Multivariate Nonlinear Mixed Model to Analyze Longitudinal Image Data: MRI Study of Early Brain Development

Shun Xu, Martin Styner, John Gilmore, Joseph Piven
University of North Carolina at Chapel Hill
Chapel Hill, NC, USA
xushun@email.unc.edu

Guido Gerig
University of Utah
Salt Lake City, Utah, USA

Abstract

With great potential in studying neuro-development, neuro-degeneration, and the aging process, longitudinal image data is gaining increasing interest and attention in the neuroimaging community. In this paper, we present a parametric nonlinear model to statistically study multivariate longitudinal data with asymptotic properties. We demonstrate our preliminary results in a combined study of two longitudinal neuroimaging data sets of early brain development to cover a wider time span and to gain a larger sample size. Such combined analysis of multiple longitudinal image data sets has not been conducted before and presents a challenge for traditional analysis methods. To our knowledge, this is the first multivariate nonlinear longitudinal analysis to study early brain development. Our methodology is generic in nature and can be applied to any longitudinal data with nonlinear growth patterns that can not easily be modeled by linear methods.

1. Introduction

With the advances of medical imaging and data acquisition techniques, longitudinal image data, in which subjects are scanned and measured repeatedly over time, become available. Such longitudinal data are of increasing interest to the neuroimaging community due to their great potential in studying processes like neuro-development, neuro-degeneration, and aging. Recently, special effects have been put into studying early human brain growth in the first few years of life [9], because it is the most dynamic and perhaps the most important phase of postnatal brain development. The ability to study brain growth at a period when it undergoes critical modication is essential to shed light on our understanding of brain development.

Traditionally, cross-sectional studies were conducted to compare the mean measurements of two or more age groups and to find out whether there are any significant differences among the age groups [13, 1]. However, this type of study does not provide a growth model that integrates the continuum of time, nor does it tell us the trend of how individuals and populations change over time.

Researchers have tried to apply traditional or nonparametric regression methods to retrieve growth information of the population [4, 7], which is a reasonable way to approximates the true population growth. But we need to realize that the effect of growing or aging is indeed an inherently within-individual effect, and the true hidden population growth trajectory should really be the average of all individual growth trajectories. Moreover, traditional regression is not suitable for the analysis of longitudinal data because repeated observations from the same individual in longitudinal data violate the Gauss-Markov assumption of independence. Ignoring the correlation between repeated measurements when it exists in longitudinal data could result in less precise estimation and incorrect inference about the growth pattern [8]. Indeed, compared to cross-sectional data, in which each subject is measured only once, longitudinal data possess some unique properties that pose challenges for analysis:

1. **Temporal correlation between repeated measurements.** It is a typical feature in longitudinal studies that repeated measurements of any given subject are correlated, because they are measured from the same subject over time. Moreover, the value of temporal correlation depends on the distance in time between measurements. For example, we would expect two measurements that are close in time to be more alike than those that are far apart. Generally speaking, the closer the two repeated observations are in time, the larger the correlation, and vice versa.

2. **Missing data.** In clinical studies, scheduled imaging sessions are often missed or image data need to be excluded due to insufficient quality. Moreover, as illustrated in Fig. 1, we can combine several longitudinal data sets, each of which provide longitudinal samples

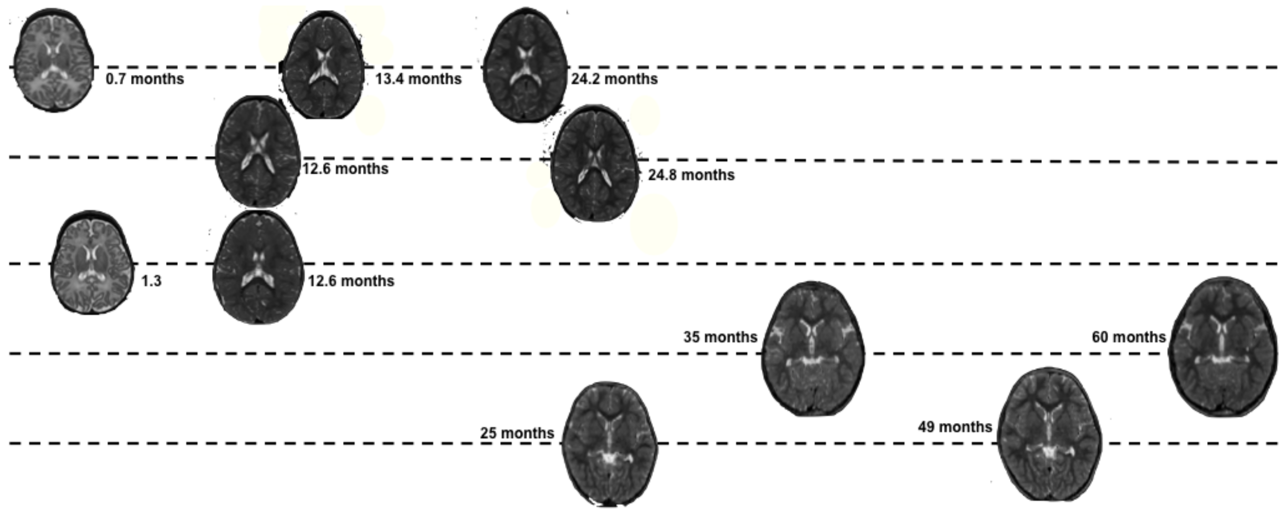

Figure 1. An illustration of properties of longitudinal data. Data are selected from a combination of two longitudinal studies. Repeatedly scanned MR T2 brain images of 5 children between birth to 5 years old are shown. We can observe the following properties: (1) temporal correlation: similarity between repeated scanned images; (2) missing data: no children have a complete set of scans that span 5 years; (3) uneven sampling of time points; (4) multivariate features; (5) Asymptotic growth pattern: growth of brain slows down after some time.

for a certain short period of life, to study longitudinal development of a wider time span or to gain a larger sample size. Such studies are bound to have missing data since time periods of different data sets do not necessarily overlap.

