Title
Identifying Longitudinal Trends within EEGExperiments

Authors
Hasenstab, Kyle
Sugar, Catherine A.
Telesca, Donatello
et al.

Publication Date
2015-01-27

Peer reviewed

Identifying Longitudinal Trends within EEG Experiments

Kyle Hasenstab[1], **Catherine A. Sugar**[1,2,3], **Donatello Telesca**[2], **Kevin McEvoy**[3], **Shafali Jeste**[3] **and Damla Şentürk**[1,2*]

[1]Department of Statistics, University of California,
Los Angeles, California 90095, U.S.A.
[2]Department of Biostatistics, University of California,
Los Angeles, California 90095, U.S.A.
[3]Department of Psychiatry and Biobehavioral Sciences, University of California,
Los Angeles, California 90095, U.S.A.
email: dsenturk@ucla.edu*

Summary

Differential brain response to sensory stimuli is very small (a few microvolts) compared to the overall magnitude of spontaneous electroencephalographam (EEG), yielding a low signal-to-noise ratio (SNR) in studies of event-related potentials (ERP). To cope with this phenomenon, stimuli are applied repeatedly and the ERP signals arising from the individual trials are averaged at the subject level. This results in loss of information about potentially important changes in the magnitude and form of ERP signals over the course of the experiment. In this paper, we develop a meta-preprocessing step utilizing a moving average of ERP across sliding trial windows, to capture such longitudinal trends. We embed this procedure in a weighted linear mixed effects model to describe longitudinal trends in features such as ERP peak amplitude and latency across trials while adjusting for the inherent heteroskedasticity created at the meta-preprocessing step. The proposed unified framework, including the meta-processing and the weighted linear mixed effects modeling steps, is referred to as MAP-ERP (Moving-Averaged-Processed ERP). We perform simulation studies to assess the performance of MAP-ERP in reconstructing existing longitudinal trends and apply MAP-ERP to data from young children with autism spectrum disorder (ASD) and their typically developing counter parts to examine differences in patterns of implicit learning, providing novel insights about the mechanisms underlying social and/or cognitive deficits in this disorder.

Keywords: event-related potentials data; heteroskedasticity; repeated measurements; signal-to-noise ratio; smoothing; weighted linear mixed effects models

1 Introduction

Both spontaneous electroencephalographam (EEG) signals and event-related potentials (ERP), which represent EEG recorded in response to stimuli, are noninvasive methods for measuring brain activity with very high time resolution. They have been in use since the 1950's in diverse biomedical settings including epilepsy, sleep disorders, multiple sclerosis, brain tumors, lesions, major affective disorder, schizophrenia, alcoholism, bipolar mood disorder, assessment of surgical outcomes, confirmation of brain death and clinical trials for drug development (Gasser and Molinari, 1996; Tierney et al., 2012). An ERP waveform (curve/signal/morphology) consists of characteristic components that span time. A typical ERP waveform from an implicit learning paradigm, containing the commonly studied P3 and N1 phasic components in this literature, is given in Figure 1 (a). Note that ERP waveforms can contain multiple phasic components, and the focus of analysis may be on different features in other applications. In our working example the N1 dip has a short latency (time-delay) and signifies early category recognition, while the P3 peak is task dependent due to its long latency and is traditionally related to cognitive processes such as signal matching, decision making and memory updating (Bugli and Lambert, 2006; Jeste et al., 2014).

The analysis of ERP data is usually performed in the time domain. Differential brain response to sensory stimuli is very small (on the order of a few microvolts) as a fraction of spontaneous EEG, yielding a low signal-to-noise ratio (SNR) in ERP studies. To cope with this well-known phenomenon, stimuli are applied repeatedly and the resulting ERP waveforms are averaged across the trials for each subject (Gasser and Molinari, 1996; Delorme and Makeig, 2004; Tierney et al., 2012). Other common preprocessing steps include artifact detection (of irregularities in the signals caused by events such as blinks, saccades or muscle contractions), bad channel replacement (a smoothing procedure performed when signals from an individual scalp electrode are compromised), referencing (to standardize measurements to an average

across all the electrodes) and baseline corrections (to standardize a subject's measurements to their own baseline period).

Since EEG has high time resolution and low spatial resolution (the space component corresponds to the ERP signals measured simultaneously at different electrodes placed on the skull), primary analyses typically focus on the amplitude (microvolts) and latency (milliseconds) of specific peaks (e.g., N1, P3) in the averaged ERP curves for each subject instead of spatial features (Luck, 2005). As a result of averaging ERP across trials within subjects, the traditional methods fail to capture potentially informative changes in ERP signals which may emerge over the course of the experiment. In this paper we focus on modeling the longitudinal component of the data structure corresponding to the trials resulting from the repeated stimuli. We propose a meta-preprocessing step based on applying a moving average across trials to preserve the longitudinal information in ERP data. The proposed meta-preprocessing procedure strikes a balance between the need to average over trials to enhance the signal-to-noise ratio and 'over-averaging' to the point where the data for each subject are reduced to a single overall ERP and all longitudinal information is lost. This procedure is embedded in a weighted mixed effects regression framework to allow modeling of the resulting longitudinal data. The proposed unified framework comprising of the meta-processing and the weighted linear mixed effects modeling steps is referred to as MAP-ERP (Moving-Averaged-Processed ERP) throughout the manuscript.

Our working example is from a study on neural correlates of implicit learning in young children with autism spectrum disorder (ASD) (Jeste et al., 2014). ASD is a neurodevelopmental disorder defined by impairments in social behavior, communication and the presence of restricted interests. Implicit learning is characterized by detection of regularities in one's environment without a conscious awareness or intention to learn (Travers et al., 2010; Jeste et al., 2014). The goal of the study is to provide insights about pathways to core deficits in ASD

through a better understanding of implicit learning, which is thought to play a critical role in social behavior (Jeste et al., 2014). The study involved 2 to 5 year old typically developing (TD) and ASD children, exposed to a continuous stream of six colored shapes (pink squares, blue crosses, yellow circles, turquoise diamonds, gray triangles and red octagons; see Figure 1 (b)). The shapes were organized into three pairs such that the sequence within the pairs was always the same but the pairs themselves occurred in random order across the experiment. For example, a pink square was always followed by a blue cross (pair) but the symbol that followed a blue cross could be any of the three 'first' symbols of a pair. As a result, the transitions within a pair (pink square to blue cross) were 'expected' or could be 'learned' while the transitions between pairs were 'unexpected' or not predictable. Each transition from one shape to the next was considered a stimulus, resulting in an ERP waveform. Differences in the ERP signals (amplitude, shape, timing) between the expected and unexpected trials were thought to reflect the degree of implicit learning in young children while longitudinal changes in this contrast would indicate how the learning process evolves. The paradigm included 120 trials (repeated stimuli) for each of the two conditions, resulting in 240 ERP waveforms per child.

