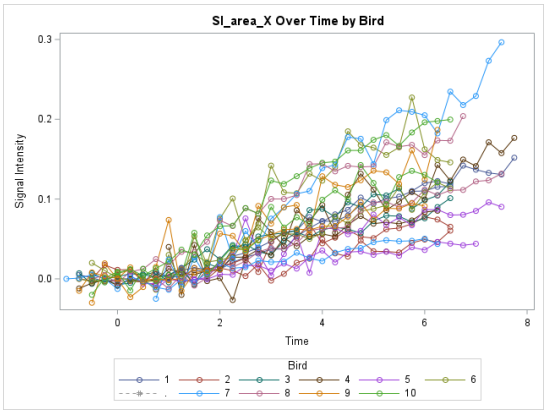

Figure 3: Individual-bird profiles for the signal intensity in *area_X*

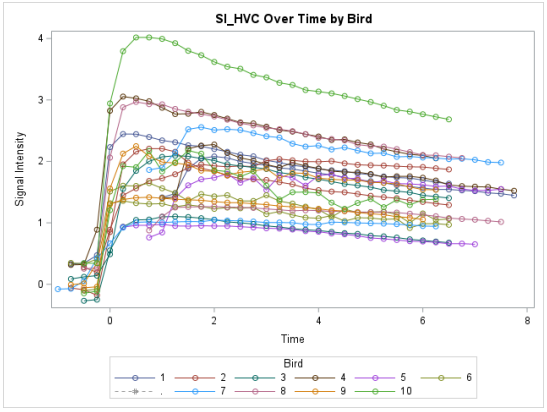

Figure 4: Individual-bird profiles for the signal intensity in the high vocal center

3 Methodology

While established non-linear modeling approaches exist for the RA and *Area_X* signal intensities, previous applications of these models have not incorporated the correlational structure present within individual birds. The situation differs for measurements of the signal intensity in the HVC region, where suitable non-linear functional forms have not been well-established in the literature. Visual analysis of our data patterns indicates that the underlying relationships would not be adequately captured by conventional linear modeling techniques. We therefore look into two different modeling approaches to tackle this problem. We start by looking at fractional polynomial models, which increase the range of curve shapes available and provide more flexibility than standard polynomial functions. Second, considering the conceptual similarities between our HVC signal intensity problem and pharmacokinetic processes, we investigate the suitability of two-compartment (bi-exponential) pharmacokinetic models. Incorporating the within-subject correlation structure built-in our repeated measures design, these modeling techniques seek to appropriately account for the non-linear nature of the data [8].
All the models will be fitted using SAS procedure NLMIXED, using several integration techniques such as adaptive Gaussian quadrature, non-adaptive Gaussian quadrature and Laplace approximation.

As stated in [8], random-effects models can be fitted by maximum likelihood estimation (MLE) which involves maximization of the marginal likelihood, obtained by integration out the random effects. Each subject i has a contribution to the likelihood such as:

$$f_i(y_i|\beta, D, \phi) = \int_{-\infty}^{\infty} \prod_{j=1}^{n_i} f_{ij}(y_{ij}|b_i, \beta, \phi) f(b_i|D) db_i, \tag{1}$$

y_{ij} is the j^{th} outcome measured for subject i, $i = 1, ..., N$, $j = 1, ..., n_i$ and y_i is the n_i dimensional vector of all measurements available for subject i. ϕ is a scalar parameter and β is a p-dimensional vector of unknown fixed regression coefficients. f is a real integrable function, $f(b_i|D)$ is the density of the $N(0, D)$ distribution of the random effects b_i and D is the random-effects covariance matrix.

From (1) the likelihood for β, D and ϕ is computed as

$$\begin{aligned} L(\beta, D, \phi) &= \prod_{i=1}^{N} f_i(y_i|\beta, D, \phi) \\ &= \prod_{i=1}^{N} \int_{-\infty}^{\infty} \prod_{j=1}^{n_i} f_{ij}(y_{ij}|b_i, \beta, \phi) f(b_i|D) db_i. \end{aligned} \tag{2}$$

The primary challenge in maximizing (2) is that there are N integrals over the q-dimensional

random effects b_i. Unlike in linear mixed models, where these integrals can be evaluated exactly due to the availability of explicit antiderivatives, non-linear mixed models typically require these integrals to be approximated numerically, as explicit antiderivatives are generally unavailable. In such cases the definite integrals required for the marginal likelihood can be approximated numerically, hence the need for numerical integration techniques.
One subdivision of numerical approximation is approximation of the integrand (replacing the complex functions inside the integral with simpler functions). The main point of approximating the integrand is to obtain integrals which are easier to integrate so that explicit antiderivatives can be obtained, thereby making it possible to maximize the approximated likelihood. The numerical integration methods we will use include adaptive Gaussian quadrature, non-adaptive Gaussian quadrature and Laplace approximation.
According to [9], Gaussian quadrature is a technique to approximate integrals that are centered about the empirical Bayesian estimates of the random effects. The number of quadrature points can be selected given a desired standard of accuracy. Adaptive and non-adaptive Gaussian quadrature is designed for the approximating integral of the form below:

$$\int_{-\infty}^{\infty} f(z)\phi(z)dz \tag{3}$$

In (3), $f(z)$ is a known, smooth function and $\phi(z)$ is the density of the standard normal distribution.

$$\int_{-\infty}^{\infty} f(z)\phi(z)dz \approx \sum_{q=1}^{Q} w_q f(z_q). \tag{4}$$

In (4), Q is the number of quadrature points and a larger Q results in a more accurate approximation. w_q is the quadrature weight chosen appropriately, z_q are quadrature points (nodes) and solutions to the Q^{th} order Hermite polynomial. Here, the quadrature points z_q are chosen based on $\phi(z)$, independent of the function $f(z)$ in the integrand [8]. Depending on the support of $f(z)$, the nodes, z_q will or will not lie in the region of interest, this is a potential short coming of this approach.
Following equation (2), the likelihood contribution for subject i is given by:

$$f_i(y_i|\beta, D, \phi) = \int_{-\infty}^{\infty} \prod_{j=1}^{n_i} f_{ij}(y_{ij}|\beta, b_i, D, \phi) f(b_i) db_i \tag{5}$$

In non-adaptive Gaussian quadrature, $\int_a^b f(z)\phi(z)dz$ is approximated directly by the weighted sum in equation (4). It does so by choosing nodes (quadrature points) in areas of high density and when $\phi(z)$ is continuous, the quadrature rule is exact if $f(z)$ is a polynomial of up to $2Q-1$. As mentioned earlier, more accurate approximation is obtained with a higher number of quadrature points so their number can be increased or decreased to get the desired level of accuracy but this could potentially increase the computational

burden.

In this study, the number of quadrature points Q, z_q and weights w_q will be found using the Gauss-Hermite quadrature (as used in SAS). This approximation can be viewed as an extension of the classical Gaussian quadrature to approximate integrals of the form

$$\int_{-\infty}^{\infty} \exp(-z^2) f(z) dz \approx \sum_{q=1}^{Q} \tilde{w}_q f(z_q) \tag{6}$$

where with respect to the Gauss-Hermite quadrature, z_q are the roots of the Hermite polynomials. The weights are given by:

$$\tilde{w_q} = \frac{2^{Q-1} Q! \sqrt{\pi}}{Q^2 [H_{Q-1}(z_q)]^2}. \tag{7}$$

It is not uncommon that some quadrature points z_q may lie outside the region of interest when analyzing non-linear longitudinal data that is why the classical Gaussian quadrature is adapted to ensure that more quadrature points lie in the region of interest. This is the origin of adaptive Gaussian quadrature in which the nodes can be rescaled and centered so that $f(z)\phi(z)$ is normally distributed. Adaptive Gaussian quadrature for the entire integral over b_i centers the integral at the empirical Bayes estimate of b_i that minimizes

$$-\log[f(y_i|x_i, b_i, \phi) p(b_i|D)] \tag{8}$$

in which ϕ and D are equal to their current estimates. More insights about Gauss-Hermite quadrature can be found in [10].
The general idea is that the numerical approximation is applied to the likelihood contribution of each subject for the N subjects in the study at hand. As mentioned earlier, better approximation of the N integrals in the likelihood is achieved with higher values of Q. Compared to classical Gaussian quadrature, adaptive Gaussian quadrature is more accurate as it ensures that there are more quadrature points in the region of interest and it frequently uses fewer points to attain the same level of precision. By dynamically updating the quadrature points and weights at each iteration of parameter estimation, adaptive Gaussian quadrature improves the accuracy of approximating integrals involving random effects. In particular, the locations and weights of the quadrature points are scaled and centered based on the mode and curvature of the quadrature points, which are determined by the current estimates of the model parameters, such as the dispersion parameter (ϕ), the random effects covariance matrix (D), and the fixed effects (β) [1]. Although this method offers a more precise approximation than classical (non-adaptive) Gaussian quadrature, it is computationally more intensive, since the quadrature scheme must be recalculated at each step of the optimization process. It is also important to note that Laplace approximation is equivalent to adaptive Gaussian quadrature with one node ($Q = 1, z_q = 0$) as this is how it

will be applied when fitting the various models in SAS. In depth explanations about these techniques can be found in [8], pages 273-275 and in [11], pages 379-405.

After the integration methods mentioned above have yielded integrals for which we have explicitly found antiderivatives (the likelihood has been approximated), we will employ various optimization techniques such as Newton-Raphson with line search, Newton-Raphson with ridging and Quasi-Newton optimization so as to maximize the likelihood function.
The Newton-Raphson algorithm is an iterative procedure that can be used to calculate MLEs. The basic idea behind the algorithm is the following. First, construct a quadratic approximation to the function of interest around some initial parameter value (hopefully close to the MLE). Next, adjust the parameter value to that which maximizes the quadratic approximation. This procedure is iterated until the parameter values stabilize [12]. However, this convergence requires that the initial guess be close enough to the solution. Since we do not know this ideal initial guess, we have to use some alternative methods (modifications), which converge slower but guaranteed to get us closer to the solution. Once the approximation is close enough to the solution, then we can switch to a genuine Newton scheme, and in very few steps an excellent approximation is obtained. These modifications can include the line search, ridging and other Newton-like methods (Quasi-Newton), these are designed to enhance stability and help the algorithm skip problematic regions of the parameter space.
The Netwon-Raphson is highly sensitive to starting values so poorly selected starting values can lead to failure of convergence, this is something we will look out for as we fit our non-linear mixed-effects models. It is therefore advisable to use Newton-Raphson in at least the final iterations of any non-linear algorithm to take advantage of its fast local convergence, but it will have to be modified in order to converge globally [13].

The non-convergence of the classical Newton-Raphson method can be eliminated by Newton's method with line search in which the Newton correction is used to generate a direction of search [14]. In the Newton-Raphson algorithm parameter estimates at each stage are updated by utilizing the gradient (first derivative) and Hessian (second derivative matrix) of the log-likelihood function. However, when a line search procedure is added to the Newton-Raphson procedure the step size along the Newton direction is adaptively chosen at each iteration by calculating the log-likelihood at several different candidate step sizes and choosing the one that increased the log-likelihood the most. Therefore, Newton-Raphson with line search combines the fast local convergence of Newton's method with the robustness of line search strategies, improving the stability and convergence of non-linear optimization algorithms.

Newton-Raphson ridge is another optimization technique used to stabilize the Newton-Raphson algorithm by adding a small positive constant to the diagonal elements of the

Hessian (second derivative) matrix. This is meant to prevent numerical instability during parameter estimation, especially in the presence of multicollinearity or near-singular matrices which are common challenges in non-linear mixed models due to their complexity. This modification will ensure that the Hessian matrix is invertible and also that the parameter estimates are not out of control [15], [16], [17].

Sometimes the computation of the Hessian matrix is computationally expensive, in such scenarios the (dual) Quasi-Newton optimization proves to be more efficient. The Quasi-Newton approach does not require the computation of the Hessian matrix at each step but rather uses its approximation unlike the line search and ridge modifications described earlier [18]. This approach significantly reduces computational burden, especially for complex models (such as ours) or large datasets, while still ensuring efficient and reliable convergence [19].

3.1 Model of Van der Linden et al, (2002)

3.1.1 Model for SI_RA

In [20], the following parametric shape for a single bird's profile was employed:

$$SI_{ij}(RA) = \frac{(\phi_{0i} + \phi_{1i}G_i)t_{ij}^{\eta_{0i}+\eta_{i1}G_i}}{(\tau_{0i} + \tau_{1i}G_i)^{\eta_{0i}+\eta_{1i}G_i} + t_{ij}^{\eta_{0i}+\eta_{1i}G_i}} + \gamma_{0i} + \gamma_1 G_{1i} + \varepsilon_{ij}. \tag{9}$$

From (9), $SI_{ij}(RA)$ is the measurement for bird i at time j, G_i is an indicator of group membership (1 for testosterone-treated birds and 0 for control birds), and t_{ij} is the time measurement. The maximal signal intensity, sometimes termed SI_{max}, is denoted by ϕ_{0i} for an untreated bird and $\phi_{0i} + \phi_{1i}$ for a treated one. The time required to reach 50% of this maximum (T_{50}) is τ_{0i} and $\tau_{0i} + \tau_{1i}$, respectively. The shape of the curve is determined by the parameters η_{0i} and τ_{1i}. Finally, ε_{ij} is a measurement error term, typically assumed to follow a normal distribution. The genesis of this model is rooted in knowledge about Mn axonal transport and changes induced in the bird's brain caused by testosterone treatment. More details can be found in [21], [20], and [22].
In [20], they utilized Model (9) to each bird under investigation and then proceeded to apply ANOVA (analysis of variance) to the estimates parameters. This approach rests on the assumption that the measurements within the same bad are not correlated. In order to account for this within bird correlation, we introduce the above model into the mixed-effects framework in which the parameters are split into fixed and random effects [8].
In this model, all parameters were assumed to be different from one songbird to another, since the non-linear model was fitted to each bird separately. We are now in a position to analyze all data together, separating out averaged (fixed) effects from bird-specific (random) effects, using the replacements below:

$$\phi_{0i} + \phi_{1i}G_i \longrightarrow \phi_0 + \phi_1 G_i + f_i, \tag{10}$$

$$\eta_{0i} + \eta_{1i}G_i \longrightarrow \eta_0 + \eta_1 G_i + n_i, \tag{11}$$

$$\tau_{0i} + \tau_{1i}G_i \longrightarrow \tau_0 + \tau_1 G_i + t_i, \tag{12}$$

The ϕ, η, and τ parameters are fixed effects, while the vector (f_i, n_i, t_i) is a bird-specific vector of random effects, assumed to follow a trivariate normal distribution with mean $\mathbf{0}$ and variance D. Combining Model (9) with replacements (10) to (12), we obtain:

$$SI_{ij}(RA) = \frac{(\phi_0 + \phi_1 G_i + f_i)t_{ij}^{\eta_0+\eta_1 G_i+n_i}}{(\tau_0 + \tau_1 G_i + t_i)^{\eta_0+\eta_1 G_i+n_i} + t_{ij}^{\eta_0+\eta_1 G_i+n_i}} + \gamma_0 + \gamma_1 G_i + \varepsilon_{ij}. \tag{13}$$

The parameters have the same meaning they had in Model (9). The residual error terms ε_{ij} are assumed to be mutually independent and independent from the the random effects, and to be drawn from a normal distribution, $N(0, \sigma^2)$.

