Regression Spline Mixed Models for Analyzing EEG Data and Event-Related Potentials

Karen E. Nielsen and Rich Gonzalez
Department of Statistics, University of Michigan, Ann Arbor, MI

UNIVERSITY OF MICHIGAN

BIO•SOCIAL METHODS Collaborative
New models. New perspectives. New discoveries.

EEG Background

Electroencephalography (EEG) is the measurement of electrical activity of the brain via electrodes placed on the scalp. Neuronal action potentials within the brain result in electrical patterns which are detected by the scalp electrodes.

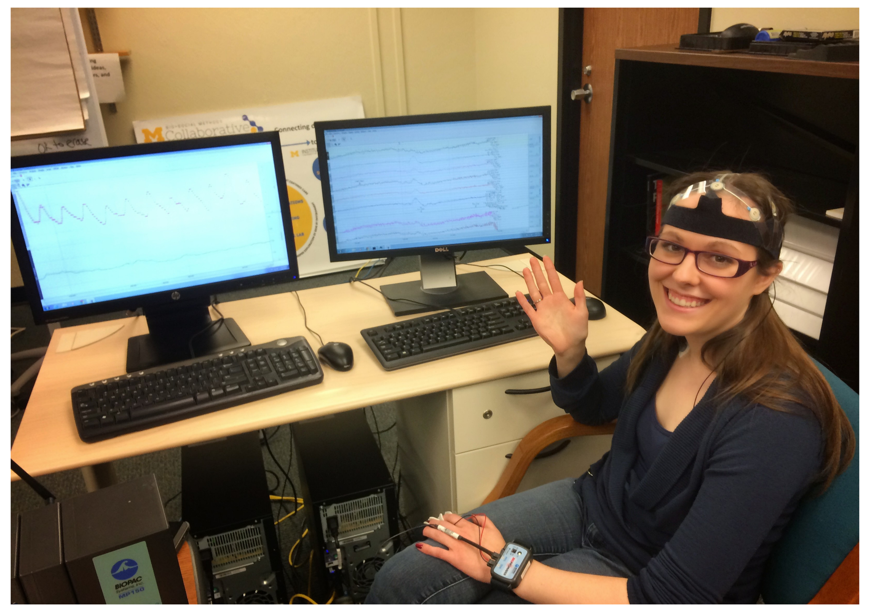

Figure 1 : The EEG cap sends data to a computer to be viewed in real time or analyzed later.

Since these electrodes are placed on the scalp and not the brain directly, and because the electrical potential of a single neuron is too small to capture individually, EEGs are noisy representations of brain activity. This presents a unique set of challenges to researchers. Statisticians are accustomed to working with noisy data, and thus possess skills that may provide valuable new insights into neuronal data.

Event-Related Potentials

This work focuses on Event-Related Potentials (ERPs). An ERP is a brain response to a time locked stimulus and is measured using EEG. ERPs are generally recorded in the first second after stimulus presentation, and an ERP waveform consists of one or a few components. Some of these components are well known, and variance in amplitude within each component is important during analysis. For example, a P300 is a positive deflection of voltage that occurs at approximately 300 milliseconds after a novel stimulus is presented. Historically, this waveform was inspected during polygraph tests for lie detection.

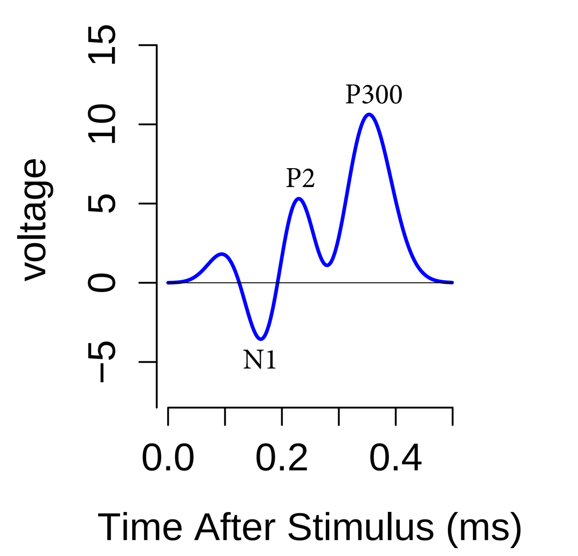

Figure 2 : Sample ERP data with known components

Notice here that we have broken from convention - in most EEG literature, negative voltages are plotted upwards on the y axis. We are plotting positive values upwards throughout this work.