The goal of the original study was to look at differences in implicit learning between TD and ASD children in order to provide insights about mechanisms underlying social and/or cognitive deficits in this disorder. While the traditional analysis simply compares the average difference between expected and unexpected trials, determining whether the groups differ in terms of how the ERP signals change as the children learn the shape patterns over the course of the task is also important, e.g. for seeing whether implicit learning occurs at different speeds. Due to the low signal-to-noise ratio, it is also not feasible to analyze the original longitudinally collected ERP curves on a trial by trial basis. As an example, in Figure 2 (top left panel), an ERP waveform is plotted for one subject from a single trial in the right frontal

region of the brain. Due to the low signal-to-noise ratio in a single trial, the N1 dip and P3 peak are not recognizable in their standard respective time intervals as defined in Jeste et al. (2014). Now consider an average of 30 ERP waveforms from adjacent trials, plotted in Figure 2 (top right panel). The N1 dip and P3 peak are easily recognizable due to the enhancement in the signal-to-noise ratio. Our proposed meta-preprocessing procedure relies on moving averages of ERP waveforms over trials to preserve trends in implicit learning, focusing on the longitudinal analysis of features such as the P3 amplitude. These features are then assessed via a weighted mixed effects model, which adjusts for inherent heteroskedasticity created at the meta-preprocessing step.

Previous studies in neuroscience and biomedical engineering have acknowledged that ERP morphology may change over the course of a task. However, most prior work has focused on controlling for longitudinal trends (Coppola et al., 1978; Gasser et al., 1983; Woestenburg et al., 1983; Mocks et al., 1984; Mocks et al., 1987; Turetsky et al., 1989) rather than viewing them as one of the central research questions. Moreover, most of the work on modeling longitudinal trends has been limited to parametric forms (Woestenburg et al., 1983; Krieger et al., 1992; Krieger and Sclabassi, 1994; Rossi et al., 2007; De Silva et al., 2012). In contrast, our proposed meta-preprocessing step, utilizing the moving average idea described above, does not make any parametric assumptions. It is only after extracting the longitudinal features of the ERP morphology via a flexible nonparametric approach that we utilize a parametric model, the proposed weighted linear mixed effects regression, to describe changes in ERP features over the trials. Techniques such as sub-ensemble averaging and block averaging, which combine disjoint subsets of ERP waveforms across trials, require long recording sessions and still result in coarse-grained repeated measurements that do not take full advantage of the information on continuous longitudinal trends that is inherently available in the raw data (Verleger et al., 1985; Bruin et al., 2001). These studies provide snapshots of the longitudinal information,

while the proposed moving average captures the continuum of longitudinal dynamics.

The paper is organized as follows. In Section 2, we outline a novel meta-preprocessing step for ERP experiments which preserves the longitudinal structure of the data while enhancing the signal-to-noise ratio. A weighted linear mixed effects modeling approach following the meta-preprocessing step of Section 2 is outlined in Section 3. In Section 4, we present a simulation to assess the performance of MAP-ERP, the proposed unified modeling framework. The analysis of the ERP data from the motivating implicit learning study in ASD is given in Section 5. We conclude with final remarks in Section 6.

2 Proposed Meta-preprocessing Step to Preserve Longitudinal Information in ERP Data

Our working example is the first to date to study implicit learning in young children with ASD via the use of EEG. EEG data were recorded for 120 trials per condition (expected and unexpected) for each subject at 128 electrodes and were preprocessed using NetStation 4.4.5 software (Electrical Geodesics, Inc.). The standard preprocessing steps outlined in Web Appendix A produced trajectories of ERP waveforms for 37 ASD and 34 TD children to be fed into the proposed meta-preprocessing algorithm. The number of trials with usable data ranged from 10 to 120 per subject per condition. The EEG signals were sampled at 250Hz, producing 250 within-trial time points per waveform spanning 1000ms. To illustrate the methods, we focus on data from 12 electrodes in the frontal region which are described in Jeste et al. (2014).

The proposed meta-preprocessing step is a moving average designed to increase the signal-to-noise ratio of ERP to a level at which curve attributes such as peak amplitudes and latency are identifiable, while preserving longitudinal information. Let $X_{ijk\ell}(t)$ represent the microvoltage of the ERP of subject i from electrode j, on trial k in condition ℓ (expected/unexpected) observed at time t, $i = 1, \ldots, N$, $j = 1, \ldots, J$, $k \in K_i$, $\ell \in L_{ik}$ and $t = 1, \ldots, T$, where N is the total number of subjects, J is the total number of electrodes, T is the total number of

time points within a trial, K_i is the set of non-missing trials for subject i and L_{ik} is the set of non-missing conditions at trial k for subject i. The maximum number of conditions per subject per trial, denoted by L, is 2 in our application. The maximum possible number of trials per subject per condition, denoted by K, is 120. Further, let B_k represent overlapping sets of trials of varying lengths with the maximum number of trials within a set denoted by b,

$$B_k = \begin{cases} [1, 2k-1], & \text{if } k < \frac{b}{2}, \\ [k - b/2 + 1, k + b/2], & \text{if } \frac{b}{2} \leq k \leq K - \frac{b}{2}, \\ [2k - K, K], & \text{if } j > K - \frac{b}{2}. \end{cases}$$

Sets not on the boundary contain b elements and the number of elements shrink linearly towards 1 at the boundaries. The B_k are used as sliding trial windows in the moving average of ERP within the meta-preprocessing step. Since the goal is to extract continuous longitudinal trends within EEG experiments, including overlapping sets of trials helps to target the continuum of features across several ERPs. Alternative options include kernel smoothing; we chose to utilize a moving average in our applications to simplify the quantification of heterogeneity in these averages due to missingness. As will be outlined in Section 3, in the proposed weighted linear mixed effects model framework, longitudinal attributes captured from averaging over smaller numbers of ERP waveforms, as in the boundary sets or sets with a larger proportion of missing trials, will receive lower weights. This will help to mitigate boundary effects and place more weight on intervals with more trials, and hence more information.