We will therefore build a model for the second period, where treatment has been administered because no treatment effect would be expected in the first period (measurements are taken prior to treatment).
Just as applied by [23] we will use Model (13) to analyze the data for which the general form has 8 fixed-effects parameters and 7 variance components (3 variances in D, 3 covariances in D, and σ^2). The model obtained from backward selection is as follows:

$$SI_{ij}(RA) = \frac{(\phi_0 + f_i)t_{ij}^{\eta_0+\eta_1 G_i}}{(\tau_0 + t_i)^{\eta_0+\eta_1 G_i} + t_{ij}^{\eta_0+\eta_1 G_i}} + \varepsilon_{ij}. \tag{14}$$

The same model will be fitted for the signal intensity of $area\_X$.

3.2 Fractional polynomials

This model will be fitted for the signal intensity of the high vocal center (HVC) up to the cubic term in time (fractional polynomials will be applied to the time covariate) and it's formulated as below;

$$\begin{aligned} SI\_HVC_{ij} =& (\beta_0 + \beta_{og} * G_i) \\ &+ \beta_1 + \beta_{1g} * G_i) * t_{ij} \\ &+ (\beta_2 + \beta_{2g} * G_i) * t_{ij}^2 \\ &+ (\beta_3 + \beta_{3g} * G_i) * t_{ij}^3 \\ &+ b_{0i} + b_{1i} * t_{ij} + b_{2i} * t_{ij}^2 + b_{3i} * t_{ij}^3 \\ &+ \varepsilon_{ij}. \end{aligned} \tag{15}$$

where SI_HVC_{ij} is the observed outcome for bird i at time j (t_{ij}[1]) is the observed time at which measurement j was taken for bird i and group membership (G_i[2]). The fixed effects include β_0, β_1, β_2, and β_3, representing the population-level intercept, linear, quadratic, and cubic slopes over time, respectively. Group-specific differences are captured by β_{0g}, β_{1g}, β_{2g}, and β_{3g}, which quantify treatment effects on the intercept and slopes. Individual variability is modeled through subject-specific random effects b_{0i}, b_{1i}, b_{2i}, and b_{3i} for the intercept, linear, quadratic, and cubic time terms respectively. This model incorporates time (t_{ij}), squared time (t_{ij}^2), and cubed time (t_{ij}^3) as predictors, with residual errors ε_{ij} assumed to follow a normal distribution: $\varepsilon_{ij} \sim N(0, \sigma^2)$. The random effects vector $(b_{0i}, b_{1i}, b_{2i}, b_{3i})$ is assumed to follow a multivariate normal distribution with mean vector 0 and diagonal covariance matrix with variances $d_{00}, d_{11}, d_{22}, d_{33}$. The interaction terms to allow for group-specific differences in each of the present effects, this is meant to show if there is an effect of testosterone on the signal intensity of the birds. The model will be fitted in SAS using the NLMIXED procedure.

In an attempt to find a model that fits and describes the data well, another model with different fractional polynomials will be fitted and from among the two models, the best one for the current data will be chosen according to the Akaike information criterion (AIC) and Bayesian information criterion (BIC) and the model with the lowest of these two values will be chosen.
This model includes fractional polynomials in terms of the logarithm and square root of time and it is adopted from [8]. The model is formulated as follows:

$$\begin{aligned} SI\_HVC_{ij} =& (\alpha_0 + \alpha_1 G_i + a_i) \\ &+ (\lambda_0 + \lambda_1 G_i + l_i) ln(t_{ij}) \\ &+ (\delta_0 + \delta_1 G_i + d_i) t_{ij}^{0.5} + \varepsilon_{ij}. \end{aligned} \tag{16}$$

where:

- SI_HVC_{ij} = observed outcome for bird i at time j.
- α_0 = The overall intercept (baseline value of SI_{ij} when all other variables are zero and for the reference group).
- α_1 = The additional effect on the intercept for being in group $G_i = 1$ (the difference in baseline between the groups).
- λ_0 = The effect of $\log(t_{ij})$ (log of time) on the outcome for the reference group.
- λ_1 = The additional effect of $\log(t_{ij})$ for group $G_i = 1$ (interaction between group and log-time).

[1] t_{ij} carries the same meaning in all subsequent models
[2] G_{ij} retains the same definition throughout

- δ_0 = The effect of the square of the time on the outcome for the reference group.
- δ_1 = The interaction between group and square root of time (additional effect of $t_{ij}^{0.5}$ for the treated group, $G_i = 1$).
- a_i = Subject-specific random intercept.
- l_i, d_i = Subject (bird)-specific random effect for the log-time term and the square root time term.
- ε_{ij} = The residual error for each observation and $\varepsilon_{ij} \sim N(0, \sigma^2)$.

The model will be fitted following the same procedure as that of the former model. The results from the chosen model will be evaluated to identify any differences resulting from using the various integration techniques.

3.3 Bi-exponential model

The HVC outcome is connected to pharmacokinetic theory which studies dispersion of compounds through a living organism. HVC can therefore be viewed as the central compartment and $area\_X$ and RA are the areas to which manganese is dispersed, therefore a bi-exponential model seems like a good choice. As stated by [8], the mixed-effects approach yields the model below:

$$Y_{ij} = \exp(\beta_{i1} \exp[-\exp(-\beta_{i2} t_{ij}] - \exp(\beta_{i3}) \exp(-\exp[-\beta_{i4} t_{ij}] + \varepsilon_{ij}. \tag{17}$$

Splitting the β parameters into fixed and random effects and also including the group effect leads to:

$$\begin{aligned} Y_{ij} =& e^{(\beta_1 + \gamma_1 G_i + b_{1i})} \exp[-e^{(-\beta_2 + \gamma_2 G_i + b_{2i}) t_{ij}}] \\ & - e^{(\beta_3 + \gamma_3 G_i + b_{3i})} \exp[-e^{(-\beta_4 + \gamma_4 G_i + b_{4i}) t_{ij}}] + \varepsilon_{ij}, \end{aligned} \tag{18}$$

where:

- Y_{ij} = The outcome for subject i at time j.
- β_1 = Baseline log-amplitude of the first exponential component for the reference group.
- β_2 = Baseline log-rate (decay) parameter of the first exponential component for the reference group.
- β_3 = Baseline log-amplitude of the second exponential component for the reference group.

- β_4 = Baseline log-rate (decay) parameter of the second exponential component for the reference group.
- γ_1, γ_3 = Additional effect of being in group $G_i = 1$ on the log-amplitude of the first and second component respectively.
- γ_2, γ_4 = Additional effect of being in group $G_i = 1$ on the log-rate of the first and second component respectively.
- b_{1i}, b_{3i} = subject-specific random effects for the log-amplitude in the first and second components respectively.
- b_{2i}, b_{4i} = subject-specific random effects for the log-rate in the first and second components respectively.
- ε_{ij} = residual error for the subject i at time j and $\varepsilon_{ij} \sim N(0, \sigma^2)$.

The amplitude parameters control the starting value or height of each exponential component and the rate parameters control how quickly each exponential component decays over time (higher rates mean faster decay). The group effects (γ terms) describe how group membership affects the amplitude or rate for each component.
This model will be fitted using the SAS procedure NLMIXED and a fixed effects only model will be fitted first so as to obtain starting values for the subsequent models. Given the model complexity, we will increase the complexity of the model gradually first by adding one group effect to each parameter one at a time and then the random effects one at a time as well. The resulting model will then be reported and its goodness of fit investigated.

3.4 Two-stage approaches

For each of the above described models, a two-stage approach will be fitted and this is because the two-stage approach simplifies fitting of complex models such as the ones we are trying to fit. As described by [24], in the first stage, the individual-level models (the non-linear models) will be fitted separately for each subject including only the time covariate and no other covariates, providing subject-specific estimates of all the models. In the second stage, these individual parameter estimates will be treated as observed outcomes and a linear regression model including the other covariates of interest and subject-specific random effects in order to estimate the population-level fixed effects, assess group differences, and quantify between-subject variability. The multivariate regression model in stage two is of the form:

$$\beta_i = K_i\beta + b_i \tag{19}$$

where:

- β_i = The subject-specific regression coefficients (parameter estimates) from stage one.
- K_i = A $(q \times p)$ matrix of known covariates.
- β = A p-dimensional vector of unknown regression parameters.
- b_i = Random effects which are assumed to be independent with a q-dimensional normal distribution with mean vector zero and general covariance matrix D.

After the second stage, the D matrix will be approximated by the empirical variance-covariance matrix in order to measure the relationships between the individual-specific model parameters, we will compute the sample covariance matrix using the Stage 1 parameter estimates for each subject. Specifically, after extracting the estimated values for each individual, we will use the PROC CORR procedure with the COV option in SAS to determine the covariance matrix. It is however important to note that the estimated matrix is not the true D matrix, it only gives us an idea about the true D matrix matrix.

3.5 Other models

This subsection describes other models that could potentially fit the signal intensity of HVC even better. These were selected from the CurveExpert professional software as the best performing options for SI_HVC_{ij}.

3.5.1 Gaussian model

We will model the outcome variable (SI_HVC) as a non-linear function of time using a Gaussian (bell-shaped) curve with both fixed and random effects, implemented in SAS procedure NLMIXED. This model allows the amplitude, peak position, and width of the Gaussian curve to vary by group and individual (bird). The model equation is:

$$SI\_HVC_{ij} = (a + a_g G_i + b_{1i}) \exp\left(-\frac{(t_{ij} - (b + b_g G_i + b_{2i}))^2}{2((c^2) + c_g G_i + b_{3i})} \right) + \varepsilon_{ij}, \tag{20}$$

where:

- SI_HVC_{ij} = The outcome for subject i at time j.
- a = Baseline amplitude (height) of the Gaussian curve for the reference group.
- a_g = Additional amplitude effect for the comparison group (testosterone-treated birds).
- b = Baseline peak position (center of the curve) for the reference group.
- b_g = Additional peak position effect for the comparison group.

- c = Baseline width (spread) of the curve for the control birds.
- c_g = Additional spread effect for the treated group.
- b_{1i}, b_{2i}, b_{3i} = Subject-specific deviation in amplitude (height), peak position (center) and width (spread) respectively.
- ε_{ij} = Residual error and $\varepsilon_{ij} \sim N(0, \sigma^2)$.

3.5.2 Gompertz model

The Gompertz model is widely used for sigmoidal (S-shaped) growth or decay processes and allows for both fixed and random effects on the curve parameters. The model has the following parametric shape:

$$SI\_HVC_{ij} = a \exp[-\exp((b + b_g G_i + b_{1i}) - (c + c_g G_i)t_{ij}] + \varepsilon_{ij}, \tag{21}$$

where:

- SI_HVC_{ij} = The outcome for subject i at time j.
- a = Asymptote (maximum value the curve approaches).
- b = Baseline location parameter (related to the timing of the inflection point, when growth is fastest).
- b_g = Additional effect on b for the comparison group.
- c = Baseline rate parameter (growth or decay rate).
- c_g = Additional effect on c for the treated group.
- b_{1i} = Subject-specific random effect on the location parameter b (individual timing differences).
- ε_{ij} = Residual error and $\varepsilon_{ij} \sim N(0, \sigma^2)$.

The rate parameter c, controls how quickly the curve rises or falls. This model allows the growth (or decay) trajectory of each subject (bird) to follow a flexible S-shaped curve, with group-level differences (control and treated group differences) in curve shape and timing, and individual-level variation in the timing of the curve's inflection point.

3.5.3 Logistic model

We will model the outcome variable (SI_HVC_{ij}) as a non-linear function of time using a logistic mixed-effects model, implemented in SAS PROC NLMIXED. The model equation is:

$$SI\_HVC_{ij} = \frac{a}{1 + (b_g G_i + b_{1i}) \exp[(-c + c_g G_i + b_{2i}) t_{ij}]} + \varepsilon_{ij}. \tag{22}$$

In model (22) SI_HVC_{ij} is the response for subject i at time j. The parameter a represents the upper asymptote (maximum response), b and c are the baseline location/shape and rate parameters for the reference group respectively, while b_g and c_g are the additional effects for the comparison group. Subject-specific random effects b_{1i} and b_{2i} allow each individual to have their own curve shape and rate. Residual errors ε_{ij} are assumed to be normally distributed with mean zero and variance σ^2.

3.5.4 Ratkowsky model

This model is used for sigmoidal (S-shaped) growth and it allows for fixed and random effects on the curve parameters just like in the Gompertz model described earlier. The model is formulated as follows:

$$SI\_HVC_{ij} = \frac{a + a_g G_i + b_{1i}}{1 + \exp[(b + b_g G_i) - (c + c_g G_i + b_{3i}) t_{ij}]} + \varepsilon_{ij}, \tag{23}$$

where:

- SI_HVC_{ij} = The outcome for subject i at time j.
- a = Baseline upper asymptote (maximum response) for the reference group.
- a_g = Additional effect on a for the comparison group.
- b = Baseline location parameter for the reference group.
- b_g = Additional effect on b for the comparison group.
- c = Baseline rate parameter for the reference group.
- c_g = Additional effect on c for the treated group.
- b_{1i}, b_{3i} = Subject-specific random effects on the asymptote and the rate parameter respectively.
- ε_{ij} = Residual error and $\varepsilon_{ij} \sim N(0, D)$.

Baseline location parameter, b controls the inflection point or the time at which the curve rises most steeply for the reference group. Baseline rate (slope) parameter, c controls how rapidly the response increases over time for the control group.