3. **Unbalanced or uneven sampling of time points.** Measuring schedules vary from subject to subject. It is unlikely to have an equally-spaced and complete clinical data set where each individual has the same time points with exactly the same time intervals. Temporal correlations, affected by this unbalanced property, also vary from subject to subject, which makes population modeling of temporal correlation challenging.

4. **Multivariate features and spatial correlation.** Multiple features can be computed from each image, such as the three dimensional size of the head, diffusion properties of white matter tracts, or cortical thickness at different locations. Correlations often exist among these features due to neighboring effects, thus we call them spatial correlations.

5. **Asymptotic growth pattern.** The asymptotic behavior is not a general longitudinal characteristic, but is prevailing in human growth data [3]. Polynomial and other parsimonious linear models may approximate a short period of human growth, but they are unsuitable when the growth curve asymptotes to an upper or lower bound [12]. Nonlinear models are required to study such longitudinal data as brain development in neuroimaging.

These unique properties of longitudinal data demand a more flexible and powerful model than traditional regression. Mixed-effects models [8] have received a great deal of attention because of the flexibility they offer in handling the unbalanced repeated measurements with missing data. Compared to linear mixed models, nonlinear mixed-effects models are considerably more difficult and computationally intensive to fit because its nonlinearity offers no close-form solution to the growth parameters [5]. In the following sections, we start off introducing the linear mixed model in section 2.1 and then extend it to the nonlinear case in section 2.2. A joint modeling scheme of nonlinear mixed models to statistically study longitudinal data with multiple features is then studied in section 2.3. Details of estimation and inference are given in section 2.4, and individual growth prediction are presented in section 2.5. In section 3, we apply the multivariate nonlinear longitudinal method to a combined study of two longitudinal image data sets whose time periods barely overlap, and then validate our result using a third data set whose time points were not measured in the previous two. Residual plots are also used for further validation.

Note that the multivariate nonlinear longitudinal method we use is generic in nature and thus can be applied to any longitudinal data with single or multiple features that exhibit nonlinear growth patterns. However, the purpose of this paper is not to provide a complete solution to the complex problem. Instead, this is work in progress that serves as a scientific exploration into applying multivariate nonlinear longitudinal methods to image data, and as a precursor to higher dimensional longitudinal shape analysis.

2. Method

The spirit of mixed models is to average individual growth trajectories rather than individual points such as that in traditional regression. An effective strategy is to first reduce the unbalanced repeated values into a few summaries, and then analyze each summary variable as a function of independent variables [8]. We illustrate this idea with linear mixed models.

2.1. Linear Mixed Models

The linear mixed model is a two-level model [8]. The first level is the individual level, which lets us think of a unique trajectory for each individual. Then, the summary variables of individual trajectories, e.g. intercepts, slopes, etc., are used to compute an average population growth, while their variances are considered to represent among-individual variability.

Individual (first stage) model: For the ith subject, if its measured time points are $t_{i1}, \cdots, t_{in_i}$, and the corresponding univariate features over time are $y_{i1}, \cdots, y_{in_i}$, then the model for individual i, $i = 1, \cdots, m$, is

$$y_{ij} = \beta_{0i} + \beta_{1i} t_{ij} + e_{ij}, j = 1, \cdots, n_i, \tag{1}$$

where β_{0i} and β_{1i} are the two summary variables describing the growth profile of individual i. We can see that no matter how many repeated measurements (e.g. n_i) each individual has, or how close or far apart the measurements are, the number of parameters used (e.g. β_{0i}, β_{1i}) to summarize individual trajectories is always the same (it is 2 in this illustrating case). Thus, the missing data and unbalanced properties of longitudinal data described in section 1 is appropriately handled.

Population (second stage) model: In the population level, we think of parameters of individual trajectories as random variables, varying around the "centered" average population parameters, e.g. mean intercept β_0 and mean slope β_1:

$$\boldsymbol{\beta_i} = \begin{bmatrix} \beta_{0i} \\ \beta_{1i} \end{bmatrix} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} b_{0i} \\ b_{1i} \end{bmatrix} = \boldsymbol{\beta} + \boldsymbol{b_i}. \tag{2}$$

Here b_{0i} and b_{1i} are called *random effects* describing how the intercept and slope for the ith subject deviate from their mean values. Doing simple substitution from Eqn. 1 and 2, we get:

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + b_{0i} + b_{1i} t_{ij} + e_{ij} = f(\boldsymbol{\beta}, \boldsymbol{b_i}, t_{ij}) + e_{ij}, \tag{3}$$

where $f(\boldsymbol{\beta}, \boldsymbol{b_i}, t_{ij})$ is a linear function of random effects $\boldsymbol{b_i}$. To write the individual i's entire response vector (i.e. n_i repeated measurements) into a matrix form, we let

$$\boldsymbol{y_i} = \begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{in_i} \end{bmatrix}, \boldsymbol{e_i} = \begin{bmatrix} e_{i1} \\ e_{i2} \\ \vdots \\ e_{in_i} \end{bmatrix}, \text{ and}$$

$$\boldsymbol{f}(\boldsymbol{\beta}, \boldsymbol{b_i}) = \begin{bmatrix} f(\boldsymbol{\beta}, \boldsymbol{b_i}, t_{i1}) \\ f(\boldsymbol{\beta}, \boldsymbol{b_i}, t_{i2}) \\ \vdots \\ f(\boldsymbol{\beta}, \boldsymbol{b_i}, t_{in_i}) \end{bmatrix} = \boldsymbol{X_i \beta} + \boldsymbol{Z_i b_i},$$

where $\boldsymbol{X_i}$ and $\boldsymbol{Z_i}$ are design matrices of the known constants. Then the individual response vector is written as

$$\boldsymbol{y_i} = \boldsymbol{f}(\boldsymbol{\beta}, \boldsymbol{b_i}) + \boldsymbol{e_i} = \boldsymbol{X_i \beta} + \boldsymbol{Z_i b_i} + \boldsymbol{e_i}. \tag{4}$$