Mixed-Effects Models

Mixed-effects models (also known as multilevel models; varying-intercept, varying-slope models; or hierarchical models) take into account the variation between groups. These models include fixed effects which do not change across groups, and random effects which do. More formally, we can write:

$$Y_{ij} = X_{ij}^T B + Z_{ij}^T \beta_i + e_i \tag{1}$$

for the outcome Y_{ij} of individuals i and groups j where X_{ij} is a design matrix for fixed effects, Z_{ij} is a design matrix for random effects, the coefficients $\beta_i \sim N(0, G)$, and errors $e_i \sim N(0, R_i)$.

In the context of ERP research, we can consider the channels to be groups within an individual's waveform (here), or each individual to represent a group of trials contributing to a population-level waveform, or we can create a hierarchical model that includes both. Any of these models can include additional variables, such as age, gender, or disease status.

Regression Splines in Context

Regression splines are one option for fitting EEG and ERP waveforms. Landmark points, such as local maxima and minima, make natural choices for knots. Smoothing splines are also an option, but resulting waveforms vary substantially based on choice of tuning parameter. Ideally, a functional form could be defined so that coefficients have meaningful interpretations to researchers.

A natural cubic spline $g(t)$ for $t_{min} \leqslant t \leqslant t_{max}$ with r knots $\tau = (\tau_1, ..., \tau_r)'$ such that $t_{min} < \tau_1 < ... < \tau_r < t_{max}$ satisfies:

- $g(t)$ is a piecewise cubic polynomial of degree $\leqslant 3$ on each $[t_{min}, \tau_1], ..., [\tau_r, t_{max}]$.
- $g(t), g'(t)$, and $g''(t)$ are continuous on $[t_{min}, t_{max}]$.
- $g^{(d)}(t_{min}) = g^{(d)}(t_{max}) = 0$ for $d = r+1, ..., 2r$. That is, $g(t)$ is linear beyond the knots.

A regression spline model can be written:

$$Q_i = \sum_{k=1}^{r} c_{ik} s_k(t_i) + e_i \tag{2}$$

where $s_k(x)$ are basis functions and $c_i = (c_{i1}, ..., c_{ir})'$ are unknown coefficients.

Here, we find basis functions for each observation (one channel for one trial) and allow these to have both fixed and random contributions in the model. Thus, Q_i carries a (suppressed) label for its channel, j, and is part of both X_{ij} and Z_{ij} in equation 1 (along with any other variables of interest).

Example

To demonstrate the use of Regression Spline Mixed Models (RSMMs) in the context of ERPs, we fit a model to a single individual's 8-channel EEG over a 1400-millisecond window using the lme4 package in R, which finds Restricted Maximum Likelihood (REML) estimates of the coefficients for the spline basis functions.