In the algorithm introduced below, we let $\widetilde{X}_{ijk\ell}(t)$ represent the cross-sectional averages of ERPs within the sliding trial windows B_k and let $\widetilde{Y}_{ijk\ell}$, $i = 1, \ldots, N$, $j = 1, \ldots, J$, $k \in M_i$, $\ell \in Q_{ik}$, represent longitudinal features (such as amplitude or latency of peaks) of ERPs captured from these cross-sectional averages, where M_i is the set of non-missing trials for subject i and Q_{ik} is the set of non-missing conditions for subject i at trial k after meta-preprocessing. Note that the notation $Y_{ijk\ell}$ is reserved to represent the true features of the underlying ERP signal to be used in subsequent sections and that different set notations,

M_i and Q_{ik}, are used to index the observed trials for subject i and the observed conditions for subject i at trial k, respectively, to accommodate differences in the missingness structure which may be induced by the proposed meta-preprocessing step on the longitudinal features. The meta-preprocessing is applied separately for each subject, electrode and condition. The proposed algorithm can be summarized in the following steps.

1. For fixed i, j, k and ℓ, consider the set of ERP waveforms $\{X_{ijk\ell}(t) : k \in B_k\}$. If the set is empty, $\widetilde{Y}_{ijk\ell}$ will be missing and the algorithm proceeds to consider the set associated with trial $k+1$.

2. Calculate the cross-sectional mean curve, $\widetilde{X}_{ijk\ell}(t) = (1/c_{ijk\ell}) \sum_{k \in B_k} X_{ijk\ell}(t)$, of the subsetted waveforms, where $c_{ijk\ell}$ equals the number of trials in the set $\{X_{ijk\ell}(t) : k \in B_k\}$.

3. Smooth $\widetilde{X}_{ijk\ell}(t)$ over t to identify the locations of the ERP features of interest.

4. Use the locations from step 3 to define the longitudinal features $\widetilde{Y}_{ijk\ell}$ for $i = 1, \ldots, N$, $j = 1, \ldots, J$, $k \in M_i$, $\ell \in Q_{ik}$. If the feature of interest is a peak latency, then step 3 will be enough for the assignment of $\widetilde{Y}_{ijk\ell}$; if on the other hand it is a peak amplitude, then the locations from step 3 but the actual amplitudes will be computed using the values of the original cross-sectional mean waveforms, $\widetilde{X}_{ijk\ell}(t)$, rather than those from their smoothed version in order to minimize bias.

5. Repeat steps 1 through 4 for $k = 1, \ldots, K$.

The meta-preprocessing step extracts features from the dense ERP curves using a moving average across trials k, producing longitudinal features across trials for each subject, electrode and condition. We emphasize that the analysis after the meta-preprocessing is performed on the longitudinal time component (trials), hence the omission of the t notation in the longitudinal features $\widetilde{Y}_{ijk\ell}$. We use a loess smooth in the third step of the proposed algorithm where

the bandwidth is selected via 10-fold cross-validation. To identify the feature locations on the smooth, which correspond to the locations of the P3 component in our application, we utilize a peak detection algorithm which identifies optima within the time interval $t \in [190\text{ms}, 350\text{ms}]$ (Jeste et al., 2014). If the peak location is on the boundary of the specified time interval, the time interval is gradually widened in the direction of the initial boundary peak until a new peak is identified that is not on the boundary (Jeste et al., 2014). If the interval is widened to twice the range of the initial interval and the peak still lies on the boundary, the peak observation is considered missing. While the locations of the features are identified on the smoothed data to minimize the effects of random noise, the features themselves, such as peak amplitudes, are assigned on the original cross-sectional averages to minimize bias. In our application, the size of the sliding trial window, b, is chosen to be 30, which corresponds to the minimum number of curves required to increase the signal-to-noise ratio to a level where the desired features are recognizable. However we also include a sensitivity analysis to show that the data analysis results are sufficiently robust to this choice. For cases where the focus of interest may be the entire ERP waveform instead of the particular features, automatic selection of b was proposed by Turetsky et al. (1989), based on minimizing the estimated mean average square error of the smoothed data.

3 Analysis of Meta-preprocessed ERP Data via a Weighted Linear Mixed Effects Model

Mixed effects regression is a powerful modeling tool which accounts for multi-level heterogeneity and complex temporal trends via the use of fixed and random effects. We propose a weighted linear mixed effects model to analyze the longitudinal features $\widetilde{Y}_{ijk\ell}$, $i = 1, \ldots, N$, $j = 1, \ldots, J$, $k \in M_i$, $\ell \in Q_{ij}$, extracted by our meta-preprocessing algorithm. We reiterate that the longitudinal components of the proposed linear mixed effects model are the extracted features defined across trials k (not ERP time t). Our application focuses on ERP data from 12

electrodes in the front of the scalp, from left, middle and right regions. Our goals in the application to the implicit learning paradigm are to model the dynamics of P3 amplitude across trials and to study differences between the TD and ASD groups. We utilize multi-level random effects at the subject and electrode region levels to model dependency of the data for a given subject and a particular electrode region for a given subject where the spatial correlations are the strongest. We further utilize spline basis functions in modeling both fixed and random effects to describe the functional dependency of the P3 amplitudes across trials in a reduced dimensional space.

In addition, our proposal includes an adjustment to the standard linear mixed effects framework to account for the heteroskedasticity induced by averaging different numbers of trials at different time points during the meta-preprocessing step. More specifically, consider the plot of the P3 amplitudes (denoted by $\widetilde{Y}_{ijk\ell}$) produced by the proposed meta-preprocessing algorithm against values of $c_{ijk\ell}$, the number of ERP waveforms averaged in the B_k trial window in step 2 of the algorithm, displayed in Figure 3 (a). There is a sharp decrease in the variance of the amplitudes with increasing $c_{ijk\ell}$, suggesting that features extracted from averages in sliding trial windows with fewer ERP waveforms are less precise, as would be expected. In order to account for this systemic heteroskedasticity, separate variance components are allowed for different bins of $c_{ijk\ell}$ in the linear mixed effects formulation below. Note that similar weighting ideas have been considered previously to correct for heteroskedasticity in the context of averaging of ERP curves but only for cross-sectional data (Stahl et al., 2010).

