HHS Public Access

Author manuscript

Med Image Comput Comput Assist Interv. Author manuscript; available in PMC 2015 June 30.

Published in final edited form as:

Med Image Comput Comput Assist Interv. 2014 ; 17(0 3): 33–40.

Subject-specific prediction using nonlinear population modeling: Application to early brain maturation from DTI

Neda Sadeghi[1], **P. Thomas Fletcher**[1], **Marcel Prastawa**[2], **John H. Gilmore**[3], and **Guido Gerig**[1]

[1] Scientific Computing and Imaging Institute, University of Utah

[2] GE Global Research

[3] Department of Psychiatry, University of North Carolina

Abstract

The term prediction implies expected outcome in the future, often based on a model and statistical inference. Longitudinal imaging studies offer the possibility to model temporal change trajectories of anatomy across populations of subjects. In the spirit of subject-specific analysis, such normative models can then be used to compare data from new subjects to the norm and to study progression of disease or to predict outcome. This paper follows a statistical inference approach and presents a framework for prediction of future observations based on past measurements and population statistics. We describe prediction in the context of nonlinear mixed effects modeling (NLME) where the full reference population's statistics (estimated fixed effects, variance-covariance of random effects, variance of noise) is used along with the individual's available observations to predict its trajectory. The proposed methodology is generic in regard to application domains. Here, we demonstrate analysis of early infant brain maturation from longitudinal DTI with up to three time points. Growth as observed in DTI-derived scalar invariants is modeled with a parametric function, its parameters being input to NLME population modeling. Trajectories of new subject's data are estimated when using no observation, only the first or the first two time points. Leave-one-out experiments result in statistics on differences between actual and predicted observations. We also simulate a clinical scenario of prediction on multiple categories, where trajectories predicted from multiple models are classified based on maximum likelihood criteria.

1 Introduction

Longitudinal data analysis can provide further insight into growth, degeneration or disease progression by analyzing change trajectories rather than snapshots in time. In this setting, individual subjects' trajectories can be compared to the normative models computed via population modeling. One can then identify the timing of deviation from typical trajectories, interventions can be targeted toward a specific developmental period, or predicted trajectories can be used for assessment of disease risk during prodromal stage or for measuring efficacy of disease-modifying therapies, for example.

The term *prediction* can be used in very different contexts and with different goals. E.g., genetics may predict risk for disease, a patient score or disease status may be predicted from imaging biomarkers, or physiological age is predicted from sets of measurements. Here, we

focus on the notion of statistical inference to predict a future observation of a new subject given a model for the temporal trajectory, comprehensive statistics from training on population data, and a set of past observations from this subject (Fig. 1). The predicted observation with confidence bounds, or more general the prediction of the whole trajectory with variability, can then be used to estimate deviation from the norm. In addition, given normative models for multiple groups, for example for different patient categories and/or controls, one can derive prediction trajectories for each group and classify based on the most likely category.

Prediction of an individual trajectory is possible even if not all the observations for all time points would be available for that subject by pooling the data from other subjects in the study along with the available observations for the individual. Analysis of longitudinal data needs to take into account the correlation due to repeated measures, variability between subjects, often unbalanced spacing due to acquisitions at different time points and missing data. All these favor the use of mixed effects models, which represent a class of statistical methods that model the correlation of measurements of an individual along with modeling the mean response of a population over time.

The proposed methodology is generic with respect to any type of data. Here, we demonstrate proof of concept with a clinical infant neuroimaging study. Longitudinal brain imaging is increasingly used in clinical studies as it provides a superior characterization of developmental trajectories compared to cross-sectional studies [2, 3]. Such studies have mostly focused on population analysis. However, individuals would likely benefit from subject-specific assessments, comparing an individual's image-derived data at each given age to the norm, and predictions of subject-specific growth trajectories and intervals based on measurements of only one or two time points, predictions which may improve early detection and therapeutic intervention.

Key aspects discussed in this paper are the selection of optimal nonlinear models to characterize temporal trajectories, building of normative models for populations, and the development of a statistical inference framework to predict future observations based on new scans' data and statistics from a reference population (Sec. 2). Predicted trajectories from multiple groups can then be used for classification and the results are presented in (Sec. 3).