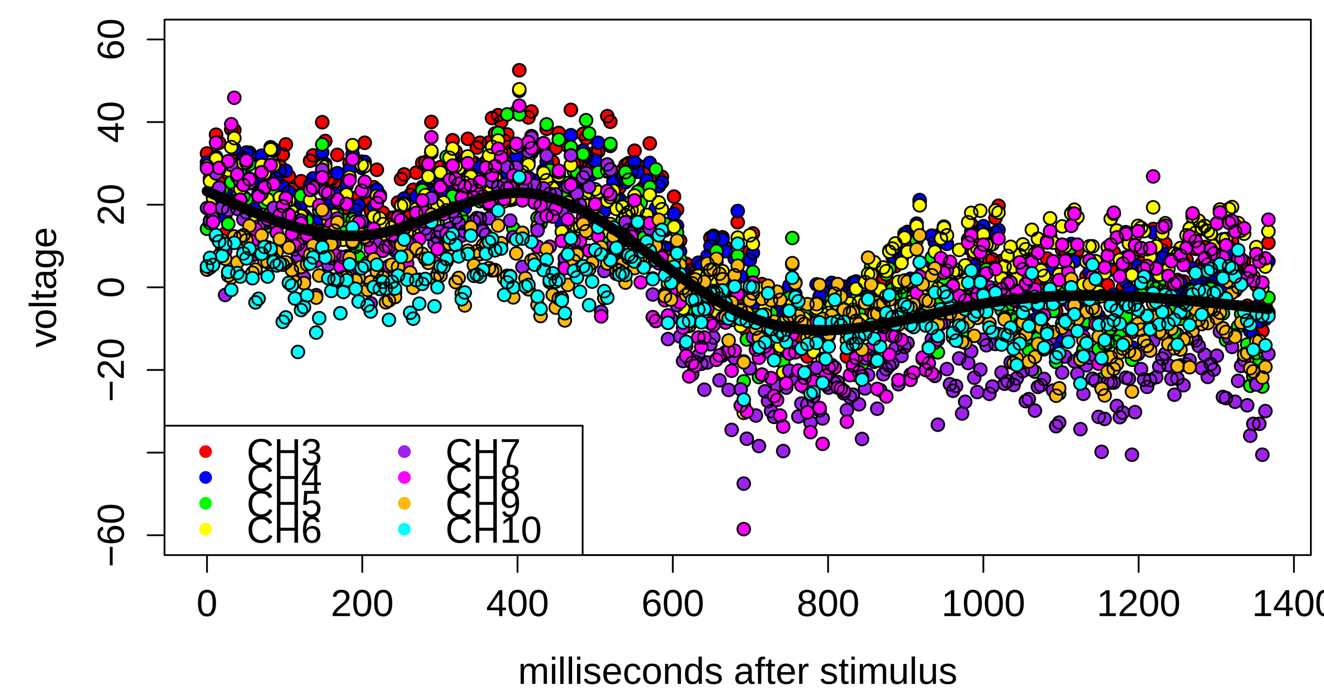

Figure 3 : Individual's fixed effect curve fit to multi-channel data

Below, curves representing the random effects for each channel are shown along with the fixed effect curve. We can see that, while the fixed effect curve is a representative summary of the overall data, each channel has its own features.