4 Results

4.1 Model of Van der Linden et al, (2002)

We built a model for the second period, where treatment had been administered because no treatment effect would be expected in the first period (measurements were taken prior to treatment).
Just as applied by [23] we use the model (13) to analyze the data for which the general form has 8 fixed-effects parameters and 4 variance components (3 variances in D, 3 covariances in D, and σ^2). However, this model failed to converge and therefore a diagonal covariance matrix was used in the final model[3]. The model is fitted using SAS procedure NLMIXED, using several integration techniques such as adaptive Gaussian quadrature, non-adaptive Gaussian quadrature and Laplace approximation.
The model obtained from backward selection is as follows:

$$SI_{ij}(RA) = \frac{(\phi_0 + f_i)t_{ij}^{\eta_0+\eta_1 G_i}}{(\tau_0 + t_i)^{\eta_0+\eta_1 G_i} + t_{ij}^{\eta_0+\eta_1 G_i}} + \varepsilon_{ij}. \tag{24}$$

Table 1: Parameter estimates from the model of Van der Linden et al, (2002) for the signal intensity in the RA region at the second period.

Effect	**Parameter**	**Estimate**	**Std. Error**	**Pr** $>\|t\|$
	ϕ_0	0.4527	0.04775	< 0.0001
	η_0	2.1825	0.08016	< 0.0001
	η_1	0.4285	0.1060	0.0037
	τ_0	2.8480	0.1762	< 0.0001
$Var(f_i)$	d_{11}	0.02245	0.01010	0.0569
$Var(t_i)$	d_{22}	0.2883	0.1339	0.0635
$\mathrm{Var}(\varepsilon_{ij})$	σ^2	0.000185	0.000017	< 0.0001

The results in Table 1 reflect the estimates obtained using adaptive Gaussian quadrature and Laplace approximation, for non-adaptive Gaussian quadrature $Var(t_i)$ is significant ($p < 0.001$). The untreated birds have a significant estimated average baseline maximum value of 0.4527 ($p < 0.0001$). The curve shape (steepness) parameter estimate for the control birds is highly significant with p-value less than 0.0001 (mean = 2.1825). The average effect of testosterone on the curve shape is estimated at 0.4285 and there is sufficient evidence to support this effect, $p = 0.0037$, indicating that treated birds have a faster signal rise. The average estimated time to half of the maximal signal intensity for the untreated birds is 2.8480 and this result is statistically significant, $p < 0.0001$. The variance of the random

[3] All subsequent models were fitted with a diagonal variance-covariance matrix except the two-stage approaches.

effect on the maximal signal intensity is estimated as 0.02245 and it is marginally significant ($p = 0.0569$) meaning that there is insufficient evidence to support between-bird differences in their maximum signal intensity. Variance of the random effect on the time to half of the maximum signal intensity is estimated to be 0.2883 and it is not statistically significant ($p = 0.0635$), therefore there is no evidence to support a difference in this effect between control and treated birds. The within-bird (residual) variance is very small and highly significant, thereby indicating that the model fits the data well. The fitted values plotted against the observed values of the SI_RA support our choice of model as shown in Figure 5.

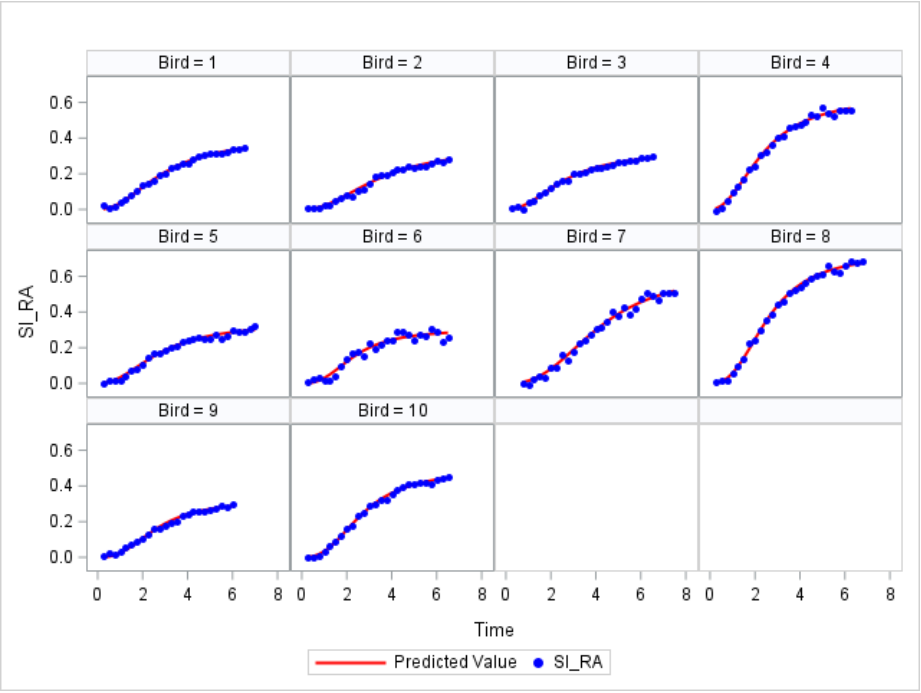

Figure 5: Fitted curves for the signal intensity in the RA region by bird at the second period.

A similar model was fit for the signal intensity of the $Area\_X$ but all attempts led to inconclusive results. Even changing the integration and optimization techniques yielded unstable results, therefore no parameter estimates were obtained for this outcome.

4.2 Fractional polynomials

Given their flexibility and allowance for various parametric shapes we fit the model with fractional polynomials wherein the fractional polynomials were applied to the time covariate up to the third degree polynomial. We used several integration techniques such as adaptive Gaussian quadrature (with 5 quadrature points), non-adaptive Gaussian quadra-

ture ($q = 5$) and Laplace approximation ($q = 1$). The models using adaptive Gaussian quadrature and Laplace approximation were unable to converge to a stable solution but non-adaptive Gaussian quadrature reached convergence. However, following the description of the fractional polynomials in Section 3, this model was compared to the model that included the logarithmic and square root polynomials of time. The results of this comparison are shown in Table 2.

Table 2: Model selection for the fractional polynomials model for the signal intensity of the high vocal center (HVC).

Model	**-2loglik**	**AIC**	**BIC**
Cubic	-196.4	-170.4	-166.5
Log-sqrt	-265.3	-245.3	-242.3

From Table 2, the ideal model was the latter model with logarithmic and square root polynomials of time because it had the smallest values of AIC and BIC meaning it fit the data better than the model with up to the third degree polynomial. The selected model has the following output:

Table 3: Parameter estimates for the fractional polynomials model for the signal intensity of the high vocal center (HVC).

Effect	**Parameter**	**Estimate**	**Standard Error**	$Pr > \|t\|$
	α_0	2.4606	0.4772	0.0013
	α_1	0.9505	0.6750	0.2019
	λ_0	0.2898	0.1640	0.1206
	λ_1	0.2421	0.2325	0.3324
	δ_0	-0.6343	0.1757	0.0086
	δ_1	-0.4040	0.2491	0.1489
$SD(a_i)$	d_a	1.1321	0.5095	0.0617
$SD(l_i)$	d_l	0.1318	0.0622	0.0719
$SD(d_i)$	d_d	0.1490	0.0707	0.0731
Residual SD	σ^2	0.0023	0.0002	< 0.0001

For the reference group at time zero, the mean SI_HVC is 2.4606 and the group effect (α_1) was obtained as 0.9505 ($p = 0.2019$) meaning that there is insufficient evidence to support the claim that being in group 1 (treated) affects the baseline SI_HVC. Each unit increase in log(time) increases SI_HVC by 0.2898 in the reference group and this result is not statistically significant, $p = 0.1206$. In the treated birds, the effect of log(time) is further increased by 0.2421 and this result is not statistically significant ($p = 0.3324$). The effect of $\sqrt{\text{time}}$ for the reference group is estimated as -0.6343 and this effect is statistically significant ($p = 0.0086$). In the testosterone-treated birds, the effect of $\sqrt{\text{time}}$ is less nega-

tive and non-significant (effect $= -0.4040$, $p = 0.1489$), indicating that there is no evidence of a different trajectory compared to the control birds (reference group). d_a, d_l, d_d are the standard deviations of the random effects on the intercept ($SD = 1.1321$, $p = 0.0617$), the log(time) ($SD = 0.1318$, $p = 0.0719$) and square-root(time) ($SD = 0.1490$, $p = 0.0731$) and they all indicate that there is no evidence of substantial individual variability in how SI_HVC changes with log(time) and square-root(time). The within-bird (residual) variance is very small and highly significant indicating that the model fits the data well. To verify this outcome the fitted values were plotted against the observed values as shown in Figure 6.

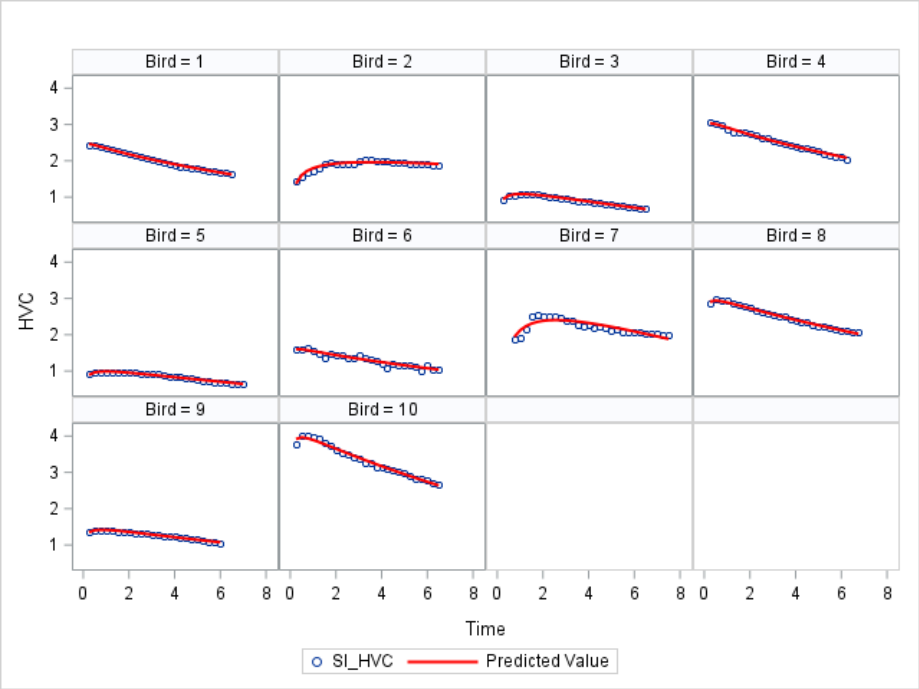

Figure 6: Fitted vs observed curves for the signal intensity if the high vocal center (HVC) by bird at the second period.

The results in Table 4 reflect the model using adaptive Gaussian quadrature and it is worth noting that they carry the same meaning as those from Table 3. The only statistically significant effects are the intercept, α_0 (effect $= 2.4606$, $p = 0.0013$) and the square-root of time effect (effect $= -0.6343$, $p = 0.0086$). This differs from the findings of the same model fit with non-adaptive Gaussian quadrature.
Laplace approximation yields the same results as those from the model in which we employed adaptive Gaussian quadrature, therefore the same interpretations of the model parameter estimates hold. The model using non-adaptive Gaussian quadrature model also

Table 4: Model parameter estimates (non-adaptive Gaussian quadrature).

Effect	**Parameter**	**Estimate**	**Standard Error**	$Pr > \|t\|$
	α_0	2.4325	0.08421	$< .0001$
	α_1	0.6390	0.1197	0.0011
	λ_0	0.8425	0.0579	$< .0001$
	λ_1	0.3789	0.0798	0.0021
	δ_0	-0.0347	0.0781	0.6706
	δ_1	-0.5939	0.1108	0.0010
$SD(a_i)$	d_a	0.0793	0.1108	$< .0001$
$SD(l_i)$	d_l	0.2778	0.0205	$< .0001$
$SD(d_i)$	d_d	0.1172	0.0065	$< .0001$
Residual SD	σ^2	0.0125	0.0011	$< .0001$

presents a good fit as shown in Figure 7. Figure 14 shows that the normality assumption is reasonably fulfilled for this model.

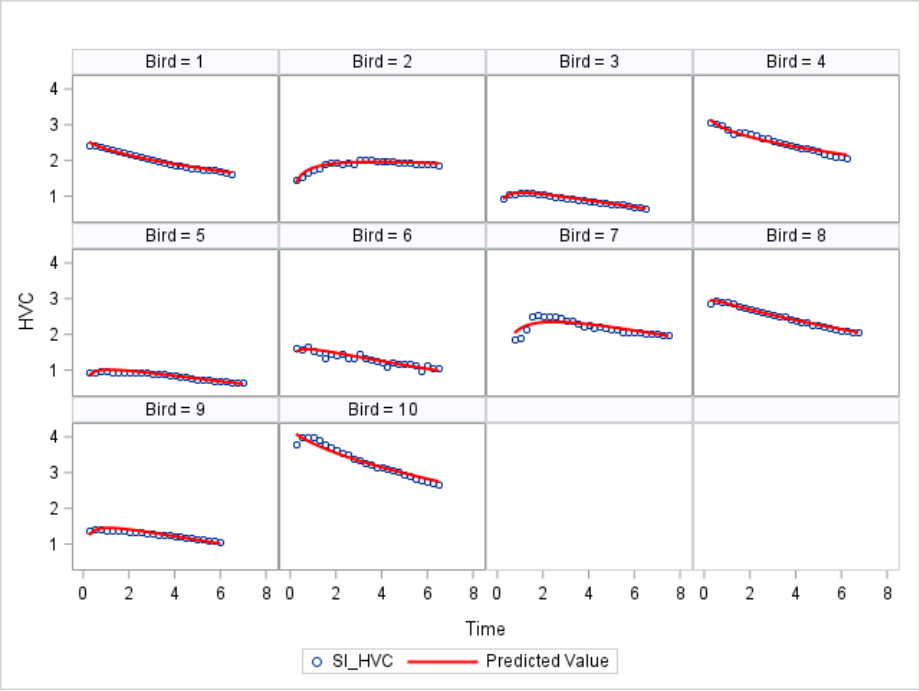

Figure 7: Fitted vs observed curves for the signal intensity if the high vocal center (HVC) by bird at the second period.

Looking at the AIC and BIC of these two models in Table 5 suggests that the model using adaptive Gaussian quadrature is a better fit to the data.