2 Method

Reference population

To prepare discussion of the prediction scheme, we briefly describe the nonlinear mixed effects model (NLME) [4]. In the mixed effects model, the observed data are assumed to be a combination of both fixed effects, β, parameters associated with the entire population (or at least within a subpopulation), and random effects, b, that are specific to an individual drawn at random from the population. Random effects account for the heterogeneity that is present in the population as these effects vary among subjects. NLME is a generalization of linear mixed effects and nonlinear regression, some or all of the fixed or random effects enter the model nonlinearly. Each individual's response is modeled as:

$$y_i = f(\phi_i, t_i) + e_i, \quad (1)$$

where $\phi = A_i\beta + B_i b_i$, and b_i are random effects with distribution $N \sim (0, \Psi)$. A_i and B_i are design matrices that indicate whether a specific fixed or random effect should be included in the model. The function f can be any nonlinear function, and e_i is the measurement error and is assumed normally distributed $N \sim (0, \sigma^2)$. Random effects and measurement errors are assumed to be independent.

The likelihood function for the mixed effects model is written as:

$$L\left(\beta, \Psi, \sigma^2 | y\right) = \prod_{i=1}^{M} p\left(y_i | \beta, \Psi, \sigma^2\right) \quad (2)$$

Since non-observable random effects are part of the model, we must integrate out random effects; thus, the marginal density of y_i becomes:

$$p\left(y_i | \beta, \Psi, \sigma^2\right) = \int p\left(y_i | \beta, b_i, \sigma^2\right) p\left(b_i | \Psi, \sigma^2\right) db_i. \quad (3)$$

The population growth parameters β and variance components Ψ and σ^2 are estimated by maximizing the likelihood equation of (3). In general there is no closed form solution to equation 3. We approximate the integral in (3) using Taylor expansion of the model function f around conditional modes of random effects b and the current estimate of β [4].

The distribution of the maximum likelihood estimator β of the fixed effects based on the linear mixed effects approximation [4] is written as:

$$\hat{\beta} \sim N\left(\beta, \left[\sum_{i=1}^{M} \hat{X}_i^T \hat{V}_i^{-1} \hat{X}_i\right]^{-1}\right), \quad (4)$$

where $\hat{V} = \hat{Z}_i \Psi \hat{Z}_i^T + \sigma^2 I_{n_i}, \hat{X}_i = \frac{\partial f_i}{\partial \beta^T}|_{\hat{\beta}, \hat{b}_i}, \hat{Z}_i = \frac{\partial f_i}{\partial b_i^T}|_{\hat{\beta}, \hat{b}_i}$.

Choice of nonlinear function

The choice of the nonlinear function, f, is to be seen as study and data specific. Nonlinear models provide a more parsimonious model compared to their linear counterparts (i.e. polynomials). More important, as extrapolation of data beyond the observed range plays an important role in the prediction of future values, model parameters of nonlinear schemes tend to have natural interpretations [5]. Data analyzed here are radial diffusivity (RD) measurements of diffusion tensor imaging (DTI) of early brain development and age is the covariate. White matter is known to mature more rapidly in the first year of life than second, with continued maturation but at a much slower rate into adulthood [1, 6]. This favors functions that have asymptotic behavior such as exponential, logistic, or Gompertz. We used Akaike Information Criterion (AIC) [7] for model selection, where $AIC = -2log(L_i) + 2n_{par}$,

L_i is the likelihood of model i and n_{par} is the number of model parameters. Among the tested models, the Gompertz function provided the lowest AIC measure, so that this function was chosen for infant DTI modeling. Using the Gompertz function, response y is modeled as $y = ax$, where a is the final asymptotic value, parameter r specifies the decay in the growth rate, and parameter d controls the difference between the final and initial values of y. All the parameters of the Gompertz function are used as fixed effects. Parameters a and d were chosen as random effects as they provided the best model fit as measured by lower AIC.

Prediction of new individual trajectory

Upon availability of new data for an individual, we can use the reference population parameters along with the individual's available data to predict a personalized growth trajectory. We substitute $\hat{\beta}, \hat{\psi}$ and $\hat{\sigma}$ of the reference population for the unknown parameters to predict an approximate empirical Bayes' estimate of b_i. Once the subject's random effects are estimated, the individual's growth trajectory and future values can be predicted.