It is often reasonable to assume that the measurement errors are i.i.d. $\boldsymbol{e_i} \sim \mathcal{N}(0, \sigma^2 \boldsymbol{I})$ and populations of intercepts and slopes are approximately normally distributed, i.e.

$$\begin{bmatrix} b_{0i} \\ b_{1i} \end{bmatrix} \sim \mathcal{N}(0, \boldsymbol{D}), \text{ where } \boldsymbol{D} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}.$$

The conditional distribution of repeated response vector $\boldsymbol{y_i}$ given $\boldsymbol{b_i}$ is thus given by

$$\boldsymbol{y_i} | \boldsymbol{\beta}, \boldsymbol{b_i}, \boldsymbol{D}, \sigma^2 \sim \mathcal{N}(\boldsymbol{f}(\boldsymbol{\beta}, \boldsymbol{b_i}), \sigma^2 \boldsymbol{I}) = \mathcal{N}(\boldsymbol{X_i \beta} + \boldsymbol{Z_i b_i}, \sigma^2 \boldsymbol{I}), \tag{5}$$

which basically says the condition mean of individual i is $E(\boldsymbol{y_i}|\boldsymbol{b_i}) = \boldsymbol{X_i \beta} + \boldsymbol{Z_i b_i}$ with variance $var(\boldsymbol{y_i}|\boldsymbol{b_i}) = \sigma^2 \boldsymbol{I}$. Thus, the marginal density of $\boldsymbol{y_i}$ turns out to be

$$p(\boldsymbol{y_i}|\boldsymbol{\beta}, \boldsymbol{D}, \sigma^2) = \int p(\boldsymbol{y_i}|\boldsymbol{\beta}, \boldsymbol{b_i}, \boldsymbol{D}, \sigma^2) p(\boldsymbol{b_i}) d\boldsymbol{b_i} \tag{6}$$

$$= \Phi(\boldsymbol{X_i \beta}, \boldsymbol{Z_i D Z_i'} + \sigma^2 \boldsymbol{I}). \tag{7}$$

Then the population growth parameters $\boldsymbol{\beta}$ and the variance components in $\boldsymbol{D}$ and σ^2 are estimated by maximizing the likelihood based on the marginal density of all $\boldsymbol{y_i}$. The reason the integral of the marginal density in Eqn. 6 can easily be expressed in a closed form as Eqn. 7 is the fact that $\boldsymbol{f}(\boldsymbol{\beta}, \boldsymbol{b_i})$ is a linear function of $\boldsymbol{b_i}$. For nonlinear mixed models, this nice close-form solution can not be obtained.

2.2. Nonlinear Mixed Models

Similar to linear mixed models, nonlinear mixed models also assume a unique trajectory for each individual, except that it is nonlinear with regard to the growth parameters $\boldsymbol{\beta_i}$. For example,

$$y_{ij} = \frac{\beta_{1i}}{1 + \beta_{2i} \exp(-\beta_{3i} t_{ij})} + e_{ij}, j = 1, \cdots, n_i, \tag{8}$$

where $\beta_{1i}, \beta_{2i}, \beta_{3i} > 0$. This is called the *logistic growth function*, one of the common models for the process of growth [14]. If one assumes that the growth rate of the response variable (e.g. y_{ij}) relative to the response value itself declines in a linear fashion with increasing growth, it may be shown that the response value at time t_{ij} may be represented by a logistic growth function of Eqn. 8. Here, β_{1i} controls the asymptote of growth, $\frac{\beta_{1i}}{1+\beta_{2i}}$ is the intercept, and β_{3i} is the "time scale" parameter characterizing how quickly the growth value increases from the initial value to the asymptote.

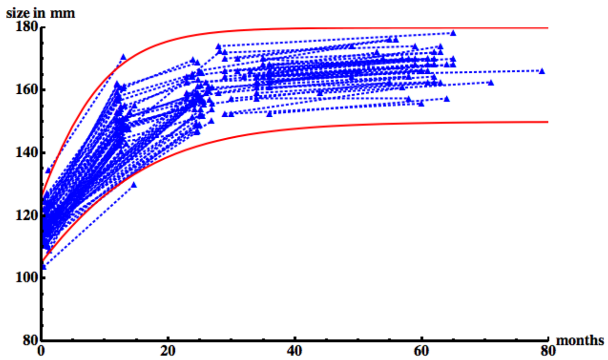

Figure 2. Illustration of individual growth trends. A spaghetti plot that connects repeated measurements of the same individual is shown. The two red upper and lower bound curves are generated by varying β_1 and β_3 only and with fixed β_2, which indicates population variance can be captured by varying only β_1 and β_3.

