Additional Proofs for Modeling Parkinson's Disease Progression from Longitudinal Voice Biomarkers: A Comparative Study of Statistical and Neural Mixed-Effects Models

Ran Tong[1], Lanruo Wang[2], Tong Wang[3], Wei Yan[4]

[1]Department of Mathematical Sciences, University of Texas at Dallas, Richardson, TX, USA

[2]Naveen Jindal School of Management, University of Texas at Dallas, Richardson, TX, USA

[3]Department of Biology, Duke University, Durham, NC, USA

[4]Department of Neurosurgery, The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China

Author-posted companion note. This document contains additional mathematical arguments related to the published article: *Modeling Parkinson's Disease Progression from Longitudinal Voice Biomarkers: A Comparative Study of Statistical and Neural Mixed-Effects Models*. Published version: *Computer Methods and Programs in Biomedicine Update*, Volume 9, 2026, Article 100242.

Published version DOI: 10.1016/j.cmpbup.2026.100242.

Important note. This document is an author-posted companion note. It is not the journal's official supplementary material and should not be treated as a substitute for the published article. The purpose of this note is to provide additional mathematical context for the empirical finding that semi-parametric mixed-effects models can be more sample-efficient than neural mixed-effects models in small-cohort longitudinal telemonitoring studies.

Abstract

This companion note gives a stylized theoretical explanation for a pattern observed in the published study, *Modeling Parkinson's Disease Progression from Longitudinal Voice Biomarkers: A Comparative Study of Statistical and Neural Mixed-Effects Models*. Semi-parametric mixed-effects models can outperform neural mixed-effects models when the number of subjects is small, even when each subject has many repeated measurements. We compare a low-dimensional statistical mixed model with a higher-dimensional neural mixed-effects model using a bias–variance argument. The result shows that when the longitudinal signal is well approximated by a smooth low-dimensional structure, and when subject-specific random effects can be estimated from repeated observations, the statistical mixed model can have a smaller prediction-risk rate. The conclusion is not that statistical models always dominate neural models. Rather, the conclusion is that there is a small-cohort regime in which the complexity cost of neural mixed-effects models can outweigh their approximation advantage.

Keywords: Parkinson's disease; voice biomarkers; longitudinal data; statistical and neural mixed-effects models; mixed-effects models; generalized additive mixed models; neural mixed-effects models; sample size; prediction risk.

1 Purpose

The published article compared statistical mixed-effects models and neural mixed-effects models for modeling Parkinson's disease progression from longitudinal voice biomarkers. The study used $m = 42$ subjects and $N = 5875$ repeated observations. The empirical result was that generalized additive mixed models (GAMMs) and linear mixed-effects models (LMMs) achieved substantially smaller test errors than generalized neural network mixed models (GNMMs), neural mixed-effects models (NMEs), and a feed-forward neural network baseline.

This note provides a mathematical explanation for that empirical pattern. The key point is that longitudinal telemonitoring data have two different sample-size dimensions. Let m denote the number of subjects, and let n_i denote the number of repeated observations for subject i. A large number of repeated observations helps estimate within-subject trajectories and low-dimensional subject-specific random effects. However, learning a flexible neural population-level function and nonlinear subject-specific neural deviations still depends strongly on the number of independent subjects. Therefore, a dataset can have a moderately large total number of observations $N = \sum_i n_i$, while still being small for training a high-capacity neural mixed-effects model.

2 Setup

Let $i = 1, \ldots, m$ index subjects and $j = 1, \ldots, n_i$ index repeated observations for subject i. Let

$$N = \sum_{i=1}^{m} n_i, \qquad n_{\min} = \min_{1 \le i \le m} n_i.$$

For each observation, let $Y_{ij} \in \mathbb{R}$ denote the response, such as the total UPDRS score, and let $X_{ij} \in \mathbb{R}^p$ denote the vector of voice biomarkers and time-related covariates.

We consider the longitudinal data-generating model

$$Y_{ij} = f_0(X_{ij}, t_{ij}) + z_{ij}^{\top} b_i + \varepsilon_{ij}, \tag{1}$$

where f_0 is the population-level mean function, $z_{ij} \in \mathbb{R}^q$ is the random-effect design vector, $b_i \in \mathbb{R}^q$ is a subject-specific random effect, and ε_{ij} is the residual error. We assume

$$b_i \sim N(0, D), \qquad \varepsilon_{ij} \sim N(0, \sigma^2),$$

with b_i independent of ε_{ij}.

For example, a random-intercept model has $q = 1$, and a random-intercept and random-slope model has $q = 2$. These are low-dimensional subject-specific structures. In contrast, a neural mixed-effects model may introduce many subject-specific deviations in hidden layers or output layers.

Let $\mathcal{F}_S$ denote a semi-parametric statistical mixed model class, such as a GAMM. Suppose it uses d linear covariates and K spline basis functions for time. Its effective population-level dimension is

$$D_S = d + K.$$

Let $\mathcal{F}_N$ denote a neural mixed-effects model class. Let W denote the effective number of population-level neural parameters after regularization, and let r denote the effective number of subject-specific neural parameters per subject. In many practical neural mixed-effects models,

$$W \gg D_S, \qquad r \geq q.$$

This reflects the larger flexibility of the neural model.