Table 5: Model comparison for the fractional polynomials model for the signal intensity of the high vocal center (HVC)

Model	**-2loglik**	**AIC**	**BIC**
Adaptive	-656.7	-636.7	-633.7
Non-adaptive	-265.3	-245.3	-242.3

4.3 Two-stage approaches

4.3.1 Fractional polynomials for outcome, SI_HVC

Using the stage 1 parameter estimates as the outcome in the linear model along with the inclusion of covariates, the second stage results are shown below:

Table 6: Stage 2 results for the signal intensity in the high vocal center (HVC).

Parameter	**Effect**	**Group**	**Estimate**	**Std. Error**	$Pr > \|t\|$
α_0	Intercept		4.5602	0.7866	0.0004
	Group	0	-1.4957	1.1124	0.2157
λ_0	Intercept		0.5474	0.1987	0.0249
	Group	0	-0.2573	0.2810	0.3866
δ_0	Intercept		-2.9798	0.6001	0.0011
	Group	0	1.1841	0.8486	0.2004

The baseline value of the signal intensity in the HVC region (when all predictors are zero) is estimated at 4.56 and this is highly statistically significant ($p < 0.001$), indicating a strong baseline effect. Furthermore, the group effect for the intercept is negative, suggesting that the control group has a lower baseline signal intensity by about 1.5 units compared to the reference group. However, this difference is not statistically significant ($p = 0.22$), so there is no strong evidence for a baseline group difference. The coefficient for the log(time) term at baseline is positive (0.5474) and significant ($p = 0.025$), suggesting that, on average, as log(time) increases, signal intensity increases. The group effect of the log(time) coefficient is -0.2573 but it is not statistically significant ($p = 0.3866$), indicating that there is no evidence that the groups differ in how log(time) affects signal intensity in the high vocal center. The coefficient for the square-root(time) term at baseline is negative (-2.9798) and highly significant ($p = 0.0011$), indicating that as square-root(time) increases, signal intensity decreases. However, the group effect for the square-root(time) coefficient is 1.1841, suggesting that the untreated birds have a less negative effect of square-root(time) on the signal intensity, but this effect is not statistically significant ($p = 0.20$). Since none of the group effects are statistically significant, there is insufficient evidence to conclude that the groups (testosterone-treated and untreated birds) differ.

To measure the relationships between the individual fractional polynomial parameters, we computed the sample covariance matrix using the Stage 1 parameter estimates for each subject. Specifically, after extracting the estimated values for α_0, λ_1 and δ_2 for each individual, we used the SAS procedure CORR with the COV option to determine the covariance matrix shown in Table 7.

Table 7: Covariance matrix of fractional polynomial parameters.

	α_0	λ_0	δ_0
α_0	3.3715	0.6354	-2.4867
λ_0	0.6354	0.1939	-0.5428
δ_0	-2.4867	-0.5428	1.9898

The variance of α_0 is 3.3715 showing that there is substantial variability in the intercept parameter across individuals (birds). For the log(time) effect, variance is 0.1939 indicating less variability in the log(time) between individuals. The variance is 1.9898 for the square-root(time) suggesting moderate variability in the square-root(time) effect across individuals. α_0 and λ_0 have a positive covariance (0.6354) meaning the birds with higher intercepts also tend to have higher log(time) effects. There is a strong negative covariance between α_0 and δ_0 (-2.4867) which indicates that birds with higher intercepts tend to have more negative square-root(time) effects. Furthermore, λ_0 and δ_0 have a negative covariance of -0.5428 therefore, birds with higher log(time) effects tend to have more negative square-root(time) effects.

4.3.2 Model of Van der Linden et al, (2002)

For SI_RA**:**

Table 8: Stage 2 results for the signal intensity of the RA region from the Model of Van der Linden et al, (2002).

Parameter	**Effect**	**group**	**Estimate**	**Standard Error**	$Pr > \|t\|$
ϕ	Intercept		0.4978	0.0722	0.0001
	group	0	-0.0929	0.1020	0.3894
η	Intercept		2.6520	0.1676	$< .0001$
	group	0	-0.3945	0.2370	0.1347
τ	Intercept		0.3633	0.0292	$< .0001$
	group	0	-0.0185	0.0413	0.6660

The mean estimated maximal signal intensity (ϕ) of the testosterone-treated and control birds did not differ significantly (difference $= -0.0929$, $p = 0.3894$). The estimated mean maximal signal intensity in the treatment group is 0.4978 whereas the control birds have

an estimated mean of 0.4049 [4]. There is no significant difference in the mean estimated curve shape parameter (η) between the treated and control birds (difference = -0.3945, $p = 0.1347$). The control birds have a mean shape parameter estimated at 2.2575 and that of the treated birds is 2.6520. The mean estimated time to reach half of the maximal signal intensity in the treated birds is not significantly different from that in the control birds (difference = 0.0185, $p = 0.6660$). In the treated birds the mean estimated time to half of the maximal signal intensity is 0.3633 while in the control birds it is estimated at 0.3488. This therefore leads to the conclusion that there is no evidence of a difference between the treated and control birds.
Consider the empirical covariance matrix below to estimate the variance-covariance (D) matrix. From Table 9, the variance of ϕ (maximal signal intensity) is 0.0255 is relatively

Table 9: Covariance matrix for the Van der Linden model parameters for SI_RA

	ϕ	η	τ
ϕ	0.0255	-0.0053	0.0017
η	-0.0053	0.1681	-0.00002
τ	0.0017	-0.00002	0.0039

small and this is similar to the variance of the time required to reach half of the maximum signal intensity (τ) which is estimated at 0.0039. On the other hand, the variance of the curve shape parameter (η) is much larger, 0.1681, indicating more variability. There is a slight negative association between ϕ and η (covariance = -0.0053). Furthermore, the covariance between ϕ and τ is 0.0017 which suggests a weak positive association and there is essentially no relationship between η and τ (covariance= -0.00002). The matrix above therefore shows that the variance in the curve shape parameter (η) is much larger than for the other parameters, and covariances between parameters are small, suggesting that individual differences in curve shape are the largest source of heterogeneity. Inspection of the residuals shows that the normality assumption is satisfied despite minor deviations in the tails, the residual plot can be found in Figure 13 in the appendix.

For SI_Area_X:
The results in Table 10 show that the average estimated maximal signal intensity for testosterone-treated birds is 0.2399 and it is statistically significant ($p = 0.0001$). The mean estimated maximal signal intensity of the control birds is 0.0992 and there is a significant difference between the treated and control birds in regards to ϕ (difference = 0.1407, $p = 0.0221$). Control birds have a higher mean curve shape parameter (η) (mean = 3.0757) than testosterone-treated birds (mean = 2.41); however, this difference was not statistically significant (difference = 0.6652, $p = 0.1790$). This implies that there is not enough evidence to draw the conclusion that testosterone treatment alters the curve shape parameter. The

[4] Φ (control birds) is obtained from: 0.4978-0.0929 = 0.4049

Table 10: Stage 2 results for the signal intensity in $Area\_X$ from the model of Van der Linden et al, (2002).

Parameter	**Effect**	**group**	**Estimate**	**Standard Error**	$Pr > \|t\|$
ϕ	Intercept		0.2399	0.0351	0.0001
	group	0	-0.1407	0.0497	0.0221
η	Intercept		2.4105	0.3193	$< .0001$
	group	0	0.6652	0.4516	0.1790
τ	Intercept		0.4722	0.0445	$< .0001$
	group	0	-0.0314	0.0630	0.6313

estimated average time to half of the maximal signal intensity is 0.4722 and it is statistically significant ($p < 0.0001$), this is higher than the value estimated for the control birds (mean = 0.4408). There is insufficient evidence to support the difference between the treated and control birds for τ (difference = -0.0314, $p = 0.6313$).

Table 11: Covariance matrix of parameters from the model of Van der Linden et al, (2002) for the signal intensity in $Area\_X$

	ϕ	η	τ
ϕ	0.0110	-0.0546	0.0068
η	-0.0546	0.5761	-0.0375
τ	0.0068	-0.0375	0.0091

The empirical covariance matrix in Table 11 shows that the variance of the maximal signal intensity is 0.011 and that of the shape parameter and time to half maximal intensity is 0.5761 and 0.0091 respectively. There is a negative association between η and ϕ (-0.0546), there is also a weak negative association between τ and η, -0.0375. Furthermore, there is a weak positive association between ϕ and τ (0.0068). From the above matrix we can conclude there is a substantial variability between individuals in regards to the shape of the curves. Most subjects have similar peak intensities and the timing to reach half-maximal intensity due to the low variability of ϕ and τ.

4.4 Bi-exponential model

The bi-exponential model presented multiple numerical and statistical challenges so much so that there were no conclusive results when it was fitted to all the outcomes (SI_RA, SI_HVC, SI_Area_X). This is because the several optimization techniques employed failed to converge to a stable solution and also the singularity of the final Hessian matrix indicating that the model was probably over-parameterized or even misspecified. To combat these challenges a fixed effects only model was fitted but even then the Hessian matrix was not positive definite (meaning the solution is a local maximum rather than the desired global

maximum) therefore rendering the estimates invalid for interpretation. More attempts were made by including the group (treatment) and random effects on a few of the parameters at a time not only to establish their importance on each of the model parameters but also to simplify the model but there was no improvement whatsoever. Iterated optimization was applied in which a pair of the model parameters were fixed at a time and the remaining pair was estimated. The newly estimated results were then used to fix the next pair of parameters. These cycles were repeated several times until the model parameter estimated were not changing any more but each cycle led to the same convergence issues described above. All of these findings indicated that this model is not suitable for the data at hand. This presented an opportunity to explore other possible models that could potentially fit the data better. This exploration was done using the CurveExpert professional software [25], the models obtained are reported in the next subsection.

4.5 Other models

4.5.1 Gaussian model

This model was fitted with adaptive Gaussian quadrature ($q = 5$) and optimization was done using Newton-Raphson ridge.

Table 12: Parameter estimates of the gaussian model for the signal intensity in the high vocal center (HVC)

Effect	**Parameter**	**Estimate**	**Standard Error**	$Pr > \|t\|$
	a	2.1609	0.5007	0.0035
	b	-3.8991	2.2298	0.1238
	c	10.8122	1.3574	< 0.0001
	a_g	0.7173	0.7082	0.3448
	b_g	-1.4175	3.0043	0.6514
	c_g	30.1986	38.8993	0.4630
Residual SD	σ	0.06638	0.0030	< 0.0001
$SD(b_{1i})$	d_1	1.1002	0.2620	0.0040
$SD(b_{2i})$	d_2	4.4678	1.1307	0.0055
$SD(b_{3i})$	d_3	54.1597	14.2809	0.0066

From Table 12[5], we observe that the average peak height, a in the reference group (control birds, group = 0) is estimated as 2.1609 and it is statistically significant ($p = 0.0035$). The estimated mean peak position, b for the reference group is -3.8991 but there is insufficient evidence to support this estimate ($p = 0.1238$). The average spread of the curve is estimated to be 10.8122 and this is highly significant with a p-value < 0.0001. The group effect on the amplitude (peak height) is estimated to be 0.7173 and it is not statistically significant

[5]SD = standard deviation ($\sqrt{variance}$)

($p = 0.3448$). Similarly the group effect on the peak position (mean $= -1.4175$, $p = 0.6514$) and on the spread of the curve (mean $= 30.1986$, $p = 0.4630$) are not statistically significant, therefore there is no evidence that the treatment affects the amplitude (peak height), peak position and spread of the curve. The standard deviations of all three random effects (d_1, d_2, d_3) are significant ($p < 0.05$), indicating substantial between-individual variability in amplitude, peak position, and curve spread. The standard deviation of the curve spread is especially large ($p = 0.0066$), suggesting that the width of the curve varies greatly among individuals. The incredibly small and highly significant residual standard deviation (sigma) indicates that the model describes the data well. This claim is backed by the goodness of fit plot in Figure 8. Inspection of residuals in Figure 15 shows that the normality assumption is reasonably fulfilled.

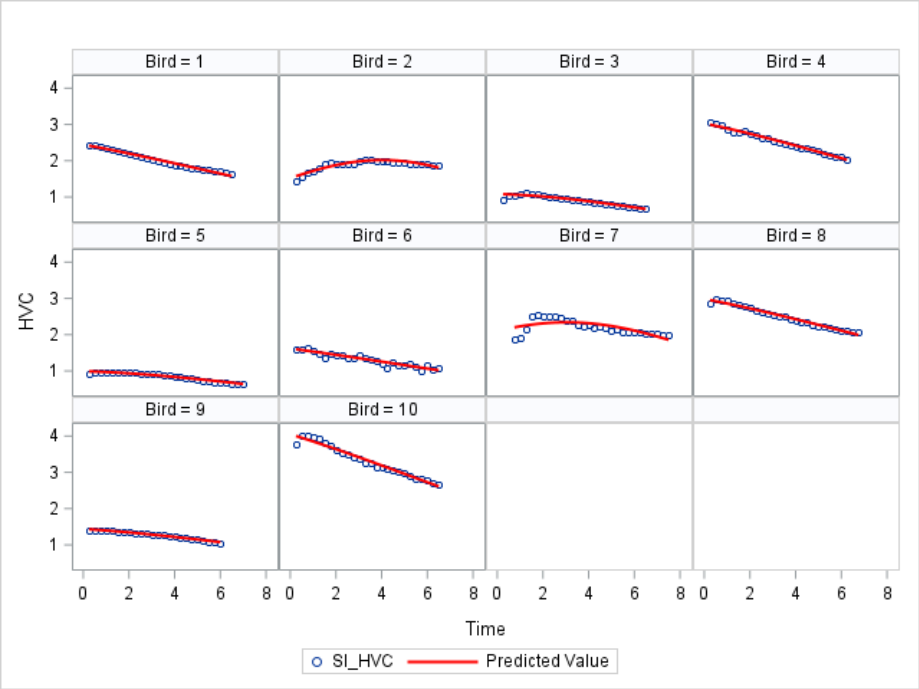

Figure 8: Fitted vs observed curves for the signal intensity in the high vocal center (HVC) by bird at the second period for the Gaussian model.

4.5.2 Gompertz model

A Gompertz model including random and group effects on all the model parameters was fit but it did not display a good fit to the data as shown in Figure 9. This inspired a model reduction procedure in which the group effect was added to each of the model parameters one at a time in order to establish their relevance in the model. This same process was repeated for the choice of random effects and the final model without convergence issues

and a good fit to the data was one that included one random effect and a group effect on two of the model parameters. It was fitted using non-adaptive Gaussian quadrature ($q = 5$) and Newton-Raphson ridge. The estimates of this model are presented in Table 13.