Similarly to the linear case, we think of individual growth parameters $\beta_{1i}, \beta_{2i}, \beta_{3i}$ as random variables, varying around the "centered" average population parameters:

$$\boldsymbol{\beta_i} = \begin{bmatrix} \beta_{1i} \\ \beta_{2i} \\ \beta_{3i} \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} b_{1i} \\ 0 \\ b_{3i} \end{bmatrix}. \qquad (9)$$

The distribution of random effects are modeled normally as

$$\boldsymbol{b_i} = \begin{bmatrix} b_{1i} \\ b_{3i} \end{bmatrix} \sim \mathcal{N}(0, \boldsymbol{D}), \text{ where } \boldsymbol{D} = \begin{bmatrix} \sigma_{11} & \sigma_{13} \\ \sigma_{31} & \sigma_{33} \end{bmatrix}.$$

Note that we could choose to exclude some random effects (e.g. set to zero) if the population of individual growth trends do not exhibit significant variation with regard to the corresponding growth parameters. As seen in the illustrative example of Fig. 2, the variation of individual trajectories can mostly be captured by β_{1i}, which controls the asymptote and intercept, and β_{3i} which controls the "growth rate". It is dangerous to include extra random effects in the model when there is little variation in them, because they would result in a non-full rank covariance matrix $\boldsymbol{D}$, and would pose challenges for parameter estimation. Doing simple substitution from Eqn. 8 and 9, we get:

$$y_{ij} = \frac{\beta_1 + b_{1i}}{1 + \beta_2 \exp(-(\beta_3 + b_{3i})t_{ij})} + e_{ij} = f(\boldsymbol{\beta}, \boldsymbol{b_i}, t_{ij}) + e_{ij}. \qquad (10)$$

The repeated measurement vector of individual i can again be written as

$$\boldsymbol{y_i} = \boldsymbol{f}(\boldsymbol{\beta}, \boldsymbol{b_i}) + \boldsymbol{e_i}, \text{ where } \boldsymbol{f}(\boldsymbol{\beta}, \boldsymbol{b_i}) = \begin{bmatrix} f(\boldsymbol{\beta}, \boldsymbol{b_i}, t_{i1}) \\ f(\boldsymbol{\beta}, \boldsymbol{b_i}, t_{i2}) \\ \vdots \\ f(\boldsymbol{\beta}, \boldsymbol{b_i}, t_{in_i}) \end{bmatrix}. \qquad (11)$$

As the nonlinearity of $\boldsymbol{b_i}$ causes the difficulty for integration in the marginal distribution in Eqn. 6, there are several methods to approximate the integral [6]. One of the natural way is to consider a linear approximation. A Taylor series expansion of $\boldsymbol{f}(\boldsymbol{\beta}, \boldsymbol{b_i})$ about the expected value of $\boldsymbol{b_i} = 0$ to linear terms leads to

$$\boldsymbol{y_i} \approx \boldsymbol{f}(\boldsymbol{\beta}, \boldsymbol{0}) + \boldsymbol{Z_i}(\boldsymbol{\beta}, \boldsymbol{0})\boldsymbol{b_i} + \boldsymbol{e_i}, \qquad (12)$$

where $\boldsymbol{Z_i}(\boldsymbol{\beta}, \boldsymbol{0})$ is a matrix of derivatives

$$\boldsymbol{Z_i}(\boldsymbol{\beta}, \boldsymbol{0}) = \left[\frac{\partial f(\boldsymbol{\beta}, \boldsymbol{b_i}, t_{ij})}{\partial \boldsymbol{b_i}}\right]_{b_i = E(b_i) = 0}.$$

Thus the integral in the marginal distribution is analytically calculable analogous to a linear mixed model, and can be approximated by a normal distribution:

$$\begin{aligned} p(\boldsymbol{y_i}|\boldsymbol{\beta}, \boldsymbol{D}, \sigma^2) &= \int p(\boldsymbol{y_i}|\boldsymbol{\beta}, \boldsymbol{b_i}, \boldsymbol{D}, \sigma^2) p(\boldsymbol{b_i}) d\boldsymbol{b_i} \\ &\approx \Phi(\boldsymbol{f}(\boldsymbol{\beta}, \boldsymbol{0}),\ \boldsymbol{\Sigma}_i), \end{aligned}$$

where $\boldsymbol{\Sigma_i} = \boldsymbol{Z_i}(\boldsymbol{\beta}, 0)\boldsymbol{D}\boldsymbol{Z_i}(\boldsymbol{\beta}, 0)' + \sigma^2 \boldsymbol{I}$.

2.3. Joint Modeling of Nonlinear Mixed Models

To study the relationship between two or more developmental processes, an approach of jointly modeling the random effects for different features can be adopted [11]. First, the average evolution of each feature is described. We still use our logistic growth model here as an illustrative example, other nonlinear functions could also be adopted:

$$\begin{aligned} y_{ij,X} &= \frac{\beta_{1,X} + b_{1i,X}}{1 + \beta_{2,X} \exp(-(\beta_{3,X} + b_{3i,X})t_{ij,X})} + e_{ij,X} \\ y_{ij,Y} &= \frac{\beta_{1,Y} + b_{1i,Y}}{1 + \beta_{2,Y} \exp(-(\beta_{3,Y} + b_{3i,Y})t_{ij,Y})} + e_{ij,Y} \\ y_{ij,Z} &= \frac{\beta_{1,Z} + b_{1i,Z}}{1 + \beta_{2,Z} \exp(-(\beta_{3,Z} + b_{3i,Z})t_{ij,Z})} + e_{ij,Z} \end{aligned}$$