We focus on prediction for a subject already observed during training. This matches a common telemonitoring use case: a patient has a longitudinal history, and the goal is to forecast a future measurement for the same patient. Define the prediction risk of an estimator $\widehat{f}$ as

$$R(\widehat{f}) = \mathbb{E}\left[\left\{Y_{ij}^{\text{new}} - \widehat{f}_i(X_{ij}^{\text{new}}, t_{ij}^{\text{new}})\right\}^2\right],$$

where the expectation is over a new repeated measurement from an already monitored subject. Let $R(f_0)$ denote the irreducible risk under the true conditional mean structure.

3 Risk comparison

Assumption 1 (Smooth low-dimensional longitudinal signal)**.** *The population-level mean function f_0 is smooth in time and can be approximated by a spline mixed model with error*

$$\inf_{f \in \mathcal{F}_S} \|f - f_0\|_2^2 = O(K^{-2\nu})$$

for some smoothness index $\nu > 0$.

Assumption 2 (Regular statistical mixed-model estimation)**.** *The statistical mixed model is identifiable, the covariance matrix of the random effects is positive definite, the design matrices have bounded eigenvalues, and the errors are sub-Gaussian. The effective population-level dimension of the statistical model is $D_S = d + K$.*

Assumption 3 (Regularized neural mixed-effects estimation)**.** *The neural mixed-effects estimator is trained in a regularized function class with effective population-level dimension W and effective subject-specific dimension r. Its estimation cost is of order $W/N + r/n_{\min}$. Its approximation error* *is*

$$A_N = \inf_{f \in \mathcal{F}_N} \|f - f_0\|_2^2.$$

Remark 1. *Assumption 3 is an effective-dimension statement. It does not require the raw number of neural weights to be exactly W. Rather, W represents the complexity left after regularization, early stopping, weight decay, architecture constraints, and optimization choices.*

Theorem 1 (A sufficient condition for statistical mixed models to dominate)**.** *Under Assumptions 1–3, the semi-parametric statistical mixed model satisfies*

$$R(\widehat{f}_S) - R(f_0) = O_p\left(K^{-2\nu} + \frac{D_S}{N} + \frac{q}{n_{\min}}\right). \tag{2}$$

The neural mixed-effects model has risk of order

$$R(\widehat{f}_N) - R(f_0) = O_p\left(A_N + \frac{W}{N} + \frac{r}{n_{\min}}\right). \tag{3}$$

Therefore, the statistical mixed model has a smaller prediction-risk rate whenever

$$K^{-2\nu} + \frac{D_S}{N} + \frac{q}{n_{\min}} = o\left(A_N + \frac{W}{N} + \frac{r}{n_{\min}}\right). \tag{4}$$

In particular, if the neural model has no meaningful approximation advantage over the spline mixed model, so that

$$A_N = O(K^{-2\nu}),$$

and if

$$D_S = o(W), \qquad q \le r,$$

then the statistical mixed model has a smaller variance-dominated risk rate when the subject-specific random effects are estimable, namely when

$$q \log m = o(n_{\min}). \tag{5}$$

Proof. The proof uses a bias–variance decomposition.

For the statistical mixed model, write

$$R(\widehat{f}_S) - R(f_0) \le C_1 \inf_{f \in \mathcal{F}_S} \|f - f_0\|_2^2 + C_2 \frac{D_S}{N} + C_3 \frac{q}{n_{\min}}.$$

The first term is the approximation error. By Assumption 1,

$$\inf_{f \in \mathcal{F}_S} \|f - f_0\|_2^2 = O(K^{-2\nu}).$$

The second term is the population-level estimation cost for a D_S-dimensional semi-parametric model. The third term is the subject-specific estimation cost for a q-dimensional random-effect vector estimated from at least $n_{\min}$ repeated observations. This gives

$$R(\widehat{f}_S) - R(f_0) = O_p\left(K^{-2\nu} + \frac{D_S}{N} + \frac{q}{n_{\min}}\right).$$

For the neural mixed-effects model, the same decomposition gives

$$R(\widehat{f}_N) - R(f_0) = O_p\left(A_N + \frac{W}{N} + \frac{r}{n_{\min}}\right).$$

Here A_N is the approximation error of the neural class, W/N is the population-level estimation cost of the effective neural model, and $r/n_{\min}$ is the cost of estimating subject-specific neural deviations.

Comparing the two rates gives condition (4). If $A_N = O(K^{-2\nu})$, then the neural model does not have a large approximation advantage over the spline mixed model. If also $D_S = o(W)$ and $q \leq r$, then the statistical mixed model has a smaller complexity term. The condition $q \log m = o(n_{\min})$ ensures that the subject-specific random effects can be estimated with enough stability across m subjects. Under these conditions, the statistical mixed model has a smaller variance-dominated risk rate. □

4 Interpretation for the Parkinson's telemonitoring study

Theorem 1 separates two regimes. If

$$n_{\min} = O(q \log m),$$

then each subject has too few repeated observations to estimate subject-specific effects reliably. In this case, the term

$$\frac{q}{n_{\min}}$$

does not vanish fast enough. The statistical mixed model should not be expected to show a stable advantage based on random-effect estimation. Neural mixed-effects models face the same issue, often more strongly if $r > q$.