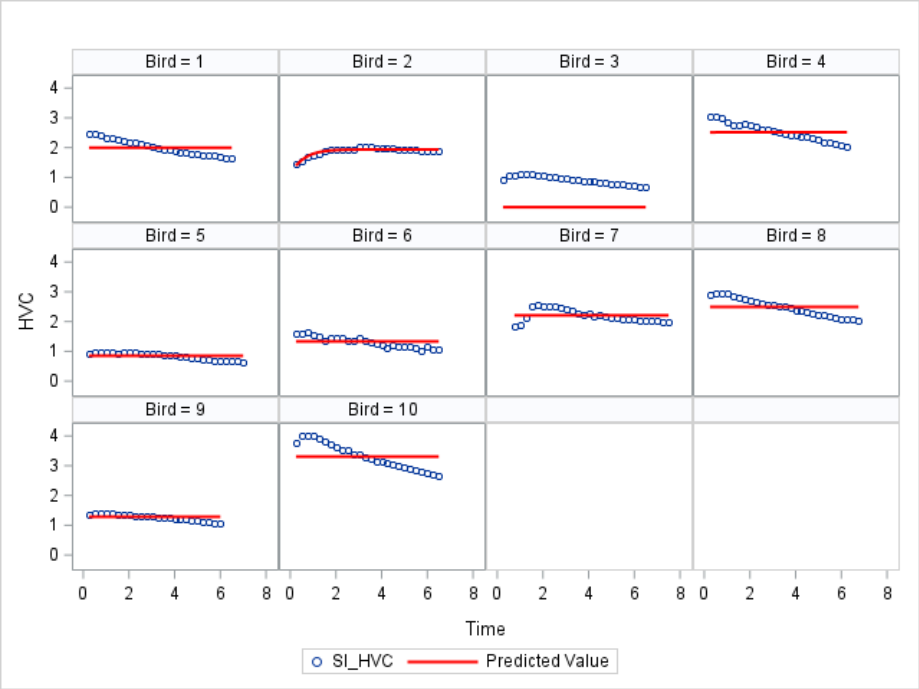

Figure 9: Fitted vs observed curves for the signal intensity in the high vocal center (HVC) by bird at the second period for the poorly fitting Gompertz model.

The average estimated maximum value for the control (reference) group is 5.3788 and this estimate is statistically significant ($p < 0.0001$). The location parameter estimate is statistically significant with mean = 0.4416 and $p-value < 0.0001$ in the control birds. The estimated average growth rate in the reference group is estimated to be -0.0394 and there is enough evidence to support this outcome ($p < 0.0001$). There is a significant difference in the location parameter between the testosterone-treated birds and the untreated birds (difference $= -0.3316$, $p < 0.0001$) with treated birds having a lower estimate, 0.11[6]. Similarly the group effect on the growth rate is statistically significant with a difference of -0.0188 and $p = 0.04$. The standard deviation of the random effect on the location parameter is estimated as 0.3531 and it is statistically significant ($p < 0.0001$) indicating that there are substantial individual differences in the location parameter. The within-bird (residual) standard deviation is small and statistically significant indicating a good model fit.

[6]location parameter estimate of the treated birds, b is given by 0.4416 - 0.3316 = 0.11

Table 13: Parameter Estimates for the signal intensity in the high vocal center (HVC) from the Gompertz model

Effect	**Parameter**	**Estimate**	**Standard Error**	$Pr > \|t\|$
	a	5.3788	0.5085	< 0.0001
	b	0.4416	0.0585	< 0.0001
	c	-0.0394	0.0062	0.0001
	b_g	-0.3316	0.0426	< 0.0001
	c_g	-0.0188	0.0078	0.0400
Residual SD	σ	0.2099	0.0092	< 0.0001
$SD(b_{1i})$	d_1	0.3531	0.0378	< 0.0001

The plot of fitted versus observed values in Figure 10 and it suggests an acceptable fit of the model to the data for some of the birds and a poor fit for the other birds. The residuals were also assessed for normality and Figure 16 shows that most points fall close to the diagonal reference line, with only slight deviations at the tails, therefore the residuals have reasonable normality.

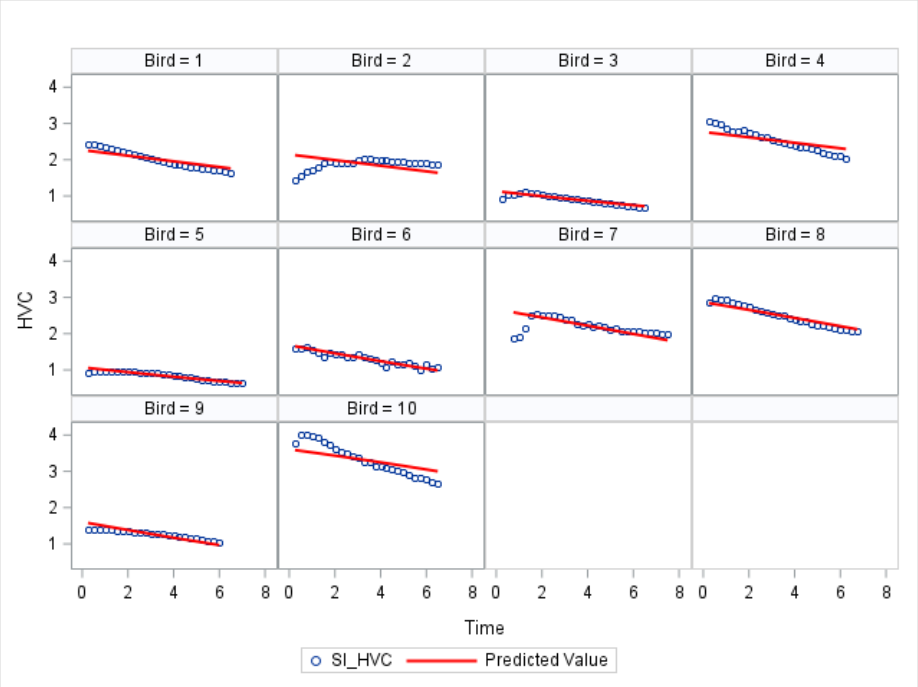

Figure 10: Fitted vs observed curves for the signal intensity in the high vocal center (HVC) by bird at the second period for the Gompertz model.

4.5.3 Logistic model

A three parameter logistic model was fit with two random effects after a model reduction procedure similar to what was done in the subsection above (Gompertz model). The model was fitted using non-adaptive Gaussian quadrature ($q = 5$) and Newton-Raphson with line search. The resulting model parameter estimates are shown in Table 14:

Table 14: Parameter Estimates for the signal intensity in the high vocal center from the logistic model.

Effect	**Parameter**	**Estimate**	**Standard Error**	$Pr > \|t\|$
	a	2.7759	0.0359	< 0.0001
	b	0.6264	0.0263	< 0.0001
	c	0.0683	0.0154	0.0021
	b_g	-0.0768	0.0148	0.0008
	c_g	-0.0014	0.0174	0.9391
Residual SD	σ	0.1804	0.0079	< 0.0001
$SD(b_{1i})$	d_1	0.3371	0.0118	< 0.0001
$SD(b_{2i})$	d_2	0.1983	0.0091	< 0.0001

Table 14 shows that the mean maximum value is estimated to be 2.7759 for the control birds and there is enough evidence to make this claim. The shape parameter is significant in the control birds with an estimated mean of 0.6264 and p-value less than 0.0001. The mean growth rate of the untreated birds is estimated at 0.0683, this value is statistically significant ($p = 0.0021$). There is a statistically significant difference in the shape parameter between the testosterone-treated and untreated birds (difference $= -0.0763$, $p = 0.0008$). The difference in growth rate between the treated and untreated birds is -0.001, this result is however no statistically significant, $p = 0.9391$. The estimated standard deviation of the random effect on the shape parameter is 0.337 with a p-value less than 0.0001 indicating substantial differences in the birds' shape parameters. There are substantial differences in the growth rate of the different birds and this is due to the highly significant estimated standard deviation of the random effect on the growth parameter ($SD = 0.198$, $p < 0.0001$). The residual (within-bird) standard deviation is small and highly significant ($\sigma = 0.18, p < 0.0001$), this result is supported by the plot of fitted versus observed values in Figure 11. Inspection of the residuals shows that the normality assumption is reasonably fulfilled in Figure 17.

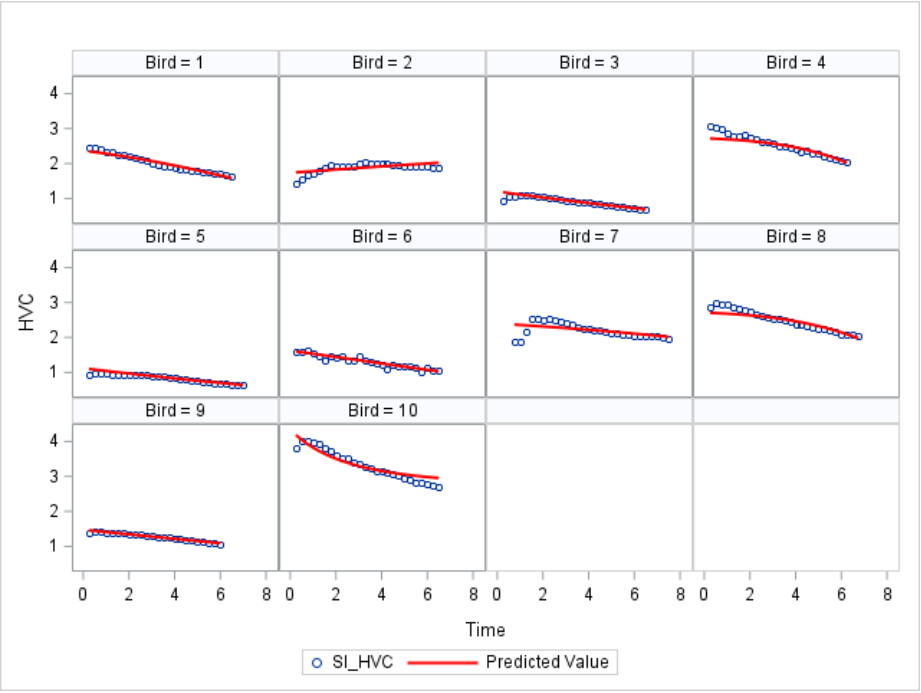

Figure 11: Fitted vs observed curves for the signal intensity in the high vocal center (HVC) by bird at the second period for the logistic model.

4.5.4 Ratkowsky model

A model including the group and random effects on all the model parameters failed to converge, and this is clearly due to the fact the model is too complex for the data at hand. To combat this issue the model was simplified by adding the group and random effects separately to the model parameters and one at a time so that they're only retained where absolutely necessary. This process led to three models each with two random effects on a different pair of model parameters and the group effect on each of the model parameters. The table below shows the loglikelihood, BIC and AIC values from each of these models and these were the basis for choosing a final model. The values in Table 15 above indicate

Table 15: Model selection for the Ratkowsky model for SI_HVC

Model	-2loglik	AIC	BIC
1	-464.4	-446.4	-443.7
2	-476.5	-458.3	-455.5
3	-233.7	-215.7	-213.0

that model 2 (with random effects on the asymptote and growth rate parameters) offers the best fit to the data with appropriate complexity given that it has the lowest AIC and BIC.

Results of this model using adaptive Gaussian quadrature ($q = 5$) and Newton-Raphson with line search are shown in Table 16. [7]

Table 16: Parameter estimates of the Ratkowsky model for the signal intensity of the high vocal center, HVC

Effect	**Parameter**	**Estimate**	**Standard Error**	$Pr > \|t\|$
	a	2.5943	0.6018	0.0026
	b	-0.8200	0.1879	0.0024
	c	0.1152	0.1830	0.5466
	a_g	1.2317	1.0077	0.2564
	b_g	0.1729	0.4456	0.7080
	c_g	-0.2363	0.2530	0.3777
Residual SD	σ	0.07307	0.003374	< 0.0001
$SD(b_{1i})$	d_1	1.3121	0.3204	0.0035
$SD(b_{3i})$	d_3	0.3882	0.1103	0.0079

The asymptote (maximum value) for the reference group (control birds) is estimated to be 2.5943, there is sufficient evidence ($p = 0.0026$) to support this estimate which means that the curves of the control birds level off at the point. There is a significant mean estimate of the curve's midpoint (mean $= -0.82$, $p = 0.0024$), suggesting that the curve rises most steeply at -0.82. The mean growth rate, c is estimated to be 0.1152 but this estimate is not statistically significant ($p = 0.5466$). The difference in the maximum value between the treated and control birds is estimated at 1.2317, this difference is however not statistically significant. Similarly, there is a non-significant difference in the curve's midpoint (mean $= 0.1729$, $p = 0.7080$) and the growth rate (mean $= -0.2363$, $p = 0.3777$) between the testosterone-treated and control birds. The standard deviations of the random effects on the maximum value ($d_1 = 1.3121, p = 0.0035$) and on the growth rate ($d_3 = 0.3882, p = 0.0079$) indicate individual differences in their maximum value and how quickly their curves rise respectively. The residual (within-bird) standard deviation, σ is small and statistically significant which is an indicator of a good model fit.

