Hi Ed and Bruce,

I wanted to show you what I’ve found re: using MLMs for different types of ERP data. When I first did this on a smaller data set (51 subjects and 3 electrodes), it looked like the trial by trial (TBT) data was too noisy to get anything of interest out of the model and that binning a small number of trials together to use as a unit of data was preferable. After running some more models with a larger data set (65 subjects and 7 electrodes) and using electrode as a grouping variable, it seems like all three approaches (individual averages, TBT and binned data) actually had very similar patterns of results.

I ran these models on some or all types of data (by the way, we’re looking at the P2 in all cases, which is quantified as the mean amplitude in the same time windows for each type of data. The only difference is how many trials get averaged together before quantifying the P2. For TBT, P2 was separately quantified for each trial. For Binned, 8 trials of the same trial type were averaged together before P2 was quantified. Each subject had about 8-15 bins for each condition. For Individual averages, all trials of the same trial type were averaged together before P2 was quantified.)

Model 1:

Grouping variables: subject, electrode

Fixed effects: race, fixation, their interaction

Random effects: all slopes by subject, intercept by electrode

Model 2:

Grouping variables: subject, electrode, bin

Fixed effects: race, fixation, their interaction

Random effects: all slopes by subject, intercept by electrode and bin

Model 3:

Grouping variables: subject, electrode, trial

Fixed effects: race, fixation, their interaction

Random effects: all slopes by subject, intercept by electrode and trial

Here are the fixed effects for the different types of data:

StartFragment

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | Individual average | Binned | Trial by trial |
| Model 1 | Race | -0.85, p<.001 | -0.82, p<.001 | -0.84, p<.001 |
|  | Fixation | -0.50, p = .012 | -0.51, p = .012 | -0.54, p = .007 |
|  | Race\*Fixation | .26, p = .175 | .26, p = .196 | .30, p = .116 |
|  |  |  |  |  |
| Model 2 | Race | - | -0.82, p<.001 | - |
|  | Fixation | - | -0.51, p = .012 | - |
|  | Race\*Fixation | - | .26, p = .192 | - |
|  |  |  |  |  |
| Model 3 | Race | - | - | -0.85, p<.001 |
|  | Fixation | - | - | -0.57, p = .004 |
|  | Race\*Fixation | - | - | .35, p = .070 |

EndFragment

As you can see, they’re pretty much the same across the different models and types of data. The individual BLUPs (predicted intercepts and slopes at each level of the grouping variables) from each model were also pretty consistent, which is totally different from the models that I ran on the smaller data set (i.e. the predicted intercept for each subject were wildly different for the individual average data and the TBT data, although the subject intercepts from the Binned data roughly matched the individual average data). Not sure why this time around the BLUPs across all types of data match. Possibly due to increasing amount of data for model to use (number of subjects and number of electrodes) or including electrode as grouping variable (with 7) instead of fixed effects (just 3)?

The nice thing about binning or using trial by trial data is that you can look at the effect of trial or bin to see change over time. I ran the following models on the Binned and TBT data to look at that.

Model 4.1:

Grouping variables: subject, electrode, bin

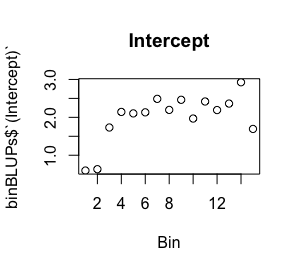
Fixed effects: race, fixation, their interaction

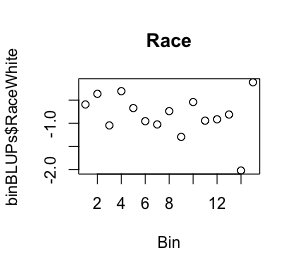
Random effects: all slopes by subject, all slopes by bin, intercept by electrode

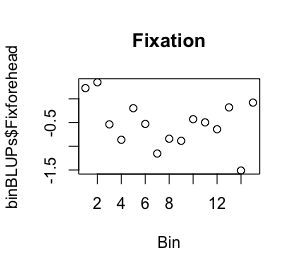
Model 4.2:

Same as Model 4.1 but with trial instead of bin

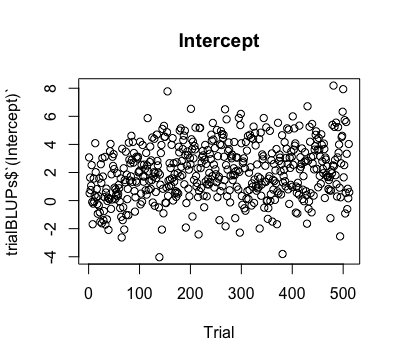
Here are the BLUPs for intercept, race, and fixation from Model 4.1 (Binned data):

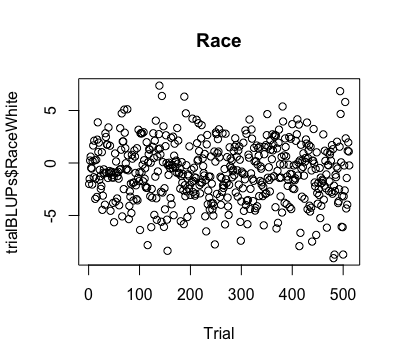


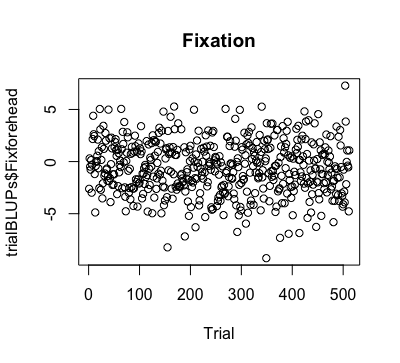




Here are the same BLUPs from Model 4.2 (TBT data):







To test the effect of bin/trial, I ran the same models with Bin as a fixed effect (with interactions with all other variables) instead of a grouping variable.

Model 5.1:

Grouping variables: subject, electrode

Fixed effects: race, fixation, bin, all interactions

Random effects: all slopes by subject, intercept by electrode

Model 5.2:

Same as Model 5.1 but with trial instead of bin

Here are the fixed effects of bin/trial and its interactions with race and fixation:

StartFragment

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Binned | Trial by trial |
| Model 5.1 | Bin | .122, p<.001 | - |
|  | Bin \* Race | -.04, p = .002 | - |
|  | Bin \* Fix | -.05, p<.001 | - |
|  |  |  |  |
| Model 5.2 | Trial | - | .003, p<.001 |
|  | Trial \* Race | - | -.002, p<.001 |
|  | Trial \* Fix | - | -.002, p<.001 |

EndFragment

So when looking at the BLUPs, it looks like the change in the effect of Race or Fixation is pretty small over time, which is confirmed by the tiny estimates of the race and fix by trial/bin interactions, but the change is highly significant (also, I’m not quite sure I’m interpreting these fixed effects correctly or whether it’s appropriate to test change in effect of Race or Fixation over time in this way). Also not sure I really see the relationships indicated by the fixed effects in the plots of the BLUPs for the TBT data though (looks just like a cloud to me…)

**Summary**:

Models using all three types of data had very similar patterns of results, so not sure if there’s an advantage to TBT or binned if you’re just interested in fixed effects. Using TBT or binned data does allow you to look at habituation/sensitivity/change in effects over time if you’re interested in that, which isn’t possible with individual averages (and these data show significant changes in P2 amplitude over time, as well as significant change in the effect of Race and Fixation over time). As for whether binned or TBT data is more advantageous, it seems like the model output is pretty similar, but models with the binned data run much faster (~2-5 minutes) than models using the TBT data (>30 minutes). So that time advantage might make binning data more advantageous.

I’ve attached the full model results, as well as the original data and R scripts I used to run the models (zipped together). Ed, they’re not all in one R script as you suggested but I tried to relabel everything so it’s clearer and hopefully easier to understand. Let me know if you want more clarification. Also, it would be possible for me to run a similar set of analyses on a different component, the N170, which is typically calculated as a peak amplitude within a time window (instead of average amplitude) to see if there are clearer differences between using different types of data for MLMs (quantifying peak amplitude is usually more sensitive to noise, so we may see more of an advantage of individual averages/binning over TBT data here). I’m not planning on doing that right away but might get to it in a couple of weeks.

If you have the time to give me feedback, it would be much appreciated! Thanks!