The repeated response vector for each feature can be written similar to Eqn. 11 and 12. Concatenating response vectors of multiple features for individual i results in:

$$\begin{bmatrix} \boldsymbol{y_{i,X}} \\ \boldsymbol{y_{i,Y}} \\ \boldsymbol{y_{i,Z}} \end{bmatrix} \approx \begin{bmatrix} \boldsymbol{f_X}(\boldsymbol{\beta_X}, \boldsymbol{0}) \\ \boldsymbol{f_Y}(\boldsymbol{\beta_Y}, \boldsymbol{0}) \\ \boldsymbol{f_Z}(\boldsymbol{\beta_Z}, \boldsymbol{0}) \end{bmatrix} + \boldsymbol{Z_{i,A}} \begin{bmatrix} \boldsymbol{b_{i,X}} \\ \boldsymbol{b_{i,Y}} \\ \boldsymbol{b_{i,Z}} \end{bmatrix} + \begin{bmatrix} \boldsymbol{e_{i,X}} \\ \boldsymbol{e_{i,Y}} \\ \boldsymbol{e_{i,Z}} \end{bmatrix}, \qquad (13)$$

where

$$Z_{i,A} = \begin{bmatrix} Z_{i,X}(\beta_X, 0) & 0 & 0 \\ 0 & Z_{i,Y}(\beta_Y, 0) & 0 \\ 0 & 0 & Z_{i,Z}(\beta_Z, 0) \end{bmatrix}.$$

Then, by imposing a joint multivariate distribution on the all the random effects, the growth patterns of different features are associated:

$$b_i = \begin{bmatrix} b_{1i,X} \\ b_{1i,Y} \\ b_{1i,Z} \\ b_{3i,X} \\ b_{3i,Y} \\ b_{3i,Z} \end{bmatrix} \sim \mathcal{N}(0, D).$$

The elements in matrix D capture both the among-individual variability with regard to growth patterns within a single feature, as well as the association between the growth patterns of different features. Crucial questions such as association and evolution between growth patterns of different features are of critical interest for neuroimaging applications. These problems can be appropriately addressed in longitudinal analysis but cannot be answered in a traditional regression analysis framework.

Similar to that described in Sec. 2.2, the marginal distribution of the joint response vector follows a normal distribution approximately

$$p(y_{i,A}|\beta_A, D, \sigma^2) \approx \Phi(f_A,\ Z_{i,A} D Z'_{i,A} + \sigma^2 I), \quad (14)$$

where

$$\beta_A = \begin{bmatrix} \beta_X \\ \beta_Y \\ \beta_Z \end{bmatrix}, \ f_A = \begin{bmatrix} f_X(\beta_X, 0) \\ f_Y(\beta_Y, 0) \\ f_Z(\beta_Z, 0) \end{bmatrix}.$$

Estimation of population growth parameters β_A and variance-covariance parameteris in D and σ^2 can then be computed by maximizing the marginal distribution of all $y_{i,A}$.

2.4. Estimation and Inference

Let us assume the joint response vectors of multiple features for different individuals $y_{i,A}$ are independent across i, then the joint probability density function is

$$\begin{aligned} m(\theta) &= \prod_{i=1}^{m} \int p(y_{i,A}|\beta_A, b_i, D, \sigma^2) p(b_i) db_i \\ &= \prod_{i=1}^{m} p(y_{i,A}|\beta_A, D, \sigma^2) \\ &\approx \prod_{i=1}^{m} \Phi(f_A,\ Z_{i,A} D Z_{i,A} + \sigma^2 I), \end{aligned}$$

where θ represents the all the unknown parameters in β_A, D, and σ^2. In particular, the function

$$h(\theta) = -\log m(\theta) \quad (15)$$

is minimized over θ numerically in order to estimate θ. Here $h(\theta)$ is referred to as the objective function for optimization.

Because objective function $h(\theta)$ is not a simple quadratic function with regard to θ, we must solve for the solution iteratively. A Quasi-Newton optimization method [2] is adopted. For each Newton step, we compute a new estimation for θ

$$\theta^{(k+1)} = \theta^{(k)} - \alpha^{(k)} \delta^{(k)},$$

where

$$\delta^{(k)} = H^{-1}(\theta^{(k)}) g(\theta^{(k)}).$$

The matrix $H^{-1}(\theta)$ refers to the Hessian $\nabla^2 h(\theta)$ and the function $g(\theta)$ refers to the gradient vector $\nabla h(\theta)$. $\alpha^{(k)}$ is called the step length, which is determined by a local optimization of the function $h(\theta^{(k)} - \alpha^{(k)}\delta^{(k)})$, called a line search [10]. Also, the computation of Hessian matrix $H^{-1}(\theta^{(k)})$ is approximated using the BFGS method [2]. The above Newton step is iteratively executed until the convergence criterion is met, resulting an estimation $\hat{\theta}$.

It turns out that the inverse of the Hessian matrix of the log-likelihood evaluation at $\hat{\theta}$ describes the asymptotic distribution of the $\hat{\theta}$ about the actual parameter value θ, i.e.

$$\hat{\theta} \sim \mathcal{N}(\theta, H^{-1}(\hat{\theta})).$$

Knowing the parameter estimate $\hat{\theta}$ and the variance of its sampling distribution $\hat{V}_\theta = H^{-1}(\hat{\theta})$, it is straightforward to conduct hypothesis testing and determine whether certain parameter is statistically significant. We can specify appropriate matrices L to represent various questions of interest. For example, if we want to study whether the asymptote of feature X is statistically different than that of feature Y, we let the corresponding element of L to be 1 and -1, e.g. $L = (1, -1, 0, \cdots, 0)$. When L consists of a single row, a general t-statistic can be constructed as follows:

$$t = \frac{L\hat{\theta}}{\sqrt{L \hat{V}_\theta L'}}. \quad (16)$$

2.5. Individual Growth Prediction

In some longitudinal studies, we may want to predict subject-specific growth trajectories, which is equivalent to predict the random effect b_i for each individual. Thus we need to compute the conditional expectation of the random effects $b_i = E(b_i|y_{i,A})$. It can be accomplished by maximizing the following objective function [6]:

$$\ell(b_i) = \frac{1}{\sigma^2}[y_{i,A} - f(\beta_A, b_i)]'[y_{i,A} - f(\beta_A, b_i)] + b_i' D b_i. \quad (17)$$