The prediction of b_i can be calculated from the posterior distribution of $p(b_i|y_i)$ using Bayes' rule:

$$p\left(b_i|y_i,\beta,\Psi,\sigma^2\right)=\frac{p\left(y_i|\beta,b_i,\Psi,\sigma^2\right)p\left(b_i|\Psi\right)}{p\left(y_i|\beta,\Psi,\sigma^2\right)} \quad (5)$$

By maximizing the log of the posterior density of b_i, we obtain the following objective function:

$$l\left(b_i\right)=-\frac{1}{\sigma^2}\left(y_i-f\left(\beta,b_i\right)\right)^T\left(y_i-f\left(\beta,b_i\right)\right)-b_i^T\Psi^{-1}b_i. \quad (6)$$

Once $\hat{b}_i$ is estimated, $E\left[\hat{b}_i\right]\simeq\hat{\Psi}\hat{Z}_i^T\hat{V}_i^{-1}\left(y_i-f\left(A_i\hat{\beta}+B_i\hat{b}_i,t_i\right)+\hat{Z}_i\hat{b}_i\right)$, we can construct continuous growth trajectories of the ith subject. The ith subject prediction for the corresponding responses y_i is: $E\left[\hat{y}_i|b_i\right]=f\left(x_i^T\hat{\beta}+z_i^T\hat{b}_i,t\right)$, where x_i represents a vector of fixed effects covariates and z_i represents a vector of covariates corresponding to random effects.

Individual's prediction interval

By knowing the sampling distribution $b_i\sim N\left(\hat{b}_i,\hat{W}\right)\quad and\quad \hat{W}=\hat{\Psi}-\hat{\Psi}\hat{Z}_i^TV_i^{-1}\hat{Z}_i\hat{\Psi}$, we can employ a Monte Carlo simulation to approximate the subject-specific prediction interval. One thousand samples of $\beta\sim N\left(\hat{\beta},\left[\Sigma_{i=1}^M\hat{X}_i^T\hat{V}_i^{-1}\hat{X}_i\right]^{-1}\right), b\sim N\left(\hat{b}_i,\hat{W}\right),\quad and\quad e\sim N\left(0,\hat{\sigma}\right)$ were generated from their respective distributions. Subsequently, 1000 trajectories were constructed by the NLME model. The prediction interval for the "new" subject can be calculated by constructing the $1-a$ range of values for a given time point, t_{ij}. At each t_{ij}, $a/2$ and $1-a/2$ percentiles were calculated as the lower and upper limits of the subject-specific interval.

Classification of a new individual

An individual's predicted trajectory is a combination of the estimated population parameters (fixed effects), and subject specific random effects. If multiple reference subpopulations are available, denoted as $c \in C$, mixed effects modeling as described earlier can be used to estimate fixed effects $\hat{\beta}_c$, variance-covariance of random effects $\hat{\Psi}_c$, and variance of noise $\hat{\sigma}_c$ for each subpopulation c. Upon availability of new scans for an individual, random effects and individual's subject's trajectory can be predicted as shown. Once random effects (trajectories) are predicted for the new individual, we classify the subject to belong to the subpopulation that has the highest likelihood given the individual's predicted random effects $\hat{b}_{ic}$ and reference subpopulation parameters $\hat{\Psi}_c$. We assign a subject to a subpopulation c where $p\left(\hat{b}_{ic}|\hat{\Psi}_c\right)$ has the highest value among C. This method takes into account not only the subpopulation trajectory $\hat{\beta}_c$ to predict $\hat{b}_{ic}$, but also the heterogeneity of subjects $\hat{\Psi}_c$ presented in the subpopulation for classification.

3 Validation and Results

The following discussion is based on clinical data from an ongoing infant DTI study but focuses on validation by mimicking its potential clinical use via leave-one-out experiments. We verify two aspects of the proposed methodology based on two scenarios: 1) predicting a future observation for a new individual, does it fall into the range of the norm and what is the difference between prediction and observed value, and 2) having models for two reference populations and thus two prediction trajectories for new individuals, what is the classification based on the more likely population (Fig. 4).

We have access to DTI data of 26 subjects with a total of 59 DTI scans (neonate: 23, year 1: 22, year 2: 14) from a normative study, with preprocessing and unbiased atlas mapping following [6]. The reference population trajectory is estimated using all subjects excluding the one used for testing. Fig. 2 shows the estimated subject growth trajectory along with the subject-specific prediction interval for RD values of a test subject, overlaid on the population model (gray). The trajectory of the individual is predicted as discussed in section 2.

Figure 2 left shows the predicted trajectory based on only the first time point (solid blue curve and light blue region) with the two left out measurements (red dots). Upon availability of more time points, future observations are predicted with increased precision (Fig. 2 right).