Let s index a partition (grouping/binning) of the range of $c_{ijk\ell} = 1, \ldots, b$. The choice of the total number of bins will be a trade-off between model parsimony and realistic representation of the heteroskedasticity in the longitudinal features. Let $\widetilde{Y}^{(s)}_{ij(r)k\ell}$ represent the longitudinal features of the meta-preprocessed ERP, from subject i, at electrode j within region r, with trial window B_k and condition ℓ, where the number, $c_{ijk\ell}$, of ERPs averaged in B_k falls into

the sth partition. Note that the region index r is added in this notation, since our proposed modeling will be addressing dependencies in the data within electrodes from the same part of the scalp. We model the $\widetilde{Y}^{(s)}_{ij(r)k\ell}$ using fixed effects and a two level random effects structure (subject and region). Fixed effects parameters include an intercept, trial (represented by a natural cubic B-spline with 5 knots), condition (expected vs. unexpected) and group (ASD vs. TD), along with all two-way and three-way intersections, totaling 24 fixed effects components. Let $\boldsymbol{w}_{ik\ell}$ be the 1×24 row vector corresponding to trial k and condition ℓ of the fixed effects matrix $\boldsymbol{W}_i$ and let $\boldsymbol{\beta}$ be the 24×1 column vector of fixed effects parameters. Further let $\boldsymbol{z}_{ij(r)k}$ represent the 1×6 row vector (including an intercept and trial, represented by a natural cubic B-spline with 5 knots) corresponding to trial k of the random effects matrix $\boldsymbol{Z}_{ij(r)}$ for electrode j in region r. Index ℓ for condition is not needed for $\boldsymbol{z}_{ij(r)k}$ in our application, since the random effects design matrix contains spline bases to model within subject correlations over trials. Vectors $\boldsymbol{b}_i$ and $\boldsymbol{b}_{ir}$ of dimension 6×1 represent subject and region level random effects. We model $\widetilde{Y}^{(s)}_{ij(r)k\ell}$ by

$$\begin{aligned} &\widetilde{Y}^{(s)}_{ij(r)k\ell} = \boldsymbol{w}_{ik\ell}\boldsymbol{\beta} + \boldsymbol{z}_{ij(r)k}\boldsymbol{b}_i + \boldsymbol{z}_{ij(r)k}\boldsymbol{b}_{ir} + \varepsilon^{(s)}_{ij(r)k\ell}, \qquad (1) \\ &\boldsymbol{b}_i \sim MVN(\boldsymbol{0}, \boldsymbol{D}^1_{6\times 6}), \quad \boldsymbol{b}_{ir} \sim MVN(\boldsymbol{0}, \boldsymbol{D}^2_{6\times 6}), \quad \varepsilon^{(s)}_{ij(r)k\ell} \sim N\left(0, \sigma^2_s\right), \end{aligned}$$

where $\boldsymbol{D}^1$ and $\boldsymbol{D}^2$ represent the random effects covariance matrices at the subject and region levels, respectively, and $\varepsilon^{(s)}_{ij(r)k\ell}$ represents the error term for the sth partition with variance σ^2_s and uncorrelated over different subjects, regions, electrodes, trials and conditions. The subject and region level random effects are assumed to be independent of the error term, leading to the following covariance structure: $\text{Var}(\widetilde{Y}^{(s)}_{ij(r)k\ell}) = \boldsymbol{z}_{ij(r)k}\boldsymbol{D}^1\boldsymbol{z}'_{ij(r)k} + \boldsymbol{z}_{ij(r)k}\boldsymbol{D}^2\boldsymbol{z}'_{ij(r)k} + \sigma^2_s$, $\text{cov}(\widetilde{Y}^{(s)}_{ij(r)k\ell}, \widetilde{Y}^{(s)}_{ij(r)k'\ell}) = \boldsymbol{z}_{ij(r)k}\boldsymbol{D}^1\boldsymbol{z}'_{ij(r)k'} + \boldsymbol{z}_{ij(r)k}\boldsymbol{D}^2\boldsymbol{z}'_{ij(r)k'}$, $\forall i, j(r), k \neq k'$ and $\text{cov}(\widetilde{Y}^{(s)}_{ij(r)k\ell}, \widetilde{Y}^{(s)}_{ij'(r)k\ell}) = \boldsymbol{z}_{ij(r)k}\boldsymbol{D}^1\boldsymbol{z}'_{ij'(r)k} + \boldsymbol{z}_{ij(r)k}\boldsymbol{D}^2\boldsymbol{z}'_{ij'(r)k}$, $\forall i, j(r) \neq j'(r), k$, for within region correlation and $\text{cov}(\widetilde{Y}^{(s)}_{ij(r)k\ell}, \widetilde{Y}^{(s)}_{ij'(r')k\ell}) = \boldsymbol{z}_{ij(r)k}\boldsymbol{D}^1\boldsymbol{z}'_{ij'(r')k}$, $\forall i, j(r) \neq j'(r'), k$, for within subject across region correlation.

We use a common design matrix $\boldsymbol{Z}_{ij(r)}$ for the subject and region random effects but they can be taken to be different in other applications. The degree of smoothness for the natural B-splines is determined through the selection of the number of knots which are typically equispaced. Automatic selection methods for the number of knots include AIC, BIC and cross-validation (Shi et al., 1996; Rice and Wu, 2001); we use AIC in our applications. Note that the mixed effects framework can easily be extended to adapt to the application at hand, for example by modeling further dependency structures or including baseline covariates. Restricted maximum likelihood (REML) is used to estimate model (1) due to its ability to produce unbiased estimates of the variance and covariance parameters. The proposed framework for addressing heteroskedasticity by using separate variance components assigns higher weights to the outcome values with lower variability (i.e. longitudinal features captured from trial windows with more ERP waveforms and therefore more information), and hence is referred to as a weighted model.

4 Simulation

We conduct simulations to study the performance of MAP-ERP, including both the meta-preprocessing and the weighted mixed effects modeling, in reconstructing longitudinal trends over trials in ERP studies. In particular, we include comparisons of our approach with a simple procedure that uses raw single-trial ERP to model longitudinal trends under various signal-to-noise ratios (SNR). Results are presented for four sample sizes, $N = 20, 40, 80$ and 160, which are typical for ERP studies, and are based on 200 Monte Carlo runs. Detailed explanations of the simulation setup along with definitions of mean error (ME) and prediction error (PE) used to evaluate proposed methodology are given in Web Appendix B.