For a fixed $\boldsymbol{\beta_A}$, the above objective function is a constant plus the log of the posterior density of $\boldsymbol{b_i}$:

$$p(\boldsymbol{b_i}|\boldsymbol{y_{i,A}}, \boldsymbol{\beta_A}, \boldsymbol{D}, \sigma^2) = \frac{p(\boldsymbol{y_{i,A}}|\boldsymbol{\beta_A}, \boldsymbol{b_i}, \boldsymbol{D}, \sigma^2)p(\boldsymbol{b_i})}{p(\boldsymbol{y_{i,A}}|\boldsymbol{\beta_A}, \boldsymbol{D}, \sigma^2)}.$$

Therefore, the $\boldsymbol{b_i}$ maximizing Eqn. 17 is the posterior mode. Once we have the estimation $\hat{\boldsymbol{b_i}}$, we can reconstruct the continuous individual growth profiles by

$$\begin{aligned} f_{i,X} &= \frac{\hat{\beta}_{1,X} + \hat{b}_{1i,X}}{1 + \hat{\beta}_{2,X}\exp(-(\hat{\beta}_{3,X} + \hat{b}_{3i,X})t)}, \\ f_{i,Y} &= \frac{\hat{\beta}_{1,Y} + \hat{b}_{1i,Y}}{1 + \hat{\beta}_{2,Y}\exp(-(\hat{\beta}_{3,Y} + \hat{b}_{3i,Y})t)}, \\ f_{i,Z} &= \frac{\hat{\beta}_{1,Z} + \hat{b}_{1i,Z}}{1 + \hat{\beta}_{2,Z}\exp(-(\hat{\beta}_{3,Z} + \hat{b}_{3i,Z})t)}. \end{aligned} \tag{18}$$

3. Application: Early Brain Development

Motivation. The motivation of our application relies on the project of building parallel head coils for neonates and young children for the acquisition of pediatric brain MRI images. We would like to know how the average head size of the population changes longitudinally. We would also like to predict the average head size at time points where we currently do not have MR scans, so that we can decide whether we need to build new coils for these ages even though we do not have existing scans at these time points.

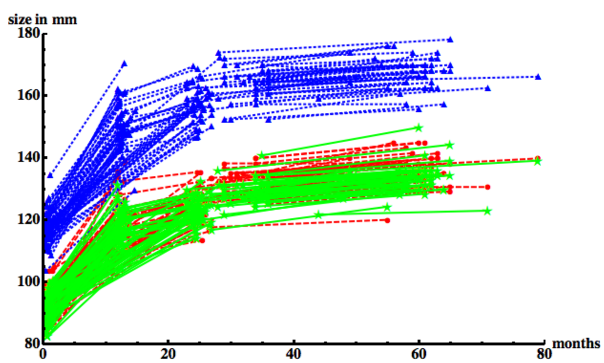

Figure 3. An illustration of individual growth trends. A spaghetti plot that connects repeated measurements of the same individual is shown. Multiple features that describe the three dimensional head size derived from each MRI brain image of neonates and young children is illustrated. X dimension: red dots connected by dashed lines; Y dimension: blue triangles connected by dotted lines; Z dimension: green stars connected by solid lines.

Data. For our study presented here, we apply the multivariate nonlinear mixed model analysis described in section 2 to a combination of two longitudinal neuroimaging studies, one of which contains repeated MR scans of neonates and young children aged between 0 to around 2 years old, while the other provides repeated MR scans of young children between 2 to around 6 years old. Out of the 236 subjects in our combined data set, 128 had 1 MR scan, 97 had 2 scans, and 11 had 3 scans. By combining the two data sets, we are able to study the longitudinal change of average head size during a wider time period.

As preliminary analysis, we consider the longitudinal growth of head size in X (left-right), Y (anterior-posterior), Z (superior-inferior) dimensions. We use the same probabilistic atlas of 1 year old as a reference to apply linear registration to all the image data and to obtain the X, Y, Z dimensions of the head size for each individual over time. Fig. 3 shows the spaghetti plot of the X, Y, Z head size for all individuals. We can observe from the plot that our data is unbalanced and incomplete and that the trend of growth displays an asymptotic property, which is shared by all the three dimensions of the head size and by all subjects.

Population growth curves. We use the logistic growth function as in Eqn. 8 to model the asymptotic growth pattern of early brain development. The joint modeling scheme of nonlinear mixed model described in section 2.3 is applied to study the growth of head size dimensiosn X, Y, Z and to obtain the population growth trajectories as a nonlinear function of time (Fig.4). The β_2 parameter that contributes to part of the intercept of the growth curve, as well as the β_3 parameter that determines how quickly the value grows from the initial value to the asymptote value, are tested to be not statistically different for all three dimension X, Y, Z. But the β_1 parameter that controls both the asymptote and the intercept of the growth curve are tested to be statistically different ($p<0.05$). This result means the X, Y, Z dimensions of the head share similar growth pattern, while each has its own initial and asymptote value.