Figure 3 shows the difference between observed and predicted values for RD of posterior thalamic radiation (PTR). The population trajectory was constructed using scans of 25 subjects. We then test our methodology for predicting RD at years 1 and 2, i.e. closeness of the predicted trajectory (solid blue line) to the observed measures (red dots) shown in Fig. 2. Predicting 1 year values was based on 19 subjects with at least neonate and 1 year time points, whereas prediction at 2 years was obtained from 9 subjects that have three time points available. Without any observation available for a new individual, the predicted RD value at years 1 and 2 are the population averages. However, as one observation becomes available the RD values at years 1 and 2 can be predicted with more accuracy (Fig. 3

middle). With two observations, variability of differences between predicted and observed RD values is further reduced (Fig. 3 right).

The predicted subject trajectories can also be used for classification. To illustrate a clinical scenario with two populations (e.g. controls vs. disease), we construct an example using two different regions representing two categories. Figure 4 illustrates the concept of classification into the categories PTR or splenium. We construct the population trajectories for RD of both PTR and splenium of 25 subjects, and then predict left out subject's RD trajectory for a test region as if the region could be either splenium or PTR. Looking at the test region's first time point (Fig. 4), it seems that it would be splenium, but the second time point is more similar to PTR; illustrating that analysis at single time points can easily lead to contradictory results. We predict the RD trajectories of the test region based on the PTR population (purple dashed line in Fig. 4) and splenium (yellow dashed line). With the two predictions for the test region, we then use the classification method of section 2 to assign the most likely category. The experiment is repeated for all the subjects with available scans at neonate and year 1. Overall, only one subject's splenium was misclassified as PTR and vice-versa, compared to the overlapping distributions and also conflicting classifications at single time points.

4 Discussion and Conclusion

This paper describes a framework for prediction based on a statistical inference approach in the mixed-effects-modeling setting, with emphasis on its application to *discrete-time data*, the use of *nonlinear parametric* temporal functions, and employing nonlinear mixed-effects-modeling (NLME). Given a new, unseen individual, estimated subject-specific trajectories and prediction intervals not only take into account parameter estimates of the normative population but also consider existing observations from the new individual. Experimental tests with a leave-one-out scheme clearly demonstrate that the resulting subject-specific prediction interval, representing uncertainty, is steadily narrowing from using no additional observation over the use of one and then two individual measurements. Here, we are making the assumption that an individual will have a temporal trajectory similar to that of the reference population. This allows transfer of information from the estimated population model to the prediction of a new individual trajectory, thus specifically tailored to the new subject.

Although current work focuses on application to clinical longitudinal neuroimaging studies, we here decided to demonstrate the potential use of prediction-based classification in a simulation procedure. Whereas comparing trajectories of two different anatomical regions may be seen somewhat artificial in view of standard clinical studies with control and patient populations, it nevertheless highlights the interesting properties and new potential of the proposed scheme. Results clearly demonstrate that classification based on estimated trajectories with use of reference population statistics and few observations is more powerful than cross-sectional classification based on data at single time-points.

Acknowledgments

Supported by NIH grants: Conte Center MH064065, NA-MIC EB005149, ACE HD055741, and CAMID NIDA DA022446-01.

References

1. Dubois J, Dehaene-Lambertz G, Perrin M, Mangin JF, Cointepas Y, Duchesnay E, Le Bihan D, Hertz-Pannier L. Asynchrony of the early maturation of white matter bundles in healthy infants: quantitative landmarks revealed noninvasively by diffusion tensor imaging. Hum Brain Mapp. Jan. 2008 29:14–27. [PubMed: 17318834]
2. Giedd JN, Snell JW, Lange N, Rajapakse JC, Casey BJ, Kozuch PL, Vaituzis AC, Vauss YC, Hamburger SD, Kaysen D, Rapoport JL. Quantitative magnetic resonance imaging of human brain development: ages 4-18. Cereb. Cortex. 560; 6:4–551.
3. Kraemer HC, Yesavage JA, Taylor JL, Kupfer D. How can we learn about developmental processes from cross-sectional studies, or can we? Am J Psychiatry. Feb; 2000 157(2):163–171. [PubMed: 10671382]
4. Lindstrom ML, Bates DM. Nonlinear mixed effects models for repeated measures data. Biometrics. Sep.1990 46:673–687. [PubMed: 2242409]
5. Pinheiro, JC.; Bates, DM. Mixed-Effects Models in S and S-Plus. Springer; 2000.
6. Sadeghi N, Prastawa M, Fletcher PT, Wolff J, Gilmore JH, Gerig G. Regional characterization of longitudinal DT-MRI to study white matter maturation of the early developing brain. Neuroimage. Mar.2013 68:236–247. [PubMed: 23235270]
7. Sakamoto. A new look at the statistical model identification. IEEE Transactions on Automatic Control. 723; 19:6–716.