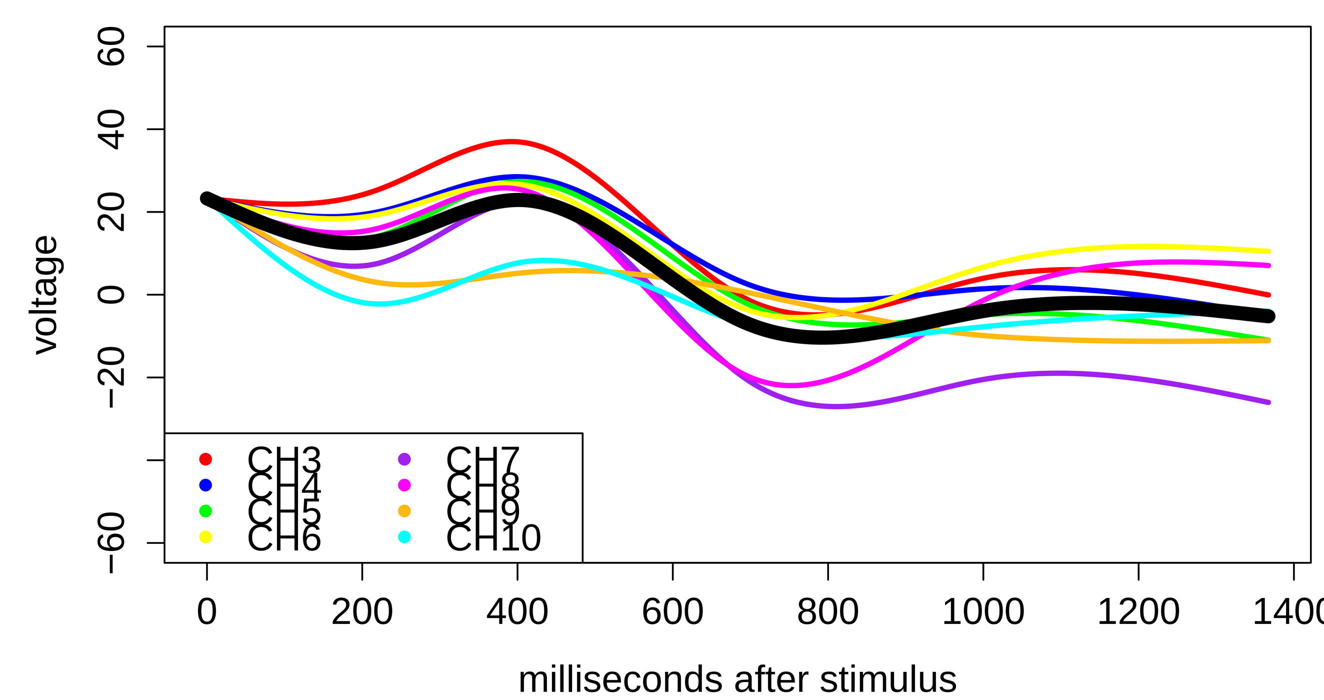

Figure 4 : Individual's random (colors) and fixed (black) effect curves fit to multi-channel data

As shown in Figure 4, each channel has its own features. Since regression splines place a particular set of assumption on the waveform, it is important to assess the fit of curves at several levels in the model hierarchy. Below, data from each channel is shown with the overall fixed effect and individual random effect curves.