The medians, 2.5th and 97.5th percentiles of the MEs and PEs for the two modeling procedures are given in Table 1 for varying SNRs and sample sizes. As expected, since the

meta-processing step enhances the overall SNR, MAP-ERP leads to consistently smaller ME and PE compared with the single-trial approach across all SNR settings and sample sizes. This is also displayed in Figure 4, which gives the estimated fixed effects means and pointwise confidence intervals from the run with the median ME value for each approach for three SNR settings at $N = 40$. Asymptotic pointwise confidence intervals for the estimated fixed effects means are formed using variance-covariance estimates of the model components. The true fixed effects mean trajectory lies within the 95% pointwise confidence intervals based on MAP-ERP and outside those based on the single-trial approach, further emphasizing the effectiveness of MAP-ERP at capturing longitudinal trends.

The single-trial modeling approach shows a clear improvement in both ME and PE with increasing SNR, but these summary metrics don not change with increasing sample size. For low SNRs, the single-trial approach primarily captures the noise component of the raw ERP, producing highly inflated and inaccurate amplitude estimates. Increases in SNR significantly reduce the magnitude of noise within the raw ERP, effectively shrinking the amplitude estimates produced by the single-trial approach towards the true values as depicted by the relatively sharp decreases in the two performance metrics. However the single trial approach still cannot handle the noise in the ERP trajectories effectively enough where the performance can show improvement within the considered range of sample sizes.

The ME values for MAP-ERP are consistently low regardless of the varying SNR, showing the effectiveness of the proposed meta-preprocessing step in enhancing the signal, especially compared to the single-trial approach. Even though the proposed method is effective for a wide range of SNRs, there are still some subtle issues. Specifically, there are two opposing dynamics, noise and latency jitter, which are still observable in the estimated mean fixed effects trajectories. While noise tends to inflate the amplitude estimates, latency jitter (the fact that the timing of the P3 peak differs across trials and subjects) tends to dampen the

amplitude estimates due to possible misalignment in averaging. As illustrated in Figure 4 (a), SNR=0.4 corresponds to the case with the largest amount of noise. Here the noise effects are dominating latency jitter and the estimated mean is slightly above the true curve especially on the right tail. While noise and latency jitter effects more or less cancel each other at SNR=0.6, where the estimated curve is the closest to the true curve, the effects of latency jitter dominate for SNR=0.8 and the estimated mean curve lies below the true curve. As a result the effect of sample size on ME of MAP-ERP is observable at SNR=0.4 where the noise effects are the dominating dynamics; in this case ME decreases with increasing sample size as expected. However, since increasing sample size does not particularly help with the effects of latency jitter, ME does not significantly improve with increasing sample size for SNR=0.6 or 0.8. Dampening effects of the latency jitter are discussed further in the Conclusion Section.

Similarly, the PE for MAP-ERP improves with increasing SNR due to the reduced levels of noise, but is constant over sample sizes due to the flexible modeling of subject specific effects in the proposed mixed effects model via the inclusion of the spline terms. The proposed modeling allows for reconstruction of complex subject specific functional trends which leads to small PE values across sample size, especially compared to the single curve approach.

5 Analysis of the ERP Data from the Implicit Learning Paradigm

In order to compare the ASD and TD groups in the expected and unexpected conditions over the course of the task, the longitudinal P3 peak amplitudes $\widetilde{Y}^{(s)}_{ij(r)k\ell}$ from 12 electrodes were obtained using our proposed meta-processing procedure for $N = 71$ subjects with up to $K = 120$ trials obtained per condition and were then analyzed via the weighted linear mixed effects model described in Section 3. Five knots were used for the spline fits. This choice was made using a combination of AIC and subject matter expertise, and was found to reflect the complexity of the data well without undersmoothing. Six equal sized groups (i.e. $s = 1, \ldots, 6$)

of $c_{ijk\ell} \in [1, 30]$ were used to model effects of heteroskedasticity. The random effects for the subject and region levels were allowed to have a unique variance for each random effect but were assumed to be independent. The number of averaged curves $c_{ijk\ell}$ was also included as a predictor in the mixed model but was not found to be significant and is hence omitted from the final analysis presented here. For comparison, we model P3 amplitudes using both weighted and unweighted linear mixed effects models, where the unweighted model does not allow for separate variance components with varying $c_{ijk\ell}$. Both models were fit using SAS PROC MIXED-REML.

The two model fits lead to largely similar fixed effects mean trajectories except at the boundaries where the unweighted fit displays more extreme values. This is not surprising since the reduction in $c_{ijk\ell}$ at the boundaries produces noisy cross-sectional averages and inflated amplitude estimates. The weighted model effectively stabilizes amplitude estimates by down-weighting these noisy observations. In addition, even though the difference is small in magnitude, the standard errors for the fixed effects estimates from the weighted model are consistently smaller than those from the unweighted model. In standard regression analysis, conventional standard errors bias up when observations with high leverage (far from predictor mean) are associated with smaller residual variance. Consistent with this observation, most of the spline basis predictor observations with high leverage are associated with smaller residual variance. The variances of the studentized residuals from the weighted model appear to be relatively constant across $c_{ijk\ell}$ compared to those from the unweighted model which have a strong downward trend in variance for increasing $c_{ijk\ell}$. This implies that the weighted model has adjusted for the heteroskedasticity in the data effectively.

Since the main interest of the original study was in comparing ASD to TD subjects, we present results in terms of the differences on condition differences in amplitudes (i.e. (ASD expected - unexpected) - (TD expected - unexpected)). In addition to point-wise confidence

intervals formed based on estimates from the mixed effects model, we also form 90% bootstrap percentile confidence bands based on 200 bootstrap samples drawn from the original subjects' ERP trajectories with replacement. Bootstrap CI's account for the entire two-step procedure as well as variability in the sampling of the ASD and TD subjects. The contrast resulting from averaging 30 ERP in sliding trial windows in the meta-preprocessing step is displayed in Figure 5 (c). During the first 60 trials, the ASD group appears to have a larger amplitude difference than the TD group with a maximum difference between conditions around trial 30 for both groups. Group differences are found to be reliably significant based on the bootstrap bands between trials 20 and 50. Although both ASD and TD subjects appear to be able to differentiate between expected and unexpected conditions, implying implicit learning is taking place, and they do so at a similar speed, the pattern of discrimination seems to differ between the two groups. While the expected minus unexpected condition mean difference is positive for the ASD group, it is negative in TD children (plots not displayed). The absolute condition difference in the amplitudes remains smaller after trial 60 until the end of the experiment. One possible interpretation is that children may be less engaged in the task after this point in the trial. These results are consistent across different window sizes b =20, 30 and 40 for the proposed meta-preprocessing step (Figures 5 (b) through (d)). Hence, inferences based on longitudinal data produced by the meta-preprocessing algorithm appear fairly robust to moderate changes in window size in our application. We also display group differences based on analyzing single-trial data in Figure 5 (a). The bootstrap bands for this approach are very wide due to the low signal-to-noise ratios of the empirical ERP and the amplitude differences between groups are no longer significant. In contrast MAP-ERP is able to identify the regions of significant group differences due to the increase in the signal-to-noise ratio.

