Modeling Group Differences in Event Related Potentials for Adolescents with and without ADHD

A Functional Mixed Effect Model

Emilie Campos

Introduction: Motivation and Goals

Electroencephalogram (EEG) is a noninvasive technology used commonly in studying psychiatric disorders. To study a person's response to a stimulus, we measure the electrophysiological response as the event related potential (ERP). It is common to study how different diagnosis groups respond to stimuli by measuring the peak amplitudes of the ERP. We will model these ERP curves using cubic B-splines and investigate how they differ in typically developing children and children with ADHD. This analysis followed a similar analysis performed by Fischer et. al published in The Annals of Applied Statistics [1].

In a study performed by the UCLA Semel Institute for Neuroscience and Human Behavior, 363 participants aged 7-17 years old, 266 with ADHD and 97 typically developing (TD), were asked to perform several tasks in order to study their event-related potential (ERP) responses through an Electroencephalogram (EEG). A continuous performance task was used to diagnose ADHD since it requires continuous focus and tracking. Subjects are presented single letters one at a time in the center of the computer screen. These letters are size 80 with Arial bold font. The subject is instructed to press and release the spacebar as quickly as they can after viewing each letter. However, when presented the letter "X" they are to not press the spacebar button and not make any type of response. The other possible letters they may receive are A, B, C, D, F, I, L, O, and T. There are 360 continuous trials randomly presented. There are 3 different interstimulus interval (ISI) times used for each trial: 1000, 2000, or 4000 ms. Each letter is always presented for 250 ms, leaving either 750, 1750, or 3750 ms of remainder response time per given fixed ISI trial. The total task takes 14.5 minutes. For the continuous performance task, the epochs, or timecourses, were locked to the stimulus, the letter being presented. The trial conditions were labeled 'X' or 'Not X'. To reduce the amount of noise, all of the trials are averaged within a subject and a condition. There are 726 (363×2) averaged ERPs.

As a small example, we will use 20 subjects from the ADHD group and 20 from the TD group. The interest will lie in the differences between the typically developing and ADHD

children overall and the differences between the conditions "X" and "Not X" for the CZ electrode. Additionally, we will cut down every ERP to 0 to 0.5s. The Figure 1 shows the ellicited ERP waveforms broken down by group.

Model

Let i denote the individual subject, $i = 1, ..., N$ and for our data $N = 40$, and t time measured in seconds, 128 timepoints between $[0, 0.5]$. Let z_{1i} be an indicator of ADHD diagnosis, 0 for typically developing and 1 for ADHD, and z_{2i} an indicator of task condition, 0 for "X" and 1 for "Not X". Let $p(i)$ denote each unique CZ ERP signal distribution, where $p(i) = 1, ..., P$. Outcome $y_i(t)$ is the CZ ERP signal for individual i.

We expect the ERP signal to have a time-specific mean, α_t and time-specific random intercept γ_i for each individual. The change in the mean ERP signal will depend on the condition, z_1, and diagnosis, z_2. Let δ_{z_1} and β_{z_2} be the regression coefficients for the condition and diagnosis. The model for any time t is

$$y_i(t) = \alpha_t + \gamma_{it} + \delta_{z_1} + \beta_{z_2} + \epsilon_{it}.$$

The population mean α_t, random intercept γ_{it}, and the coefficients δ_{t,z_1} and β_{t,z_2} to vary smoothly over time so we will model them as cubic B-spline functions of t. Let $B(t)$ be a cubic B-spline basis over time t with $(K - 4)$ knots.