If instead

$$q \log m = o(n_{\min}),$$

then subject-specific effects can be estimated more stably. In this regime, the main comparison is driven by the population-level complexity terms:

$$\frac{D_S}{N} \qquad \text{versus} \qquad \frac{W}{N}.$$

If the true signal is well approximated by smooth low-dimensional structure and $D_S \ll W$, then the statistical model has a smaller estimation cost.

For the Parkinson's telemonitoring application, the published study used

$$m = 42, \qquad N = 5875.$$

For a random-intercept and random-slope model, $q = 2$. Then

$$q \log m = 2 \log(42) \approx 7.48.$$

The average number of repeated observations per subject is

$$\bar{n} = \frac{5875}{42} \approx 140.$$

Thus, the repeated-measurement dimension is much larger than $q \log m$. This supports the use of a low-dimensional mixed-effects structure for within-subject forecasting.

However, the number of independent subjects remains only $m = 42$. This is small for training a flexible neural mixed-effects model with a large number of population-level and subject-specific neural parameters. Therefore, the empirical finding that GAMM and LMM outperform GNMM, NME, and ANN is consistent with the risk comparison above.

The theorem does not say that statistical mixed models always outperform neural mixed-effects models. Instead, it gives a sufficient condition for when this can happen. Statistical mixed models are favored when

$$f_0 \text{ is smooth and low-dimensional,}$$

$$n_{\min} \gg q \log m,$$

and

$$D_S \ll W.$$

Neural mixed-effects models are more likely to become competitive when they have a clear approximation advantage and when the number of independent subjects is large enough to estimate flexible population-level and subject-specific nonlinear structure. This requires not only many repeated observations, but also many subjects.

In practical terms, a longitudinal telemonitoring dataset can be large in the total number of rows but still small in the number of independent subjects. In such a setting, a low-dimensional GAMM or LMM can be more data-efficient than a neural mixed-effects model.

5 Conclusion

This companion note provides a mathematical explanation for why semi-parametric mixed-effects models can outperform neural mixed-effects models in small-cohort longitudinal telemonitoring studies. The main point is a sample-size tradeoff. Repeated observations per subject help estimate low-dimensional random effects, but they do not fully replace the need for many independent subjects when fitting flexible neural population and subject-specific components.

A useful threshold is

$$n_{\min} \gg q \log m.$$

Above this threshold, low-dimensional random effects can be estimated stably. If the signal is also smooth and low-dimensional, then the statistical mixed model can have lower prediction risk than a neural mixed-effects model with much larger effective dimension.

References

[Bartlett et al.(2017)] P. L. Bartlett, D. J. Foster, and M. J. Telgarsky. Spectrally-normalized margin bounds for neural networks. *Advances in Neural Information Processing Systems*, 30, 2017.

[Breslow and Clayton(1993)] N. E. Breslow and D. G. Clayton. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88(421):9–25, 1993.

[Györfi et al.(2002)] L. Györfi, M. Kohler, A. Krzyzak, and H. Walk. *A Distribution-Free Theory of Nonparametric Regression*. Springer, 2002.

[Laird and Ware(1982)] N. M. Laird and J. H. Ware. Random-effects models for longitudinal data. *Biometrics*, 38(4):963–974, 1982.

[Lin and Zhang(1999)] X. Lin and D. Zhang. Inference in generalized additive mixed models by using smoothing splines. *Journal of the Royal Statistical Society: Series B*, 61(2):381–400, 1999.

[Mandel et al.(2023)] F. Mandel, R. P. Ghosh, and I. Barnett. Neural networks for clustered and longitudinal data using mixed effects models. *Biometrics*, 79(2):711–721, 2023.

[Ruppert et al.(2003)] D. Ruppert, M. P. Wand, and R. J. Carroll. *Semiparametric Regression*. Cambridge University Press, 2003.

[Schmidt-Hieber(2020)] J. Schmidt-Hieber. Nonparametric regression using deep neural networks with ReLU activation function. *The Annals of Statistics*, 48(4):1875–1897, 2020.

[Wainwright(2019)] M. J. Wainwright. *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*. Cambridge University Press, 2019.

[Wood(2011)] S. N. Wood. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B*, 73(1):3–36, 2011.

[Wood(2017)] S. N. Wood. *Generalized Additive Models: An Introduction with R.* Chapman and Hall/CRC, 2nd edition, 2017.

[Wörtwein et al.(2023)] T. Wörtwein, N. B. Allen, L. B. Sheeber, R. P. Auerbach, J. F. Cohn, and L.-P. Morency. Neural mixed effects for nonlinear personalized predictions. In *Proceedings of the 2023 International Conference on Multimodal Interaction*, pages 445–454. ACM, 2023.