We highlight that previously published results on the ERP data from this implicit learning paradigm completely ignored the longitudinal component of the data due to averaging over

trials (Jeste et al., 2014). In addition, the single-trial approach is too noisy to depict any group differences across trials. In contrast, MAP-ERP approach provides novel insights on a new longitudinal dimension that is typically lost in analysis of ERP data, leading to interpretable group differences over trials with respect to patterns of implicit learning.

6 Discussion

We have proposed a meta-preprocessing procedure for ERP studies which enhances the signal-to-noise ratio while still retaining longitudinal trends across trials. Longitudinal features may be an important focus in various ERP studies, such as the implicit learning paradigm analyzed in this manuscript where speed of acquisition contributes significantly to the characterization of implicit learning in young children with ASD compared to typically developing controls. While the proposal focuses on repeated ERP signal recorded in response to multiple applications of a stimuli, the proposed methodology is applicable more broadly to studies involving repetitions of a systematic signal observed with noise, such as heart beat, breath cycle or eye blinks. Following the meta-preprocessing step, we also proposed a weighted linear mixed effects model that has been shown to describe longitudinal trends in ERP features effectively in simulation studies.

The proposed method is a two-step approach comprising a meta-preprocessing step followed by a weighted linear mixed mixed effects model. One reason for using a two-step approach was to connect to the rich literature on analysis of ERP data which focuses on modeling particular features of ERP curves (e.g. amplitudes, latency) which are readily interpretable scientifically. Our meta-preprocessing step allows focusing on these readily interpretable features, while preserving longitudinal trends over trials in contrast to the common practice in this literature. Another advantage of the two-step approach is that the initial novel meta-preprocessing step is left modular to be flexibly coupled with an array of secondary analyses options for the

extracted features. One limitation of the current approach based on moving averages is that misalignment of the ERP features can potentially lead to slight underestimation of the true peak amplitudes. This is a challenging issue, since alignment would de facto require that features be identified before averaging which is impractical due to noise levels in the individual raw ERP. Even though we were able to draw combined inference for the two steps of MAP-ERP via bootstrap confidence intervals in our application, we identify the development of formal inference procedures as a direction for further research.

While the current manuscript focuses on analyzing particular features of ERP such as peak amplitudes or latencies, which is a common practice in the EEG literature, we note that recent studies have proposed using functional data analysis for analysis of the ERP curves in their entirety (Bugli and Lambert, 2006; Davidson, 2009). The proposed meta-preprocessing algorithm creates longitudinal functional data, i.e. repeated ERP waveforms over the trials. Analysis of such data using functional techniques is a very interesting open problem.

7 Supplemental Materials

The provided R code and Web Appendices (Sections 2 and 4) are available with this paper at the Biometrics website on Wiley Online Library.

Acknowledgements

We thank two referees, the associate editor and the editor for their valuable comments. This publication was made possible by the graduate research fellowship grant DGE-1144087 (KH) from the National Science Foundation.

References

Bruin, N. M. W. J., Ellenbroek, B. A., van Schaijk, W. J., Cools, A., Coenen, A. M. L. and van Luijtelaar, E. L. J. M. (2001). Sensory gating of auditory evoked potentials in rats:

effects of repetitive stimulation and the interstimulus interval. *Biological Psychology* **55(3)**, 195–213.

Bugli, C. and Lambert, P. (2006). Functional ANOVA with random functional effects: an application to event-related potentials modelling for electroencephalograms analysis. *Statistics in Medicine* **25(21)**, 3718–3739.

Coppola, R., Tabor, R. and Buchsbaum, M. S. (1978). Signal to noise ratio and response variability measurements in single trial evoked potentials. *Electroencephalography and Clinical Neurophysiology* **44(2)**, 214–222.

Davidson, D. J. (2009). Functional mixed-effects models for electrophysiological responses. *Neurophysiology* **41(1)**, 79–87.

Delorme, A. and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including component analysis. *Journal of Neuroscience Methods* **134(1)**, 9–21.

De Silva, A. C., Sinclair, N. C. and Liley, D. T. J. (2012). Limitations in the rapid extraction of evoked potentials using parametric modeling. *Transactions on Biomedical Engineering* **59(5)**, 1462-1471.

Gasser, T., Mocks, J. and Verleger, R. (1983). SELAVCO: A method to deal with trial-to-trial variability of evoked potentials. *Electroencephalography and Clinical Neurophysiology* **55(6)**, 717–723.

Gasser, T. and Molinari, L. (1996). The analysis of the EEG. *Statistical Methods in Medical Research* **5(1)**, 67–99.

Jeste, S. S., Kirkham, N., Sentürk, D., Hasenstab, K., Sugar, C., Kupelian, C., Baker, E., Sanders, A., Shimizu, C., Norona, A., McEvoy, K., Paparella, T., Freeman, S. F. N. and

Johnson, S. P. (2014). Neural correlates of visual statistical learning in young children with autism spectrum disorder (ASD). *Developmental Science*, In Press.

Krieger, D., Berger, T. W. and Sclabassi, R. J. (1992). Instantaneous Characterization of Time-Varying Nonlinear Systems. *Transactions on Biomedical Engineering* **39(4)**, 420–424.

Krieger, D. and Sclabassi, R. J. (1994). Time-varying evoked potentials. *Journal of Medical Engineering and Technology* **18(3)**, 96–100.

Luck, S. J. (2005). *An Introduction to the Event-Related Potential Technique*. Cambridge: The MIT Press.