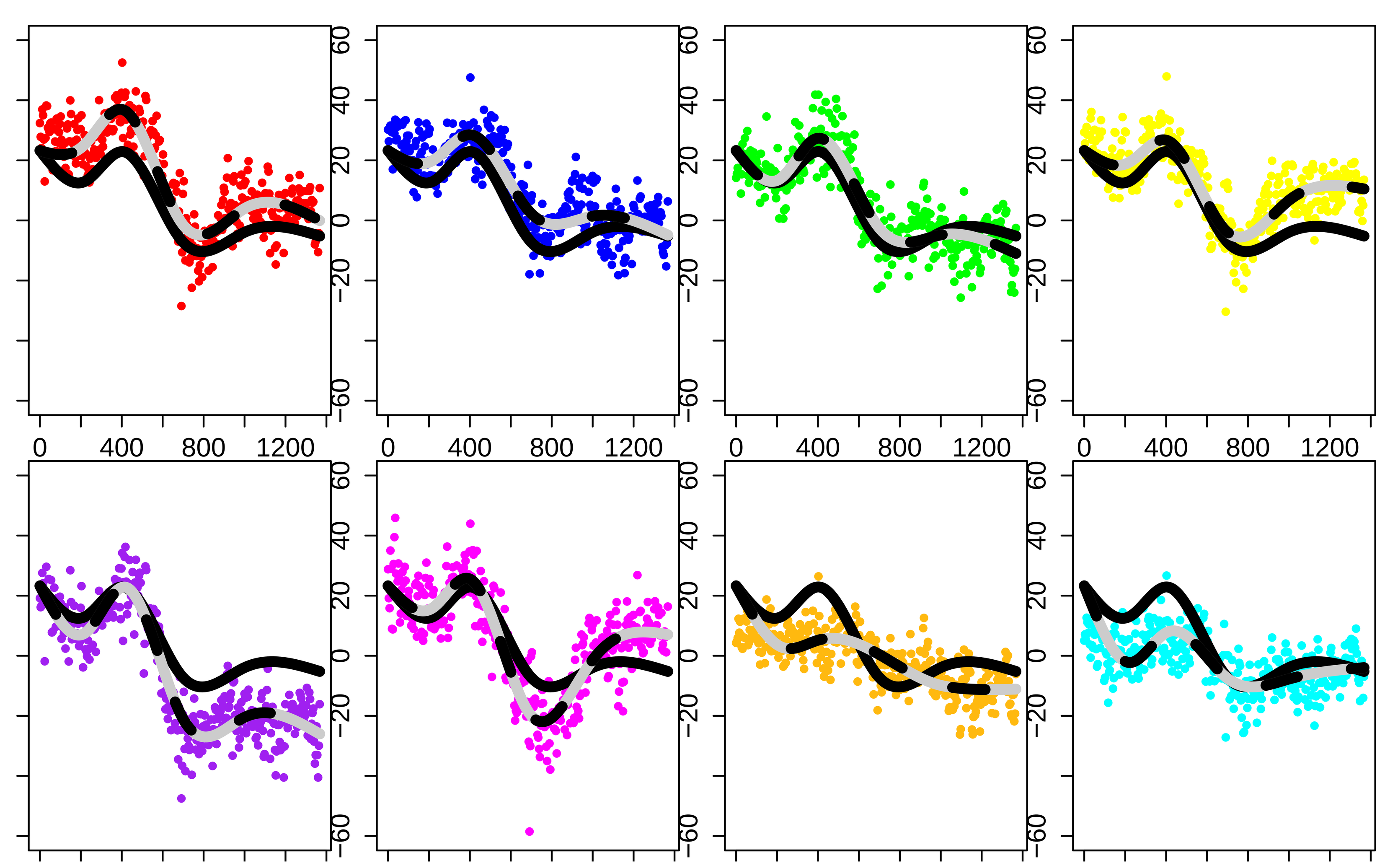

Figure 5 : Observations from each channel with curves for overall fixed (solid black) and random (dashed) effects.

In addition to a visibly representative fit to the subject's EEG, this modeling approach allows us to calculate standard errors, and thus confidence bands, on the waveform at various levels.

Other Bases

Regression splines are just one possible choice of basis set. While they are a standard approach with well-studied properties, they do not have meaningful coefficients. By using the functional form of a normal distribution kernel as a basis set, the coefficients for fixed and random effects become interpretable for researchers. The time latency (such as 100 milliseconds for a P100) and amplitude (voltage) are of particular interest. These coefficients are estimated with standard errors, meaning that confidence bands and hypothesis tests can be conducted on these values.