[7] SD = standard deviation ($\sqrt{variance}$)

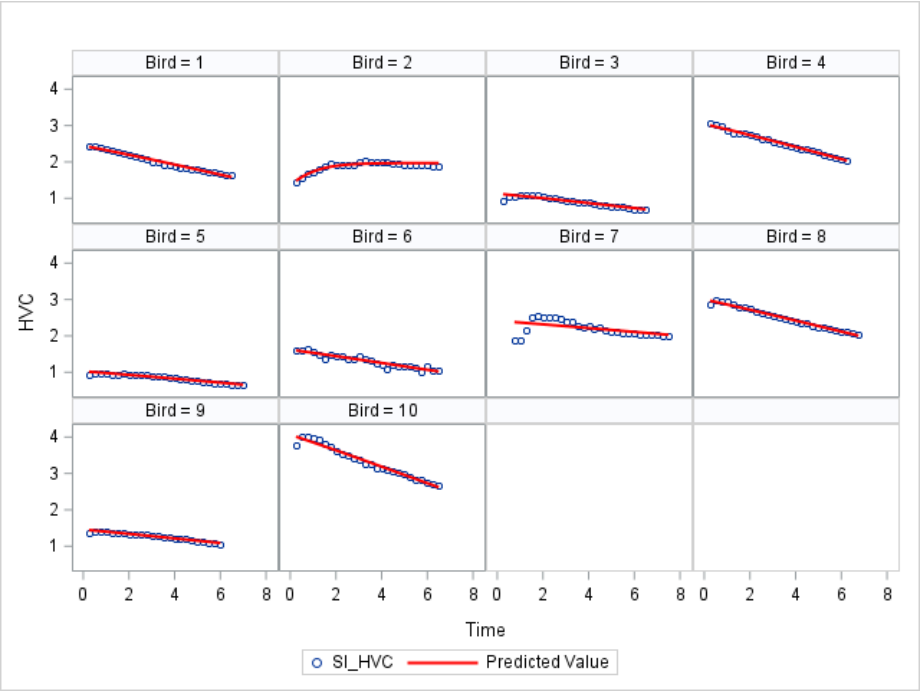

Figure 12: Fitted vs observed curves for the signal intensity in the high vocal center (HVC) by bird at the second period for the Ratkowsky model.

Figure 12 shows the fitted values from the Ratkowsky model plotted against the observed values of the signal intensity in the high vocal center (HVC) from which we can conclude that the model fit the data well with appropriate complexity. Figure 18 indicates that the normality of residuals is acceptable despite the deviations at the tails.

5 Discussion and Conclusions

The various models that were fitted for the signal intensity in the 3 regions, the high vocal center, RA and $area\_X$ all yielded different results and this section provides an overview of these results. The model of Van der Linden et al, (2002) for the signal intensity in the RA region indicates that there is a treatment effect which is shown by the significant difference in the curve shape parameter. There is also evidence of substantial between-bird variability in their maximal intensity and time to reach half of the maximum intensity. Similarly, the fractional polynomials model also indicates a treatment effect because of the significance of the baseline group effect. There is also a difference in the time effects for example the treatment group has an additional effect over the control group so the increase in SI_HVC with log(time) is steeper in the treatment group and this is the same for the square root of time. Adaptive Gaussian quadrature and Laplace approximation give different results from the non-adaptive Gaussian quadrature but since the model employing the adaptive Gaussian quadrature is a better fitting model, there is therefore no detected treatment effect and the is supported by the model parameter estimates.
From the two-stage approach of the model of Van der Linden, for the SI_RA outcome, there is no significant group difference in the estimated maximal signal intensity, the curve shape parameter as well as the time to reach half of the maximum signal intensity. Therefore, there is no evidence of a treatment effect. For the outcome SI_Area_X, there is a significant group difference in the maximal signal intensity of the control and testosterone-treated birds. There is, however, insufficient evidence to claim a group effect on the curve shape parameter as well as on the time required to reach half of the maximal intensity. As discussed in the above subsection, the bi-exponential model produced no interpretable results and all attempts to improve its fit were in vain. The Gaussian model showed no treatment effect on the peak height (amplitude), peak position and the curve spread. This is indicated by the non-significance of the group effect on each of these parameters. There is however substantial between-bird variability arising from the spread of the curve suggesting that the width of the curves varies a lot among the birds. There is evidence of a treatment effect in the Gompertz model and this is the case because there is significant group differences in the location parameter as well as in the growth rate and there is also a lot of variability in the location parameter between the birds. The logistic model indicated a significant difference in the shape parameter of the two groups but the growth rate was not significantly in the two groups. However, there is a lot of variability in the shape and growth rate between the birds. Finally, the Ratkowsky model suggested no treatment effect whatsoever because the treatment had no effect on the maximum value, the midpoint of the curve and the growth rate of the birds. The between-bird deviations are quite pronounced in the maximum value and the growth rate.

5.1 Possible limitations of the methods used

Even with convergence, each of the approaches applied possesses limitations and some of these include: Adaptive Gaussian quadrature is computationally more intense because the quadrature points and weights have to be updated at each iteration of parameter estimation, this can also take longer. While non-adaptive Gaussian quadrature can be faster since it skips the extra calculations needed for adaptive centering, this speed comes at the cost of potentially large errors in parameter estimation. Furthermore, using too few points can produce inaccurate or unstable estimates, while using many points increases computation time dramatically. The Laplace method is computationally efficient and often faster than quadrature, but this comes at the expense of accuracy when the model or data do not meet its assumptions.

The Newton-Raphson method is sensitive to starting values. Poor starting values can lead to non-convergence or convergence to local, rather than global, maxima. This method relies on the calculation and inversion of the Hessian (second derivative) matrix. If the Hessian is near-singular or not positive definite (which often happens with complex or poorly scaled models), the algorithm may fail or produce unreliable results. Even with line search, if the step size is not chosen well, the algorithm can oscillate or take excessively small steps, slowing convergence. Also, for models with many parameters, repeatedly calculating and inverting the Hessian can be computationally expensive. In Newton-Raphson ridge, the ridge modification adds a small value to the diagonal of the Hessian to improve invertibility, but this can introduce bias into parameter estimates, especially if the ridge parameter is not chosen carefully. Furthermore, it is only a partial solution because while it can help with singularity or near-singularity of the Hessian matrix, it does not solve problems related to poor starting values or convergence to local maxima. Quasi-Newton methods approximate the Hessian matrix rather than computing it directly. This can lead to slower convergence, especially for highly non-linear problems like the one we are presented with. Just like Newton-Raphson, quasi-Newton methods can struggle to converge or may converge to local optima if starting values are poor. Furthermore, because the Hessian is only approximated, more iterations are often needed to reach convergence compared to the full Newton-Raphson method. Also for highly non-linear models or those with complex likelihood surfaces, the approximation may not be accurate enough, leading to convergence issues or suboptimal solutions.

Convergence can be hampered by the high sensitivity of non-linear mixed models to initial values, particularly when little is known about the parameters. For non-linear mixed models, it is therefore advised to select suitable initial values. In addition to careful initialization, researchers should think about how computational strategies, model complexity, and estimation techniques interact. As demonstrated throughout this thesis, the choice of integral approximation and optimization technique can substantially affect both convergence and the reliability of parameter estimates.

6 Ethical thinking, societal relevance, and stakeholder awareness

6.1 Ethical thinking

Ethical considerations include maintaining integrity and ensuring transparency in all research phases. In this research, the context of the original study along with its objectives were preserved in order to maintain the integrity of the study. In an effort to ensure transparency, there is clear documentation of all methods, this also enables reproducibility of all results.

6.2 Societal relevance

This study aims to explore the statistical and computational challenges that arise when fitting non-linear mixed-effects models and by addressing some of these challenges, this research increases flexibility and reduces the computational complexity of fitting non-linear mixed-models. This study also creates better conditions for the success of more in-silico experiments thereby reducing the need for extensive animal or human testing and hence minimizing impact on living organisms. Additionally, the methodological contributions of this research have direct applications in fields such as pharmacokinetics, ecology and many other real-world problems. The project therefore supports better scientific understanding and more informed decision-making.

6.3 Stakeholder awareness

Key stakeholders in this research include researchers, data subjects (such as, patients) and policy makers. Efforts were made to engage with relevant literature to ensure that the models and methods used were appropriate for the data and research questions at hand. Researchers stand to gain from the scientific knowledge and exploration of the challenges that come with fitting non-linear mixed-models. By improving the flexibility of non-linear mixed-effects models, this project ensures that the patient-level differences will be understood better and accounted for leading to more precise predictions about the effects of a drug in different people thereby, supporting personalized medicine and improving the quality of life of patients. By addressing the numerical and statistical barriers faced when using NLMMs, policy makers stand to gain from the reliability of predictions and recommendations which supports evidence-based decisions about societal needs such as public health interventions.

7 Ideas for future research

In the future, researchers could look into spline-based models which offer more flexibility than fractional polynomials to capture complex, non-linear relationships in the data [26], [27], [28]. Another promising approach is the use of Bayesian methods, which give researchers a natural framework for quantifying uncertainty in both parameter estimation and predictions and enable them to incorporate prior information. This contrasts with conventional frequentist approaches that depend on maximum likelihood estimation and linearization. Bayesian methods are especially useful when working with very complex models or small sample sizes [29].

References

[1] J. C. Pinheiro and D. M. Bates, "Approximations to the loglikelihood function in the nonlinear mixed effects model," Department of Statistics, University of Wisconsin, Madison, Tech. Rep., 1995, Technical Report.

[2] M. Davidian and D. Giltinan, *Nonlinear Models for Repeated Measurement Data*. Chapman and Hall/CRC, 1995.

[3] L. Sheiner and S. Beal, "Evaluation of methods for estimating population pharmacokinetic parameters. i. michaelis-menten model: Routine clinical pharmacokinetic data," *Journal of Pharmacokinetics and Biopharmaceutics*, vol. 8, no. 6, pp. 553–571, 1980.

[4] M. Lindstrom and D. Bates, "Nonlinear mixed effects models for repeated measures data," *Biometrics*, pp. 673–687, 1990.

[5] R. Wolfinger and X. Lin, "Laplace's approximation for nonlinear mixed models," *Biometrika*, vol. 84, no. 2, pp. 291–295, 1997.

[6] "Nonlinear mixed-effects model." (2020), [Online]. Available: https://en.wikipedia.org/wiki/Nonlinear_mixed-effects_model.

[7] "Nonlinear multilevel mixed-effects models - stata." (2022), [Online]. Available: https://www.stata.com/features/overview/nonlinear-multilevel-mixed-effects-models/.

[8] G. Molenberghs and G. Verbeke, *Models for Discrete Longitudinal Data* (Springer Series in Statistics). New York: Springer, 2005, ISBN: 9780387289790.

[9] X. Liu, "Chapter 8 - generalized linear mixed models on nonlinear longitudinal data," in *Methods and Applications of Longitudinal Data Analysis*, X. Liu, Ed., Oxford: Academic Press, 2016, pp. 243–279, ISBN: 978-0-12-801342-7. DOI: https://doi.org/10.1016/B978-0-12-801342-7.00008-3. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780128013427000083.

[10] J. Stoer, R. Bartels, W. Gautschi, R. Bulirsch, and C. Witzgall, *Introduction to Numerical Analysis* (Texts in Applied Mathematics). Springer New York, 2010, pp. 150–159, ISBN: 9781441930064. [Online]. Available: https://books.google.be/books?id=vUmIcgAACAAJ.

[11] A. Quarteroni, R. Sacco, and F. Saleri, *Numerical Mathematics* (Texts in Applied Mathematics), 2nd. Berlin, Heidelberg: Springer, 2007, vol. 37, ISBN: 978-3-540-34658-6. DOI: 10.1007/978-3-540-34658-6.

[12] U. of Washington Department of Statistics, *The Newton Raphson algorithm for function optimization*, https://www.stat.washington.edu/adobra/classes/536/Files/week1/newtonfull.pdf, Accessed: 2025-05-31, n.d.

[13] J. E. Dennis and R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations* (Classics in Applied Mathematics). Englewood Cliffs, New Jersey: Prentice-Hall, 1983, Originally published by Prentice-Hall; reprinted by SIAM.

[14] R. Fletcher, *Practical Methods of Optimization*, 2nd. John Wiley & Sons, 1987.

[15] F. S. Kitaura and T. A. Enßlin, "Bayesian reconstruction of the cosmological large-scale structure: Methodology, inverse algorithms and numerical optimization," *Monthly Notices of the Royal Astronomical Society*, vol. 389, no. 2, pp. 497–544, Sep. 2008, ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2008.13341.x. eprint: https://academic.oup.com/mnras/article-pdf/389/2/497/2942479/mnras0389-0497.pdf. [Online]. Available: https://doi.org/10.1111/j.1365-2966.2008.13341.x.

[16] R. Fletcher and M. J. D. Powell, "A family of gradient methods for optimization," *The Computer Journal*, vol. 13, no. 4, pp. 413–420, 1970. DOI: 10.1093/comjnl/13.4.413.

[17] SAS Institute Inc., *Sas/stat® 14.1 user's guide: The nlmixed procedure*, Accessed: 2025-05-24, Cary, NC, 2023. [Online]. Available: https://support.sas.com/documentation/onlinedoc/stat/141/nlmixed.pdf.

[18] SAS Institute Inc. "The nlmixed procedure: Details - specifications and computing." Accessed: 2025-05-24. (2024), [Online]. Available: https://documentation.sas.com/doc/en/statug/15.2/statug_nlmixed_details15.htm#statug_nlmixed005544.

[19] University of South Carolina, *Chapter 13: Random effects models*, Lecture notes for STAT 770, n.d. [Online]. Available: https://people.stat.sc.edu/grego/courses/stat770/Lecture%20Notes/Chapter13.pdf.

[20] A. Van der Linden, M. Verhoye, V. Van Meir, *et al.*, "In vivo manganese-enhanced magnetic resonance imaging reveals connections and functional properties of the songbird vocal control system," *Neuroscience*, vol. 112, no. 2, 2002, ISSN: 0306-4522. DOI: 10.1016/S0306-4522(02)00070-2. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306452202000702.

[21] E. A. Brenowitz, M. D. Beecher, and K. W. Nordeen, "Neural architecture of the song control system in passerine birds: Seasonal plasticity and aging," *Journal of Neuroscience*, vol. 17, no. 18, pp. 6001–6010, 1997. DOI: 10.1523/JNEUROSCI.17-18-06001.1997.

[22] V. Van Meir, K. Verhoeven, J. Balthazart, and G. F. Ball, "Relation between hormonal and neuronal markers of song behavior in an animal model of psychiatric disorders," *European Journal of Neuroscience*, vol. 20, no. 2, pp. 421–432, 2004. DOI: 10.1111/j.1460-9568.2004.03492.x.

[23] J. Serroyen, G. Molenberghs, M. Verhoye, V. Van Meir, and A. Van der Linden, "Dynamic manganese-enhanced mri signal intensity processing in songbirds," *NeuroImage*, vol. 26, no. 2, pp. 593–600, 2005. DOI: 10.1016/j.neuroimage.2005.02.026.

[24] G. Verbeke and G. Molenberghs, *Linear Mixed Models for Longitudinal Data* (Springer Series in Statistics). New York: Springer, 2000, ISBN: 0-387-98969-9.

[25] H. Development, *Curveexpert professional*, version 2.7.3, 2025. [Online]. Available: https://www.curveexpert.net/products/curveexpert-professional/.