Proper priors are used on all parameters. Let α_t be the coefficients of $B(t)$ for the population mean, then $\alpha_t = \alpha^T B(t)$. Similarly, δ_{t,z_1} and β_{t,z_2} are the coefficients for the other fixed effects so $\delta_{t,z_1} = \delta_{z_1}^T B(t)$ and β_{t,z_2}. Now the priors are

$$\alpha | c_\alpha \sim N_K(\alpha_0, c_\alpha I_K),$$

$$\delta_{z_1} | c_\delta \sim N_K(\delta_0, c_\delta I_K),$$

$$\beta_{z_2} | c_\beta \sim N_K(\beta_0, c_\beta I_K),$$

where I_K is the $K \times K$ identity matrix. We will set the prior mean α_0 to 0 with some uncertainty c_α equal to 16. The other fixed effects have prior means set to 0 to match the null hypothesis of no difference between diagnosis groups or task conditions. However, since we can see from the individual plots that this may not be true, we again set the prior variance fairly large, equal to 16 as well. Let γ_i be the coefficients of $B(t)$ for random intercept for individual i, then $\gamma_{it} = \gamma_i^T B(t)$,

$$\begin{aligned} \gamma_i | \Sigma_\gamma &\sim N_K(0, \Sigma_\gamma), \\ \Sigma_\gamma &\sim \text{Inverse-Wishart}(f, (f - K - 1) \times M). \end{aligned}$$

We will assume that the remaining individuals in the dataset are part of a separate study and use their estimates of the matrix M found using a classical linear mixed model. In this case, f is equal to the number of individuals in that set, 324. Lastly, we want to model the error with non-constant variance that will vary smoothly over time. The variance will be modeled with a separate quadratic B-spline basis over time, $B_{\text{err}}(t)$, with $(L - 3)$ knots. The prior for the variance is

$$\begin{aligned} u_{it} &\sim N(0, \sigma_t^2), \\ \sigma^2 &= \eta^T B_{\text{err}}(t) + w_t, \\ w_t &\sim N(0, \tau_\eta^2), \\ \eta &\sim N_L(\eta_0, g^2 I_L), \end{aligned}$$

where the w_t are small normal residual terms with known variance τ_η^2 and the value of g is chosen so that 95% of the residual variances are between 0.2 and 0.7. I attempted to add an autoregressive lag one correlation and the resulting model was worse than the one without it so I left it as this.

A sensitivity analysis was conducted on the number of knots chosen in the cubic B-spline. The R and JAGS code is included in Appendix B.

Results

The autocorrelation and time series plots were inspected. The autocorrelation went down to zero instantly at lag 1 and the time series looked like grass so the convergence for this model is great. Figure 2 also highlights that our model is fitting the data well. However, the residuals plots, Figure 3, may give evidence that our variance structure may need to be changed since there seems to be a time trend in the residuals.

The coefficients of the B-splines are in Table 1. From this table we see that many of the δ coefficients have confidence intervals that cross zero, which is evidence that there is no difference in the diagnosis groups. However, we also see that many of the β coefficients do not cross zero, which is evidence against the null hypothesis so there may be a difference in task conditions.

References

[1] Heidi J. Fischer, Qunfang Zhang, Yifang Zhu, and Robert E. Weiss. Functional time series models for ultrafine particle distributions. *The Annals of Applied Statistics*, 11(1):297–319, 2017.