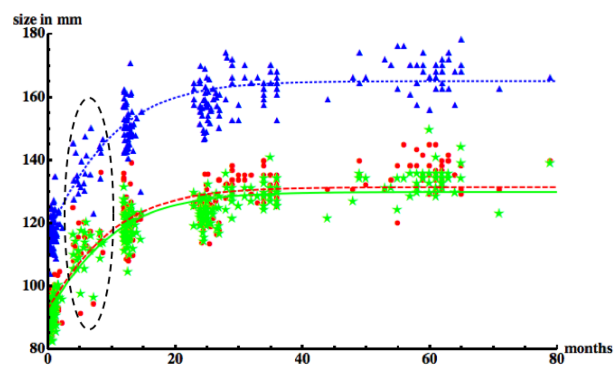

Figure 4. Population growth trajectories of head size dimension X, Y, Z plotted against the original data points ranged from age 0 to around 6 years old. A third population of 22 infants aged from 4 to 8 months old (in the black dashed ellipse) are also plotted to validate the soundness of the average growth estimation. Symbols are the same as those in Fig. 3.

We further validate our result by plotting a third population of 22 infants with age ranging from 4 to 8 months against the obtained population growth curves, as shown in Fig. 4. The head size data of the third population were not used to calculate the shown population growth curves, yet they fall nicely along the growth trend represented by the average growth trajectories.

Population growth rates. We can compute the growth rate of head size dimensions by taking the derivatives of the parametric growth curves:

$$\partial_t y = \frac{\beta_1 \beta_2 \beta_3 \exp(-\beta_3 t)}{(1 + \beta_2 \exp(-\beta_3 t))^2}, \tag{19}$$

as shown in Fig. 5. It indicates that the growth of the Y dimension is faster than those of X and Z dimensions of the head, i.e. almost 1mm faster per month. But at the same time, the growth rate of the Y dimension also slows down faster.

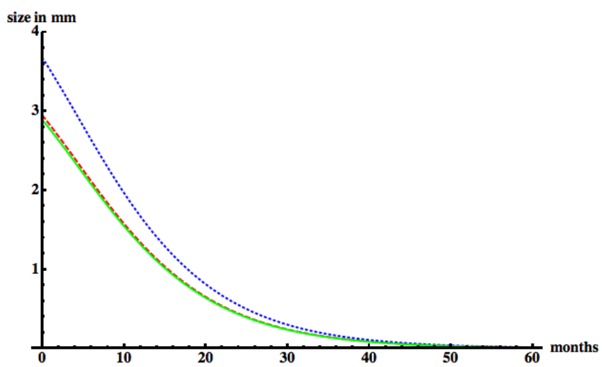

Figure 5. Growth rates of head size dimension X, Y, Z between birth to around 6 years old. X dimension: dashed red lines; Y dimension: dotted blue lines; Z dimension: solid green lines.

The faster growth rate of the Y dimension of the head can be expected if we examine the original repeated MR scans of any children. As shown in Fig. 6, the head for the neonate at 0.7 month old looks more round, while the head of the same child after two years is elongated in the anterior-posterior (Y) direction.

Individual growth trajectories. As described in section 2.5, we can predict the subject-specific growth trajectories by computing the random effects $\boldsymbol{b_i}$ for each individual. Fig. 7 shows the predicted individual growth trajectories for all 236 subjects. Remember that a lot of the individuals have very few scans, e.g. one or two scans, during the 6 year period, it is unlikely to predict their growth curves using traditional methods. On the other hand, the prediction of individual growth trajectories by longitudinal analysis can borrow "strength" across individual, while at the same time

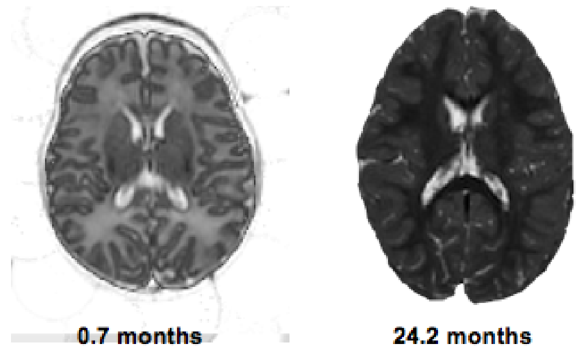

Figure 6. Repeated MR scans of the same child at age 0.7 month old and again at 24.2 months old. We can observe that the head for the neonate is more round, while the head of the same child after two years is elongated in the anterior-posterior (Y) direction.(Contrast changes between the neonate image and that of the two years old are due to the early brain myelination.)

maintain individual characteristics, e.g. individual profiles have different initial values at age 0 and different asymptote values as the children grow over time. This phenomenon justifies the approach of mixed models, which treats individual growth parameters, e.g. initial values, asymptotes, as random variables, and model them as samples from a population with variance characterizing among-individual variability.

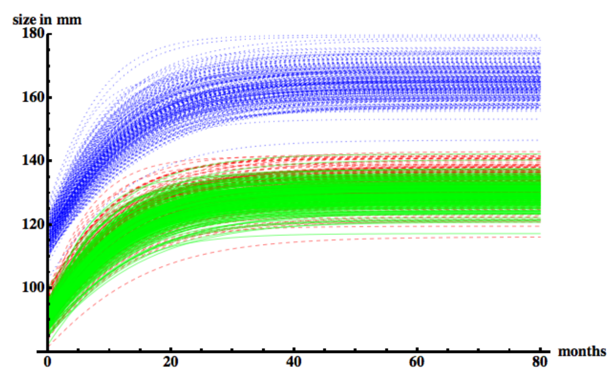

Figure 7. Individual growth trajectories for head size X, Y, Z dimensions. X dimension: dashed red lines; Y dimension: dotted blue lines; Z dimension: solid green lines.