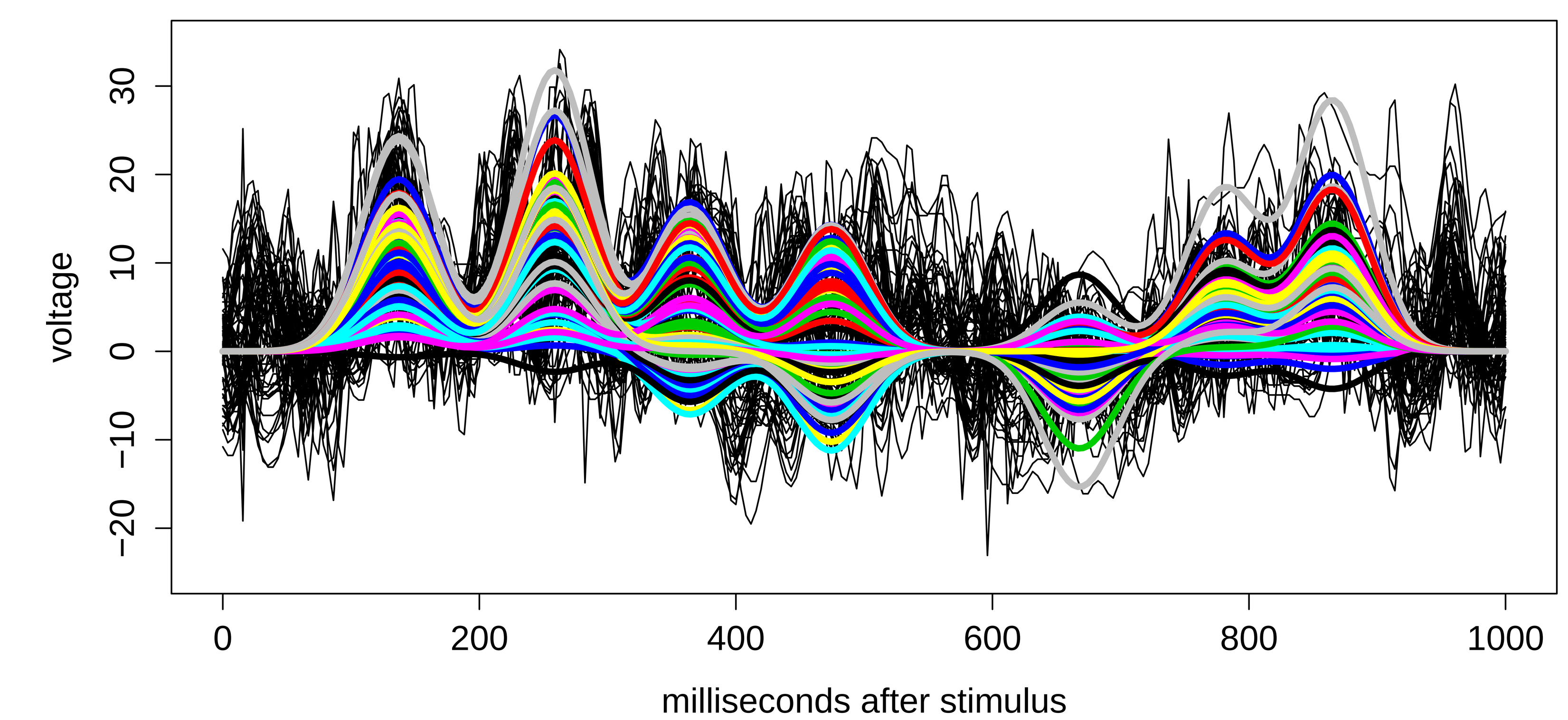

Figure 6 : Observations from each channel (black) with curves for random effects (colored).

We made use of the nlme package in R, which will fit coefficients for any functional form as both fixed and random effects. This approach currently requires the user to give the number of peaks to estimate. It is also sensitive to the choice of start points. Visual inspection usually gives a good impression of whether the model has been misspecified, and how to update these values before re-running the algorithm.