Appendix A: Tables and Figures

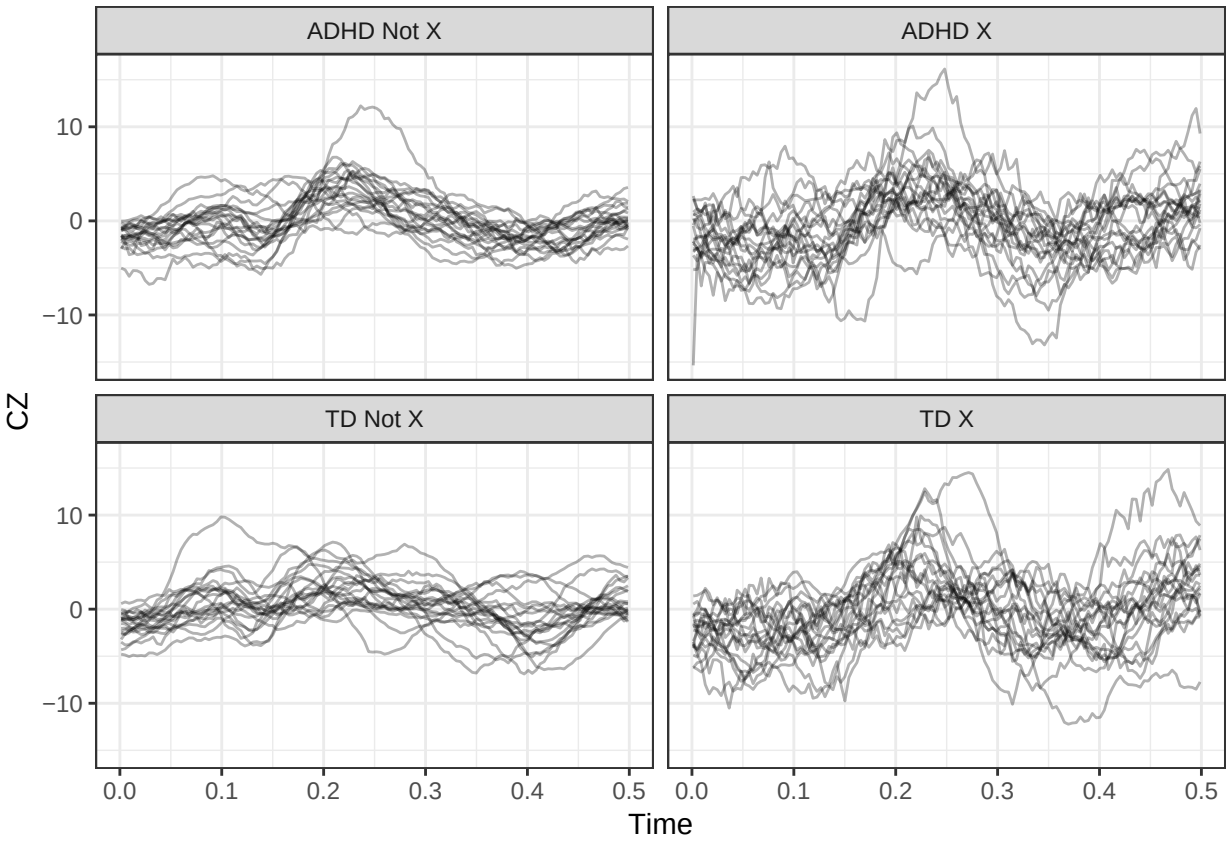

Figure 1: Individual curves for electrode CZ. Each individual in the TD and ADHD groups are plotted here for each task condition. There is clearly an average curve here, with deviations per condition and diagnosis group. We also note there is more variation in the X case due to the fact there are fewer 'X' trials than 'Not X' trials.

Table 1: Model Results

Parameter	Mean	SD	2.5%	97.5%
Overall Mean				
alpha[1]	-1.96	0.33	-2.61	-1.31
alpha[2]	-2.56	0.41	-3.36	-1.75
alpha[3]	-2.54	0.39	-3.31	-1.77
alpha[4]	-0.73	0.32	-1.37	-0.09
alpha[5]	6.93	0.49	5.97	7.87
alpha[6]	-8.22	0.48	-9.15	-7.27
alpha[7]	3.44	0.40	2.66	4.21
alpha[8]	2.19	0.33	1.55	2.84
Task				
beta[1]	-0.23	0.27	-0.76	0.29
beta[2]	1.23	0.36	0.52	1.93
beta[3]	3.44	0.28	2.89	3.97
beta[4]	0.65	0.18	0.30	1.00
beta[5]	-1.70	0.24	-2.17	-1.23
beta[6]	3.93	0.28	3.39	4.48
beta[7]	-3.26	0.23	-3.72	-2.81
beta[8]	-0.76	0.24	-1.24	-0.27
Diagnosis				
delta[1]	0.58	0.43	-0.26	1.41
delta[2]	-0.24	0.52	-1.24	0.77
delta[3]	-0.48	0.51	-1.46	0.52
delta[4]	-0.56	0.44	-1.43	0.31
delta[5]	1.38	0.66	0.09	2.68
delta[6]	-1.83	0.65	-3.10	-0.56
delta[7]	0.67	0.53	-0.38	1.72
delta[8]	-1.65	0.43	-2.49	-0.80