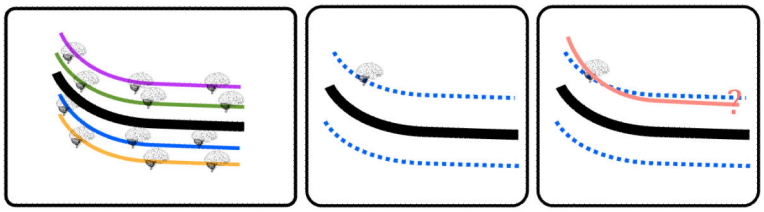

Fig. 1.

Prediction based on reference population and new individual's scan(s). Left: Population trajectory constructed based on the reference population. Middle: Population trajectory along with predicted interval. Right: Can we predict the new individual's trajectory based on new scan(s) and the reference population?

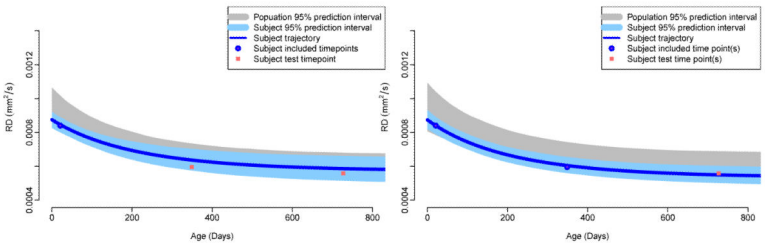

Fig. 2.

Subject prediction interval compared to the overall prediction for RD of PTR. Left: subject-specific interval calculated based on only one time point (neonate). Right: subject-specific interval calculated based on scans at neonate and 1 year. Blue dots represent observations used for prediction, and red dots actual observed measurements.

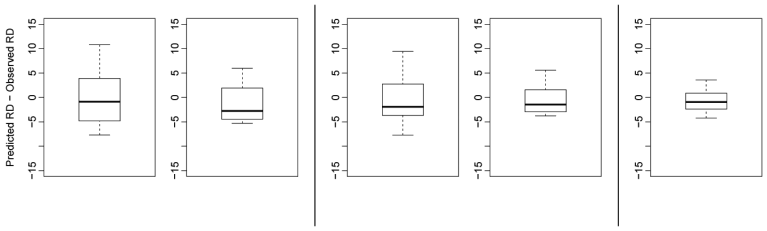

Fig. 3.

Distribution of differences between observed and predicted values for radial diffusivity (RD) of PTR. Left two figures: using population mean as predicted value for prediction at year 1 (left) and year 2 (right). Middle two figures: using RD at neonate and reference population to predict RD at year 1 (left) and year 2 (right). Right figure: using population reference and both neonate and year 1 RD to predict year 2 RD. Note: the reference population was estimated based on 26 subjects with 59 scans (minus the test subject), but only 9 subjects had data for all three time points. Plots illustrate increased prediction performance from using no observation over one and then two observations (left, middle, right). RD differences displayed here have been scaled up by 10^5.

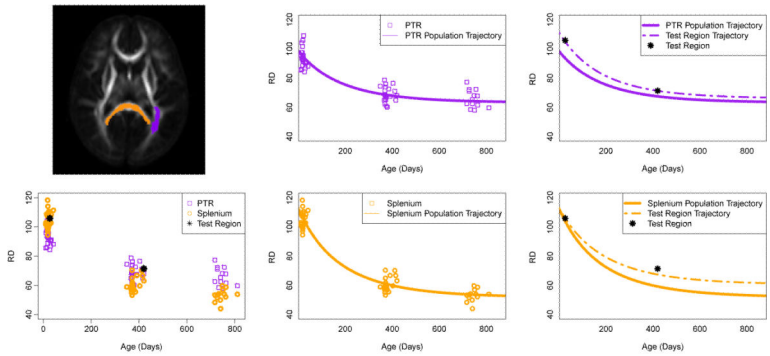

Fig. 4.

Test if a new region's RD measures are representing PTR or splenium. Left: scatterplot of longitudinal measurements of radial diffusivity (RD) for the two categories. Middle: population trajectories for PTR and splenium. Right: trajectories of the test region predicted for each category (dashed lines). The test regions' trajectory with highest likelihood determines the classification, here classified into the PTR category. RD values displayed here have been scaled up by 10^5.