[26] K. Kagerer, "A hat matrix for monotonicity constrained B-spline and P-spline regression," University of Regensburg, Department of Economics, University of Regensburg Working Papers in Business, Economics and Management Information Systems 484, Mar. 2015. [Online]. Available: https://ideas.repec.org/p/bay/rdwiwi/31450.html.

[27] I. G. Burova, "Local splines and the least squares method," *WSEAS Transactions on Systems*, vol. 23, 2024, ISSN: 2224-2678. DOI: 10.37394/23202.2024.23.21. [Online]. Available: https://wseas.com/journals/systems/2024/a425102-014(2024).pdf.

[28] I. S. Pop, V. Pop, S. Cobzac, and C. Sârbu, "Use of weighted least-squares splines for calibration in analytical chemistry," *Journal of Chemical Information and Computer Sciences*, vol. 40, no. 1, pp. 91–98, 2000, ISSN: 0095-2338. DOI: 10.1021/ci990321z. [Online]. Available: https://pubs.acs.org/doi/10.1021/ci990321z.

[29] S. Y. Lee, "Bayesian nonlinear models for repeated measurement data: An overview, implementation, and applications," *Mathematics*, vol. 10, no. 6, 2022, ISSN: 2227-7390. DOI: 10.3390/math10060898. [Online]. Available: https://www.mdpi.com/2227-7390/10/6/898.

A Checking the normality of residuals

A.1 Model of Van der Linden, (2002)

This model satisfies the assumption of normality of the residuals.

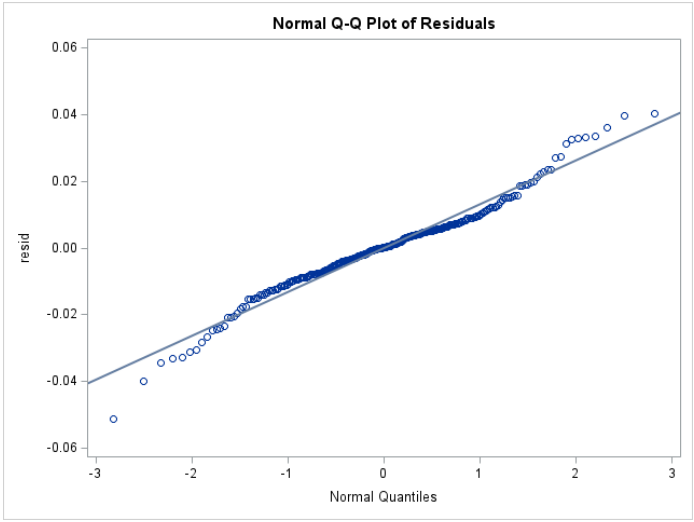

Figure 13: Normal Q-Q plot of residuals from the model of Van der Linden, (2002) for the signal intensity in the RA region.

A.2 Fractional Polynomials

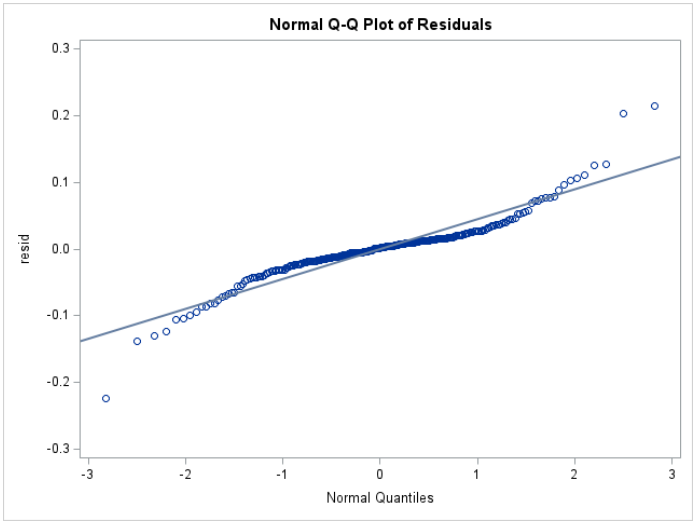

Figure 14: Normal Q-Q plot of residuals from the model with fractional polynomials for the signal intensity in the HVC region.

A.3 Gaussian model

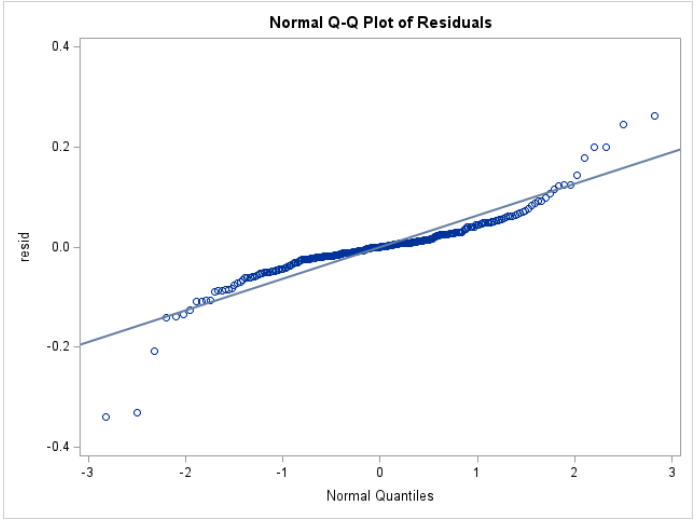

Figure 15: Normal Q-Q plot of residuals from the Gaussian model for the signal intensity in the HVC region.

A.4 Gompertz model

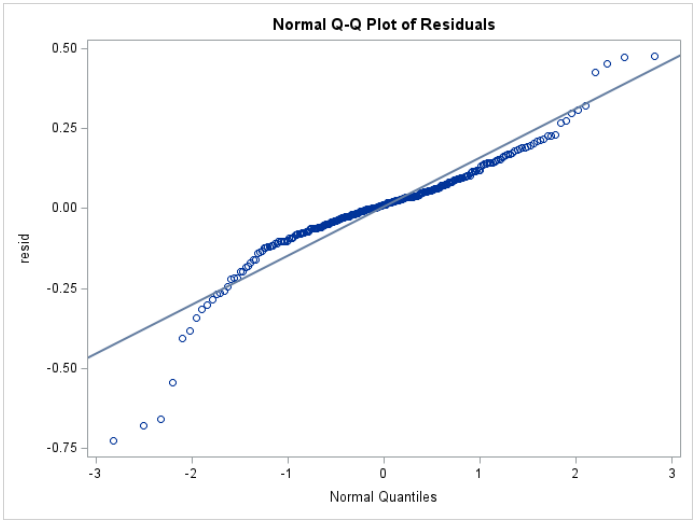

Figure 16: Normal Q-Q plot of residuals from the Gompertz model for the signal intensity in the HVC region.

A.5 Logistic model

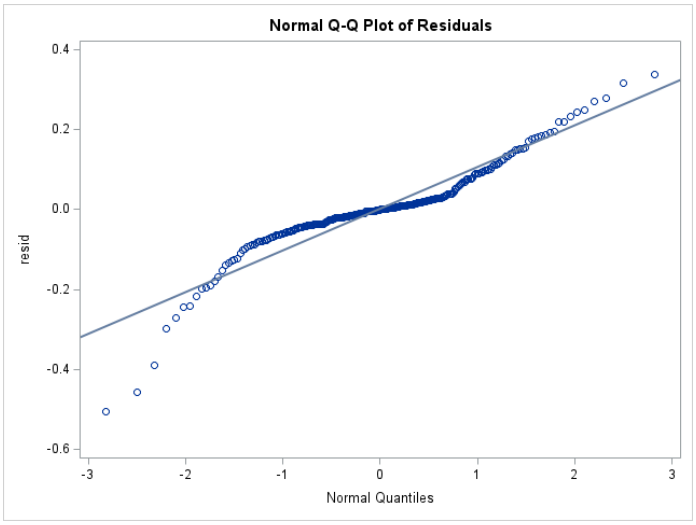

Figure 17: Normal Q-Q plot of residuals from the Logistic model for the signal intensity in the HVC region.

A.6 Ratkowsky model

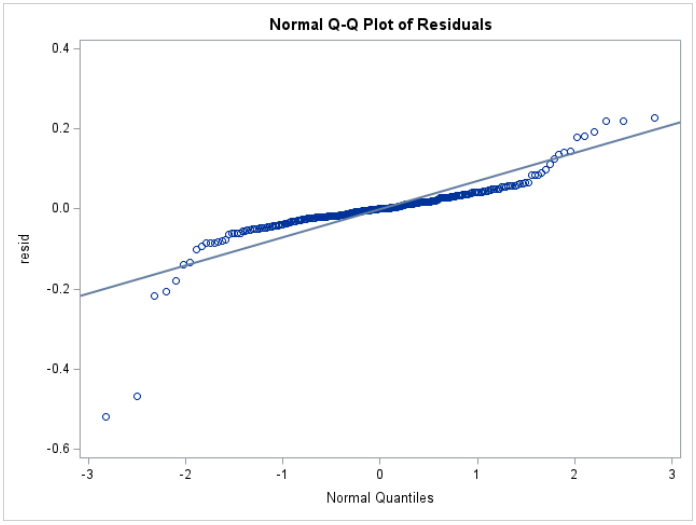

Figure 18: Normal Q-Q plot of residuals from the Ratkowsky model for the signal intensity in the HVC region.