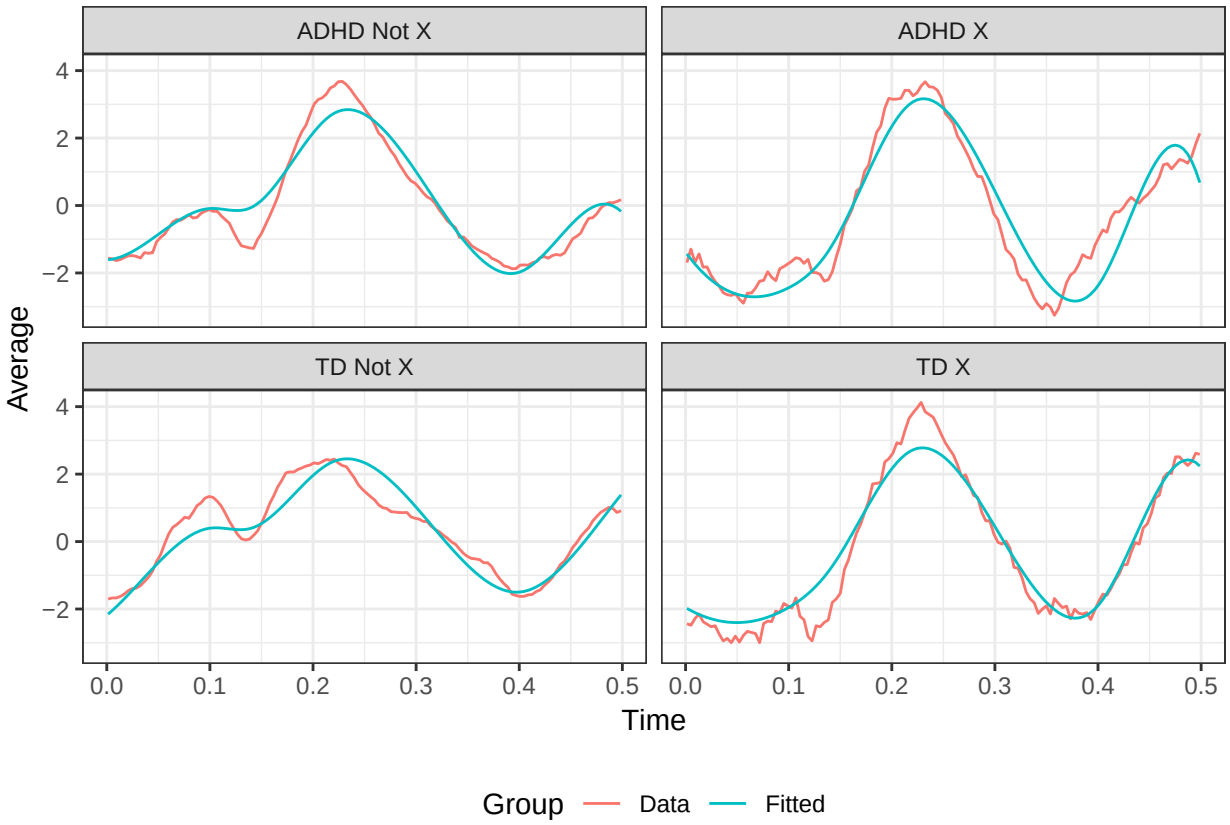

Figure 2: Comparing Fitted and Data Averages. The red curve in each plot is the average curve for that group and the blue line is the curve formed by adding the appropriate terms. For example, in the ADHD X case: $\alpha^T B(t) + \delta^T B(t) + \beta^T B(t)$. We note that the model used seems to reflect the data well.