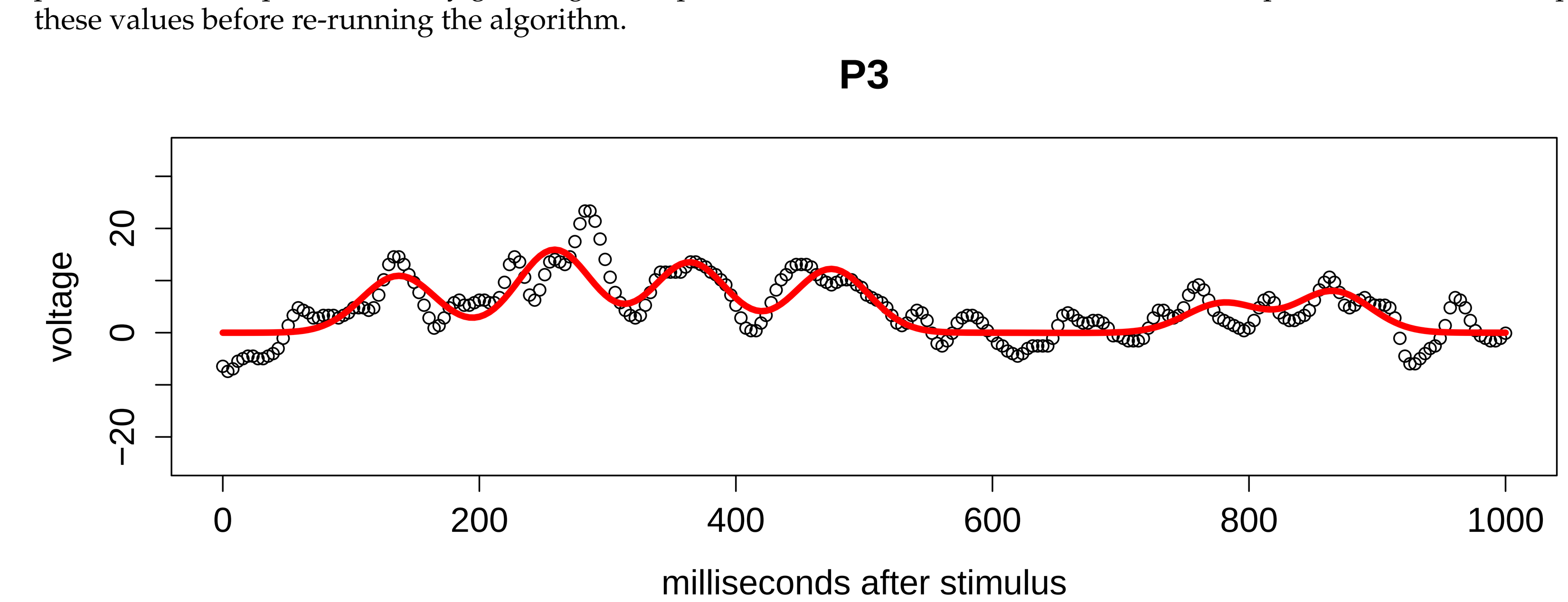

Figure 7 : Observations from a single channel (points) with a curve for the fitted model for this channel.

The data shown here is a subset of an EEG data set from Henri Begleiter at the Neurodynamics Laboratory at the State University of New York Health Center at Brooklyn, made available through the UCI repository. Shown in Figure 6 are 64 channels of EEG data for a single trial of a single subject. Fixed effects determine latencies and an average waveform, and random effects for each channel determine separate amplitudes. For example, Figure 7 shows the model fit for one channel by combining fixed and relevant random effects.

Future Work

- Establish a standardized processing stream for ERP data.
- Use random effects for baselining EEG/ERP data.
- Develop a formal test for identifying "uncertain" responses in arrow flanker and similar exercises. These responses are currently categorized as "correct" or "incorrect" and removed by hand.
- Explore the scientific accuracy of a "random walk" analogy to the paths of electrical potentials in the brain.

References

- Bates, D., Machler, M., Bolker, B., & Walker, S. Fitting Linear Mixed-Effects Models using lme4, 2014.
- Edwards, L. J., Stewart, P. W., MacDougall, J. E., & Helms, R. W. A method for fitting regression splines with varying polynomial order in the linear mixed model. Statistics in Medicine, 25(2005), 513-527.
- Fitzmaurice, G. Longitudinal Data Analysis. Chapman & Hall, 2009.
- Gelman, A. & Hill, J. Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge University Press, 2007.
- Lichman, M. UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science, 2013.
- Luck, S. An Introduction to the Event-Related Potential Technique. The MIT Press, 2005.
- Nurnberger, G. Approximation by Spline Functions. Springer, 1989.
- Pinheiro J, Bates D, DebRoy S, Sarkar D and R Core Team. nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-120, URL: http://CRAN.R-project.org/package=nlme, 2015
- Wand, M. P. A Comparison of Regression Spline Smoothing Procedures. Computational Statistics. 15(2000), 443-462.

contact: karenen@umich.edu