SAS CODE

```
/*Van der Linden model*/
libname mydata "C:\Users\chris\OneDrive\Documents\BIOSTAT 2\SEM 2\Thesis\papers
\Songbird data";

data hulp2;
set mydata.vincent03;
if time <= 0 then delete;
if periode = 1 then delete;
run;
proc sort data=hulp2; by vogel; run;

/*For RA*/
proc nlmixed data=hulp2 tech=newrap qpoints=5 MAXITER=1000 cov; /*Adaptive*/
   title 'sigmoide voor testosteron';
   parms phim=0.4 eta=2  etadiff=0.5  tau=3 d11=1E-5 sigma2=0.001 d22=1E-5;
     teller = (phim + vm) * (time ** (eta + etadiff * groep));
     noemer = ((tau + t) ** (eta + etadiff * groep)) + (time ** (eta + etadiff * groep));
     gemid = teller/noemer;
   model si_ra ~ normal(gemid,sigma2);
   random vm t ~ normal([0, 0],[d11,0,d22]) subject=vogel out = mydata.eb;
   bounds d11 > 0, d22 > 0, sigma2 > 0;
   predict gemid out=pred_nlmixed;
run;

/*normality of residuals*/
data residuals;
  set pred_nlmixed;
  resid = si_ra - pred;
run;
/*Q-Q plot*/
proc univariate data=residuals normal;
  var resid;
  qqplot resid / normal(mu=est sigma=est);
  title " ";
run;

/*Goodness of fit*/
proc sort data=pred_nlmixed;
  by vogel time;
```

```
run;
data pred_nlmixed;
   set pred_nlmixed;
   label vogel = "Bird";
run;
proc sgpanel data=pred_nlmixed;
  panelby vogel / columns=4 rows=3;
  series x=time y=Pred / lineattrs=(thickness=2 color=red) name='fitted';
  scatter x=time y=si_ra / markerattrs=(symbol=circlefilled color=blue size=8) name='obs';
  colaxis label="Time";
  rowaxis label="SI_RA";
  title " ";
run;

/* For Area_X*/
proc nlmixed data=hulp2 MAXITER=500 tech=newrap noad qpoints=5;
   title 'sigmoide voor testosteron';
   parms phim=0.5 eta=1.5  etadiff=0.5  tau=4 d11=0.01 sigma2=0.01 d22=0.01;
     teller = (phim + vm) * (time ** (eta + etadiff * groep));
     noemer = ((tau + t) ** (eta + etadiff * groep)) + (time ** (eta + etadiff * groep));
     gemid = teller/noemer;
   model si_area_x ~ normal(gemid,sigma2);
   random vm t ~ normal([0, 0],[d11,0,d22]) subject=vogel out = mydata.eb;
    predict gemid out=mydata.predicted_values;
run;

/*Fractional polynomials*/
data hulp2;
set hulp2;
time2=time*time;
time3=time*time*time;
run;
   proc nlmixed data=hulp2 maxiter=5000 tech=newrap noad qpoints=5;/*Non-adaptive*/
   parms beta0=2.15 beta1=2.41 beta2=-7.91 beta3=5.47
         beta0_g=0 beta1_g=0 beta2_g=0 beta3_g=0
         d00=0.77 d11=1.10 d22=0.97 d33=1 sigma2=0.015;
   linpred = (beta0 + beta0_g * groep)
             + (beta1 + beta1_g * groep) * time
             + (beta2 + beta2_g * groep) * time2
             + (beta3 + beta3_g * groep) * time3
```

```
            + b0 + b1 * time + b2 * time2 + b3 * time3;
   model si_hvc ~ normal(linpred, sigma2);
   random b0 b1 b2 b3 ~ normal([0,0,0,0], [d00,0,d11,0,0,d22,0,0,0,d33]) subject=vogel;
   predict linpred out=fit_diag;
run;
/**Adaptive gaussian quadrature**/
    proc nlmixed data=hulp2 maxiter=5000 tech=quanew qpoints=5;/*with groep effect*/
   parms beta0=2.15 beta1=2.41 beta2=-7.91 beta3=5.47
         beta0_g=0 beta1_g=0 beta2_g=0 beta3_g=0
         d00=0.77 d11=1.10 d22=0.97 d33=1 sigma2=0.015;
   linpred = (beta0 + beta0_g * groep)
            + (beta1 + beta1_g * groep) * time
            + (beta2 + beta2_g * groep) * time2
            + (beta3 + beta3_g * groep) * time3
            + b0 + b1 * time + b2 * time2 + b3 * time3;
   model si_hvc ~ normal(linpred, sigma2);
   random b0 b1 b2 b3 ~ normal([0,0,0,0], [d00,0,d11,0,0,d22,0,0,0,d33]) subject=vogel;
   predict linpred out=fit_diag;
run;

/*log-sqrt model*/
proc nlmixed data=hulp2 maxiter=5000 tech=newrap qpoints=5;/*adaptive gaussian quadrature*/
   title 'Log/Square-root transformation model with group effects';
   parms alpha0=2.0 alpha1=0 lambda0=0.5 lambda1=0 delta0=0.1 delta1=0
         d_a=0.1 d_l=0.1 d_d=0.1 sigma2=0.01;
     /* Create transformed time variables */
   logtime = log(time);
   sqrttime = sqrt(time);

   linpred = (alpha0 + alpha1*groep + a_i)
           + (lambda0 + lambda1*groep + l_i) * logtime
           + (delta0 + delta1*groep + d_i) * sqrttime;

   model si_hvc ~ normal(linpred, sigma2);
   random a_i l_i d_i ~ normal([0,0,0], [d_a,0,d_l,0,0,d_d]) subject=vogel;
   predict linpred out=fit_diag_logsqrt;
run;

data fit_diag_logsqrt;
   set fit_diag_logsqrt;
```

```
   label vogel = "Bird";
run;
/* Plot fitted and observed values *//*good fit*/
proc sgpanel data=fit_diag_logsqrt;
   panelby vogel / columns=4 rows=3;
   scatter x=time y=SI_HVC;
   series x=time y=pred / lineattrs=(color=red thickness=2);
   rowaxis label="HVC";
   colaxis label="Time";
   title " ";
run;

/*normality of residuals*/
data residuals;
  set fit_diag_logsqrt;
  resid = si_hvc - pred;
run;
/*Q-Q plot*/
proc univariate data=residuals normal;
  var resid;
  qqplot resid / normal(mu=est sigma=est) nolegend odstitle="Normal Q-Q Plot of Residuals";
 run;

 /**FINAL GAUSSIAN MODEL**/
proc nlmixed data=hulp2 maxiter=5000 tech=nrridg qpoints=5 cov;
*proc nlmixed data=hulp2 maxiter=1000 tech=newrap qpoints=5 cov;
   parms a=1 b=3 c=1 sigma=0.5 d1=0.1 d2=0.1 d3=0.1 ag=0 bg=0 cg=0;
    pred = (a+ag*groep+b1) * exp(-(time - (b+bg*groep+b2))**2 / (2*((c*c)+cg*groep+b3)));
      model SI_HVC ~ normal(pred, sigma**2);
      random b1 b2 b3 ~ normal([0,0,0], [d1*d1,0,d2*d2, 0,0, d3*d3]) subject=vogel;
     predict pred out=gaussfit_hvc;
run;
data gaussfit_hvc;
   set gaussfit_hvc;
   label vogel = "Bird";
run;
/*Goodness of fit check*/
proc sgpanel data=gaussfit_hvc;/*good fit**/
   panelby vogel / columns=4 rows=3;
   scatter x=time y=SI_HVC;
```

```
   series x=time y=pred / lineattrs=(color=red thickness=2);
   rowaxis label="HVC";
   colaxis label="Time";
   title " ";
run;
/*normality of residuals*/
data residuals;
  set gaussfit_hvc;
  resid = si_hvc - pred;
run;
/*Q-Q plot*/
proc univariate data=residuals normal;
  var resid;
  qqplot resid / normal(mu=est sigma=est) nolegend odstitle="Normal Q-Q Plot of Residuals";
 run;

/*FINAL RATKOWSKY MODEL*/
proc nlmixed data=hulp2 maxiter=5000 tech=newrap qpoints=5;
    parms a=0.3 b=1 c=1 sigma=0.5 d1=0.1 d3=0.1 ag=0.3326 cg=-0.3502 bg=-0.01625;
 pred = (a+ag*groep + b1) / (1 + exp((b+bg*groep) - (c+cg*groep + b3)*time));
      model SI_HVC ~ normal(pred, sigma**2);
   random b1 b3 ~ normal([0,0],[d1*d1,0,d3*d3]) subject=vogel;
      predict pred out=logistic_fit;
run;

/*Goodness of fit check*/
data logistic_fit;
   set logistic_fit;
   label vogel = "Bird";
run;
proc sgpanel data=logistic_fit;
   panelby vogel / columns=4 rows=3;
   scatter x=time y=SI_HVC;
   series x=time y=pred / lineattrs=(color=red thickness=2);
   rowaxis label="HVC";
   colaxis label="Time";
   title " ";
run;
/*normality of residuals*/
data residuals;
```

```
  set logistic_fit;
  resid = si_hvc - pred;
run;
/*Q-Q plot*/
proc univariate data=residuals normal;
  var resid;
  qqplot resid / normal(mu=est sigma=est) nolegend odstitle="Normal Q-Q Plot of Residuals";
 run;

/**Final Gompertz model**/
proc nlmixed data=hulp2 maxiter=10000 tech=nrridg noad qpoints=5;
   parms a=1 b=1 c=1 sigma=0.5 bg=0 cg=0 d1=0.1;
      pred = a * exp(-exp((b+bg*groep+b1) - (c+cg*groep)*time));
      model SI_HVC ~ normal(pred, sigma**2);
   random b1  ~ normal([0],[d1*d1]) subject=vogel;
   predict pred out=gompertz_fit;
run;

/*Goodness of fit*/
data gompertz_fit;
   set gompertz_fit;
   label vogel = "Bird";
run;

proc sgpanel data=gompertz_fit;
   panelby vogel / columns=4 rows=3;
   scatter x=time y=SI_HVC;
   series x=time y=pred / lineattrs=(color=red thickness=2);
   rowaxis label="HVC";
   colaxis label="Time";
   title " ";
run;
/*normality of residuals*/
data residuals;
  set gompertz_fit;
  resid = si_hvc - pred;
run;
/*Q-Q plot*/
proc univariate data=residuals normal;
  var resid;
```

```
  qqplot resid / normal(mu=est sigma=est) nolegend odstitle="Normal Q-Q Plot of Residuals";
 run;

 /*FINAL LOGISTIC MODEL*/
proc nlmixed data=hulp2 maxiter=5000 tech=newrap noad qpoints=5;
 parms a=2.5 b=1 c=0.5 bg=0 cg=0 sigma=0.5 d1=0.1 d2=0.1;
      pred = a / (1 + (b+bg*groep+b1) * exp((-c+cg*groep+b2) * time));
      model SI_HVC ~ normal(pred, sigma**2);
   random b1 b2 ~ normal ([0,0], [d1*d1,0,d2*d2]) subject = vogel;
   predict pred out=logistic_fit;
run;

/*Goodness of fit*/
data logistic_fit;
   set logistic_fit;
   label vogel = "Bird";
run;
proc sgpanel data=logistic_fit;
   panelby vogel / columns=4 rows=3;
   scatter x=time y=SI_HVC;
   series x=time y=pred / lineattrs=(color=red thickness=2);
   rowaxis label="HVC";
   colaxis label="Time";
   title " ";
run;
/*normality of residuals*/
data residuals;
  set logistic_fit;
  resid = si_hvc - pred;
run;
/*Q-Q plot*/
proc univariate data=residuals normal;
  var resid;
  qqplot resid / normal(mu=est sigma=est) nolegend odstitle="Normal Q-Q Plot of Residuals";
 run;

/*TWO-STAGE APPROACHES*/
/*CUBIC POLYNOMIAL MODEL*/
proc nlin data=hulp2 method=marquardt;
    by vogel;
```

```
    parms beta0=2.15 beta1=2.41
          beta2=-7.91 beta3=5.47;
    pred = beta0
           + beta1*time
           + beta2*time**2
           + beta3*time**3;
    model si_hvc = pred;
    ods output ParameterEstimates=Stage1_Params;
  run;
/***********STAGE 2*************/
proc transpose data=Stage1_Params
               out=BirdParams(drop=_NAME_)
               prefix=beta_;
  by vogel;
  id Parameter;
  var Estimate;
run;
proc sort data=hulp2(keep=vogel groep) nodupkey; by vogel; run;
data BirdParams;
  merge BirdParams hulp2(keep=vogel groep);
  by vogel;
run;
proc print data=BirdParams (obs=10); run;

/*Fit linear model to the parameter estimates*/
proc mixed data=BirdParams;
  class groep;
  model beta_beta0 = groep / solution;
run;
proc mixed data=BirdParams;
  class groep;
  model beta_beta1 = groep / solution;
run;
proc mixed data=BirdParams;
  class groep;
  model beta_beta2 = groep / solution;
run;
proc mixed data=BirdParams;
  class groep;
  model beta_beta3 = groep / solution;
```

```
run;
/* Compute the 4×4 covariance and correlation matrices */
proc corr data=BirdParams cov outp=CovCorr noprint;
  var beta_beta0 beta_beta1 beta_beta2 beta_beta3;
run;
proc print data=CovCorr noobs label;
  where _TYPE_ = 'COV';
  var beta_beta0 beta_beta1 beta_beta2 beta_beta3;
  title "Covariance Matrix of Bird-Level Parameters";
run;

/**FRACTIONAL POLYNOMIAL MODEL**/
/*Stage 1 no group effect*/
proc nlin data=hulp2 method=marquardt;
  by vogel; /* one fit per bird */
  parms alpha0=2.0 lambda0=0.5 delta0=0.1;
  logtime = log(time);
  sqrttime = sqrt(time);
  pred = alpha0 + lambda0*logtime + delta0*sqrttime;
  model si_hvc = pred;
  ods output ParameterEstimates=Stage1_Params;
run;

/***********STAGE 2*************/
proc transpose data=Stage1_Params out=BirdParams(drop=_NAME_);
  by vogel;
  id Parameter;
  var Estimate;
run;
proc sort data=hulp2(keep=vogel groep) nodupkey; by vogel; run;
data BirdParams;
  merge BirdParams hulp2(keep=vogel groep);
  by vogel;
run;

/*Fit linear model to the parameter estimates*/
proc mixed data=BirdParams;
  class groep;
  model alpha0 = groep / solution;
run;
```

```
proc mixed data=BirdParams;
  class groep;
  model lambda0 = groep / solution;
run;
proc mixed data=BirdParams;
  class groep;
  model delta0 = groep / solution;
run;
proc corr data=BirdParams cov outp=Covariance noprint;
  var alpha0 lambda0 delta0;
run;
proc print data=Covariance;
run;

/***MODEL FOR RA (Model of Van der Linden)***/
/*Stage 1 no groep effect*/
proc nlin data=hulp2 method=marquardt;
  by vogel;
  parms phim=0.4 eta=2 tau=3;
  pred = (phim) * (time**(eta))
         / ((tau)**(eta) + time**(eta));
  model si_ra = pred;
  ods output ParameterEstimates=Stage1_Params;
run;
/*Stage 2*/
proc transpose data=Stage1_Params out=Stage1_wide prefix=param_;
  by vogel;
  id Parameter;
  var Estimate;
run;
data Stage1_analysis;
  merge Stage1_wide (in=a)
        hulp2 (keep=vogel groep);
  by vogel;
  if a;
run;
proc mixed data=Stage1_analysis;  class groep;  model param_phim = groep / solution;run;
proc mixed data=Stage1_analysis;  class groep;  model param_tau = groep / solution;run;
run;
```

```
/*Bi-exponential model*/
/*For SI_Area_X */
proc nlmixed data=hulp2 maxiter=500 method=firo qpoints=5 noad;
  parms beta1=0.5 beta2=0.0 beta3=-0.5 beta4=0 sigma=0.7  beta6=0 beta8=0
   d3=0.1 d4=0.1;
  pred = exp(beta1) * exp(-exp(beta2) * time)
       + exp(beta3+beta6*groep+b3) * exp(-exp(beta4+beta8*groep+b4) * time);
  model SI_Area_X ~ normal(pred, sigma**2);
 random b3 b4 ~ normal([0,0], [d3*d3,0,d4*d4]) subject=vogel;
run;

proc nlmixed data=hulp2 maxiter=500 method=firo qpoints=5 noad;
  parms beta1=0.5 beta2=0.0 beta3=-0.5 beta4=0 sigma=0.7  beta5=0 beta6=0
   d3=0.1 d1=0.1;
  pred = exp(beta1+beta5*groep+b1) * exp(-exp(beta2) * time)
       + exp(beta3+beta6*groep+b3) * exp(-exp(beta4) * time);
  model  SI_Area_X ~ normal(pred, sigma**2);
 random b1 b3 ~ normal([0,0], [d1*d1,0,d3*d3]) subject=vogel;
run;

/*For HVC*/
proc nlmixed data=hulp2 maxiter=500 method=firo qpoints=5 noad;
  parms beta1=0.5 beta2=0.0 beta3=-0.5 beta4=0 sigma=0.7  beta6=0 beta8=0
   d3=0.1 d4=0.1;
  pred = exp(beta1) * exp(-exp(beta2) * time)
       + exp(beta3+beta6*groep+b3) * exp(-exp(beta4+beta8*groep+b4) * time);
  model SI_HVC ~ normal(pred, sigma**2);
 random b3 b4 ~ normal([0,0], [d3*d3,0,d4*d4]) subject=vogel;
run;

proc nlmixed data=hulp2 maxiter=500 method=firo qpoints=5 noad;
  parms beta1=0.5 beta2=0.0 beta3=-0.5 beta4=0 sigma=0.7  beta5=0 beta6=0
   d3=0.1 d1=0.1;
  pred = exp(beta1+beta5*groep+b1) * exp(-exp(beta2) * time)
       + exp(beta3+beta6*groep+b3) * exp(-exp(beta4) * time);
  model SI_HVC ~ normal(pred, sigma**2);
 random b1 b3 ~ normal([0,0], [d1*d1,0,d3*d3]) subject=vogel;
run;

/*For RA*/
```

```
proc nlmixed data=hulp2 maxiter=500 method=firo qpoints=5 noad;
  parms beta1=0.5 beta2=0.0 beta3=-0.5 beta4=0 sigma=0.7  beta6=0 beta8=0
   d3=0.1 d4=0.1;
  pred = exp(beta1) * exp(-exp(beta2) * time)
       + exp(beta3+beta6*groep+b3) * exp(-exp(beta4+beta8*groep+b4) * time);
  model SI_RA ~ normal(pred, sigma**2);
 random b3 b4 ~ normal([0,0], [d3*d3,0,d4*d4]) subject=vogel;
run;

proc nlmixed data=hulp2 maxiter=500 method=firo qpoints=5 noad;
  parms beta1=0.5 beta2=0.0 beta3=-0.5 beta4=0 sigma=0.7  beta5=0 beta6=0
   d3=0.1 d1=0.1;
  pred = exp(beta1+beta5*groep+b1) * exp(-exp(beta2) * time)
       + exp(beta3+beta6*groep+b3) * exp(-exp(beta4) * time);
  model  SI_RA ~ normal(pred, sigma**2);
 random b1 b3 ~ normal([0,0], [d1*d1,0,d3*d3]) subject=vogel;
run;
```