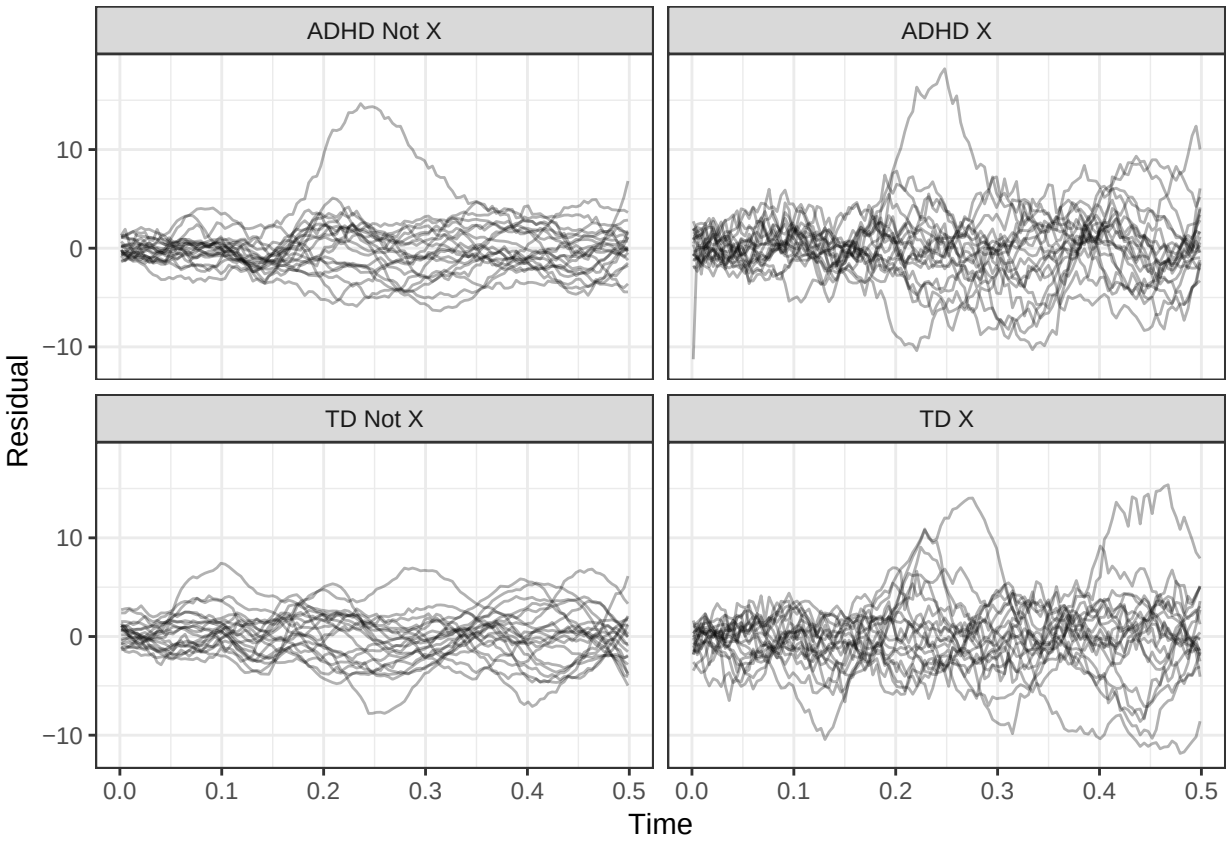

Figure 3: Residuals. The residuals for each individual should be randomly dispersed around zero but we found they have a time trend. In further analysis we would want to tweak the lag one autocorrelation between the points.

Appendix B: R and JAGS Code

```
model {
  # likelihood contribution
  for (r in 1:rows) {
    y[r] ~ dnorm(y.hat[r], tau[Subject[r],])
    y.hat[r] <- B[r, ] %*% (alpha + gamma[Subject[r],] + delta * Z[r, 1] +
                    beta * Z[r, 2])
  }

  # fixed effect
  alpha ~ dmnorm(mu0_alpha, precp_alpha)
  delta ~ dmnorm(mu0_delta, precp_delta)
  beta  ~ dmnorm(mu0_beta, precp_beta)

  # random effect
  for (j in 1:N) {
    gamma[j, 1:knots_plus_4] ~ dmnorm(mu0_gamma, Sigma_gaminv)
    sigma[j, 1] ~ dnorm(B_ncv[j, ] %*% eta, 100)
    tau[j, 1] <- 1/sigma[j, 1]
  }

  Sigma_gaminv ~ dwish(SG_prior, n_gammaprior)
  SigGam <- inverse(Sigma_gaminv)
  eta ~ dmnorm(mu0_eta, precp_eta)
}
```