Mocks, J., Gasser, T. and Tuan, P. D. (1984). Variability of single visual evoked potentials evaluated by two new statistical tests. *Electroencephalography and Clinical Neurophysiology* **57(6)**, 571–580.

Mocks, J., Gasser, T., Pham, D. T. and Kohler, W. (1987). Trial-to-trial variability of single potentials: methodological concepts and results. *International Journal of Neuroscience* **33(1-2)**, 25–32.

Quiroga, R. Q. and Garcia, H. (2003). Single trial event-related potentials with wavelet denoising. *Clinical Neurophysiology* **114(2)**, 376–390.

Rice, J. A. and Wu, C. O. (2001). Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics* **57(1)**, 253–259.

Rossi, L., Bianchi, A. M., Merzagora, A., Gaggiani, A., Cerutti, S. and Bracchi, F. (2007). Single trial somatosensory evoked potential extraction with ARX filtering for a combined spinal cord intraoperative neuromonitoring technique. *Biomedical Engineering Online* **6(2)**, 1–12.

Shi, M., Weiss, R. E. and Taylor, J. M. G. (1996). An Analysis of Paediatric CD4 Counts for Acquired Immune Deficiency Syndrome using Flexible Random Curves. *Applied Statistics* **45(2)**, 151–163.

Stahl, D., Parise, E., Hoehl, S. and Striano T. (2010). Eye contact and emotional face processing in 6-month-old infants: Advanced statistical methods applied to event-related potentials. *Brain and Development* **32(4)**, 305–317.

Tierney, A. L., Durnam, L. G., Farley, V. V., Flusberg, H. T. and Nelson, C. A. (2012). Developmental Trajectories of Resting EEG Power: An endophenotype of autism spectrum disorder. *PLoS One* **7(6)**, e39127.

Travers, B. G., Klinger, M. R., Mussey, J. L. and Klinger, L. G. (2010). Motor-linked implicit learning in persons with autism spectrum disorders. *Autism Research* **3(2)**, 68–77.

Turetsky, B., Raz, J. and Fein, G. (1989). Estimation of trial-to-trial variation in evoked potential signals by smoothing across trials. *Psychophysiology* **26(6)**, 700–712.

Verleger, R., Gasser, T. and Mocks, J. (1985). Short term changes of event related potentials during concept learning. *Biological Psychology* **20(1)**, 1–16.

Woestenburg, J. C., Verbaten, M. N. and Slangen, J. L. (1983). Stimulus information and habituation of the visual event related potential and the skin conductance reaction under task-relevance conditions. *Biological Psychology* **16(3-4)**, 225–240.

Woestenburg, J. C., Verbaten, M. N., van Hees, H. H. and Slangen, J. L. (1983). Single trial ERP estimation in the frequency domain using orthogonal polynomial trend analysis (OPTA): estimation of individual habituation. *Biological Psychology* **17(2-3)**, 173–191.

Yeung, N., Bogacz, R., Holroyd, C. B., Nieuwenhuis, S. and Cohen, J. D. (2007). Theta phase-resetting and the error-related negativity. *Psychophysiology* **44(1)**, 39–49.

Table 1: Median and (2.5th, 97.5th) percentiles of simulation performance metrics (ME and PE) for varying SNRs from 200 Monte Carlo runs with sample sizes $N = 20, 40, 80$ and 160.

		ME		PE	
N	SNR	Single	MAP-ERP	Single	MAP-ERP
20	0.4	.531 (.493, .563)	.040 (.020, .067)	.526 (.476, .572)	.164 (.141, .200)
	0.6	.225 (.190, .258)	.032 (.014, .069)	.222 (.194, .253)	.116 (.100, .137)
	0.8	.108 (.073, .148)	.039 (.013, .076)	.111 (.097, .129)	.094 (.082, .108)
40	0.4	.530 (.503, .558)	.033 (.019, .053)	.528 (.498, .559)	.164 (.144, .187)
	0.6	.224 (.198, .248)	.030 (.014, .058)	.221 (.205, .243)	.115 (.104, .129)
	0.8	.109 (.081, .135)	.037 (.013, .064)	.112 (.102, .122)	.093 (.085, .102)
80	0.4	.530 (.501, .553)	.029 (.019, .043)	.529 (.507, .551)	.164 (.150, .180)
	0.6	.222 (.203, .242)	.030 (.016, .049)	.223 (.210, .236)	.115 (.108, .125)
	0.8	.107 (.087, .127)	.037 (.021, .058)	.112 (.104, .119)	.093 (.087, .100)
160	0.4	.531 (.503, .548)	.026 (.019, .034)	.528 (.515, .545)	.163 (.152, .175)
	0.6	.224 (.203, .236)	.029 (.017, .045)	.222 (.214, .232)	.115 (.109, .122)
	0.8	.109 (.090, .122)	.037 (.025, .054)	.111 (.107, .117)	.093 (.089, .098)