Residuals. Given the fact that numerous individuals have very few time points and most of them just have one measurement, the prediction to individual growth curves could present a challenge, in a way that the prediction could be biased or erroneous. Thus, validation is necessary. In Fig. 7 we validate that all the predicted individual trajectories follow a reasonable growth pattern, i.e. no individual profile exhibit unrealistic behavior that violates the original growth trend. Next, we further validate our prediction results by looking at the residuals between the original measurements y_{ij} and the predicted value calculated by Eqn. 18:

$$r_{ij} = y_{ij} - f(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{b_i}}, t_{ij}),$$

where $f(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{b_i}}, t_{ij})$ is the individual growth trajectory computed by predicting the random effects $\hat{\boldsymbol{b_i}}$. Residual plots

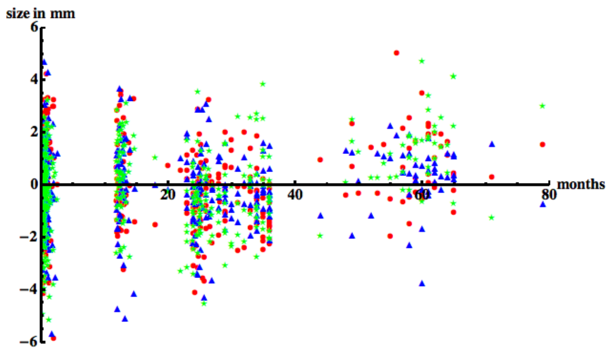

Figure 8. Residule plot between the original measurements y_{ij} and the predicted value calculated by $r_{ij} = y_{ij} - f(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{b_i}}, t_{ij})$.

for all three dimensions are shown in Fig. 8. We observe that each individual trajectory does not vary too far away from the original measurements, which is within the range of ±6mm during the whole 80 month period.

4. Discussion

We present a joint modeling scheme of nonlinear mixed models to statistically study multivariate longitudinal image data that exhibit asymptotic growth trends. We demonstrate our results using a combination of two neuroimaging longitudinal studies whose time periods barely overlap to study early brain development. We obtain population growth curves and growth rates for the head size dimensions as a nonlinear function of time and validate our result using a third population whose time points were not measured in the other two. Given the limited number of repeated measures per individual, we are able to predict the continuous individual growth trajectories for all 236 subjects. Residual plots are examined to make sure the predicted trajectories are a good fit to the original data. To our knowledge, this paper is the first multivariate nonlinear longitudinal analysis to study early brain development.

The nonlinear mixed model scheme we use is flexible enough to permit statistical hypothesis testing such as gender differences or differences between a patient and a control population. It also allows combining image-derived features with subject scores, e.g. motor, language, or cognitive scores. This enables us to link physical longitudinal brain changes with underlining brain performance or functions, which open a new research direction to our future work. Indeed, new analytic methods to study brain growth trajectories in healthy and patient populations as well as the longitudinal relationship between physical and functional features will have excellent potential to contribute to a better understanding of origin, timing and nature of morphologic differences in neurodevelopmental disorders.

References

[1] P. Aljabar, K. K. Bhatia, J. V. Hajnal, J. Boardman, L. Srinivasan, M. Rutherford, L. Dyet, D. Edwards, and D. Rueckert. Analysis of growth in the developing brain using non-rigid registration. IEEE ISBI, pages 201–204, 2006.

[2] C. G. Broyden. The convergence of a class of double-rank minimization algorithms. *Journal of the Institute of Mathematics and Its Applications*, 6:76–90, 1970.

[3] E. Courchesne, H. J. Chisum, J. Townsend, A. Cowles, J. Covington, B. Egaas, M. Harwood, S. Hinds, and G. A. Press. Normal brain development and aging: quantitative analysis at in vivo mr imaging in healthy volunteers. *Radiology*, 216:672 682, 2000.

[4] E. Courchesne, C. M. Karns, H. R. Davis, R. Ziccardi, R. A. Carper, Z. D. Tigue, H. J. Chisum, P. Moses, and K. Pierce. Unusual brain growth patterns in early life in patients with autistic disorder: An mri study. *Neurology*, 57(2):245–254, 2001.

[5] M. Davidian and Giltinan. *Nonlinear Models for Repeated Measurement Data*. Chapman and Hall, New York, 1995.

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

[7] B. Davis, P. T. Fletcher, E. Bullitt, and S. Joshi. Population shape regression from random design data. Proceeding of ICCV, 2007.

[8] P. J. Diggle, P. Heagerty, K.-Y. Liang, and S. L. Zeger. *Analysis of Longitudinal Data*. Oxford University Press, New York, 2002 (second edition).

[9] J. N. Giedd, J. Blumenthal, N. O. Jeffries, F. X. Castellanos, H. Liu, A. Zijdenbos, T. Paus, A. C. Evans, and J. L. Rapoport. Brain development during childhood and adolescence: a longitudinal mri study. *Nature Neuroscience*, 2:861–863, 1999.

[10] N. I. M. Gould and S. Leyffer. *An introduction to algorithms for nonlinear optimization*, pages 109–197. Frontiers in Numerical Analysis. Springer Verlag, Berlin, 2003.

[11] G. Marshall, R. D. Cruz-Mesía, A. E. Barón, J. H. Rutledge, and G. O. Zerbe. Non-linear random effects model for multivariate responses with missing data. *Statistics in Medicine*, 25:2817–2830, 2006.

[12] R. L. Sandland and C. A. McGilchrist. Stochastic growth curve analysis. *Biometrics*, 35:255–271, 1979.

[13] S. C. Xu, M. Styner, B. Davis, S. Joshi, and G. Gerig. Group mean differences of voxel and surface objects via nonlinear averaging. IEEE ISBI, pages 758–761, 2006.

[14] D. L. Zimmerman and V. Núñez-Antón. Parametric modelling of growth curve data: An overview. *TEST*, 10(1):1–73, 2001.