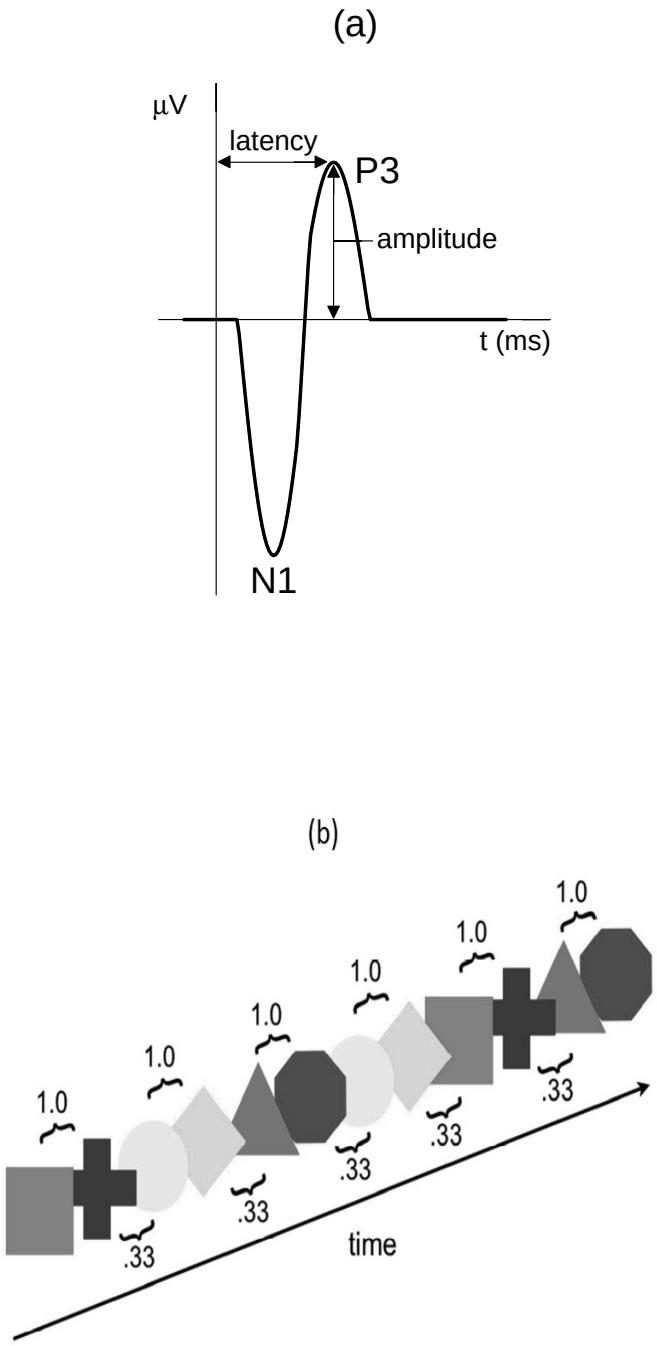

Figure 1: (a) A typical ERP waveform containing the P3 and N1 phasic components from the implicit learning paradigm. (b) Visualization of the implicit learning paradigm. The continuous stream of six colored shapes (pink squares, blue crosses, yellow circles, turquoise diamonds, gray triangles and red octagons) are organized into three familiar pairs. The 'expected' condition is defined as the transition from shape 1 to shape 2 in the familiar shape pair (with probability 1) and the 'unexpected' condition is defined as the transition from shape 2 to shape 1 of any shape pair (with probability 0.33).

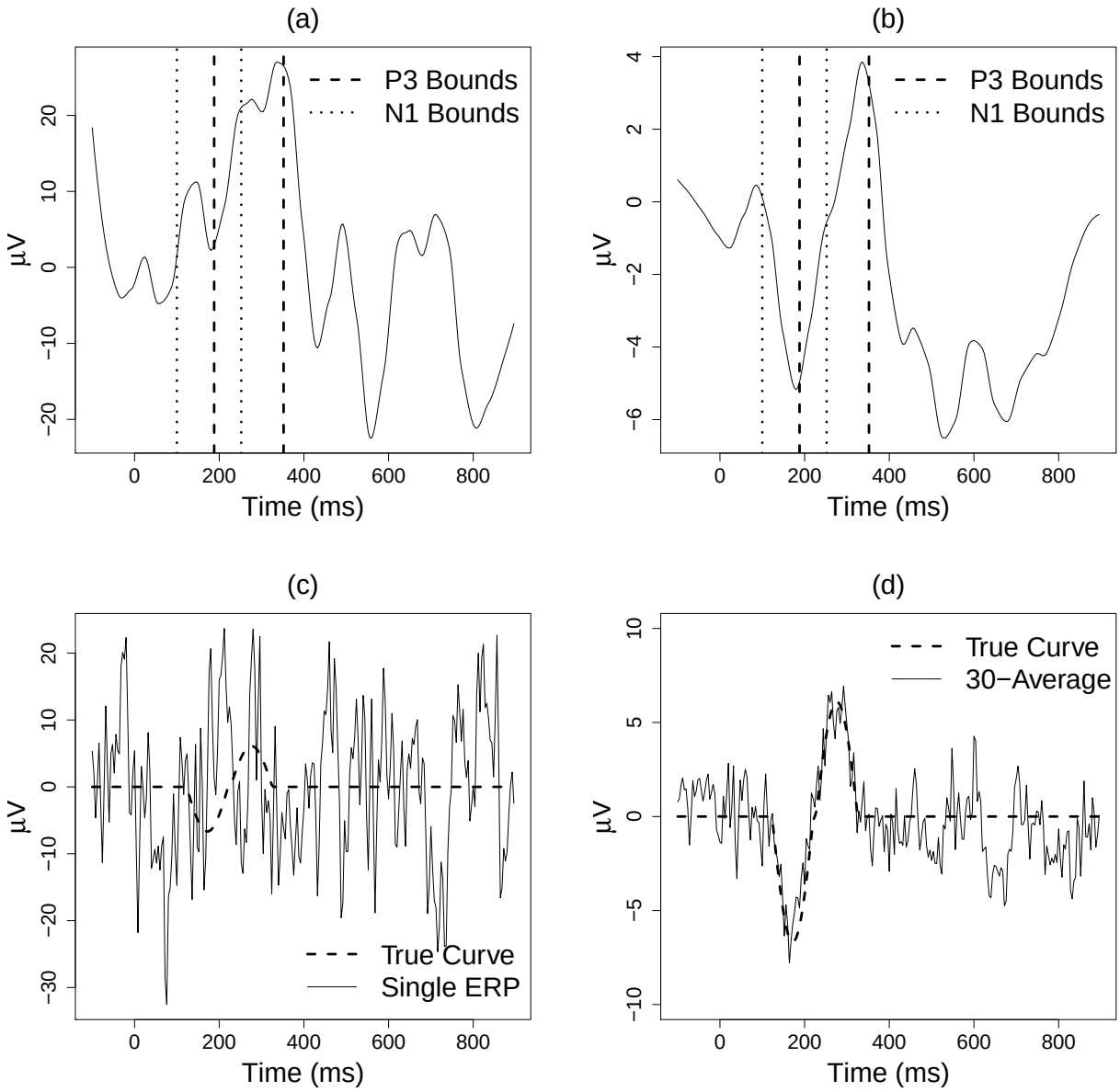

Figure 2: (a) ERP waveform for a single subject, trial, electrode and condition in the frontal region after preprocessing, (b) average of the first 30 consecutive ERP waveforms for a single subject, electrode and condition, (c) simulated ERP waveform for a single subject, trial and electrode, (d) average of 30 simulated ERP waveforms for a single subject and electrode. Vertical boundaries denoted by the dotted and dashed lines in (a) and (b) correspond to the search of the locations of the N1 and P3 components and are [100ms, 250ms] and [190ms, 350ms], respectively (Jeste et al., 2014).

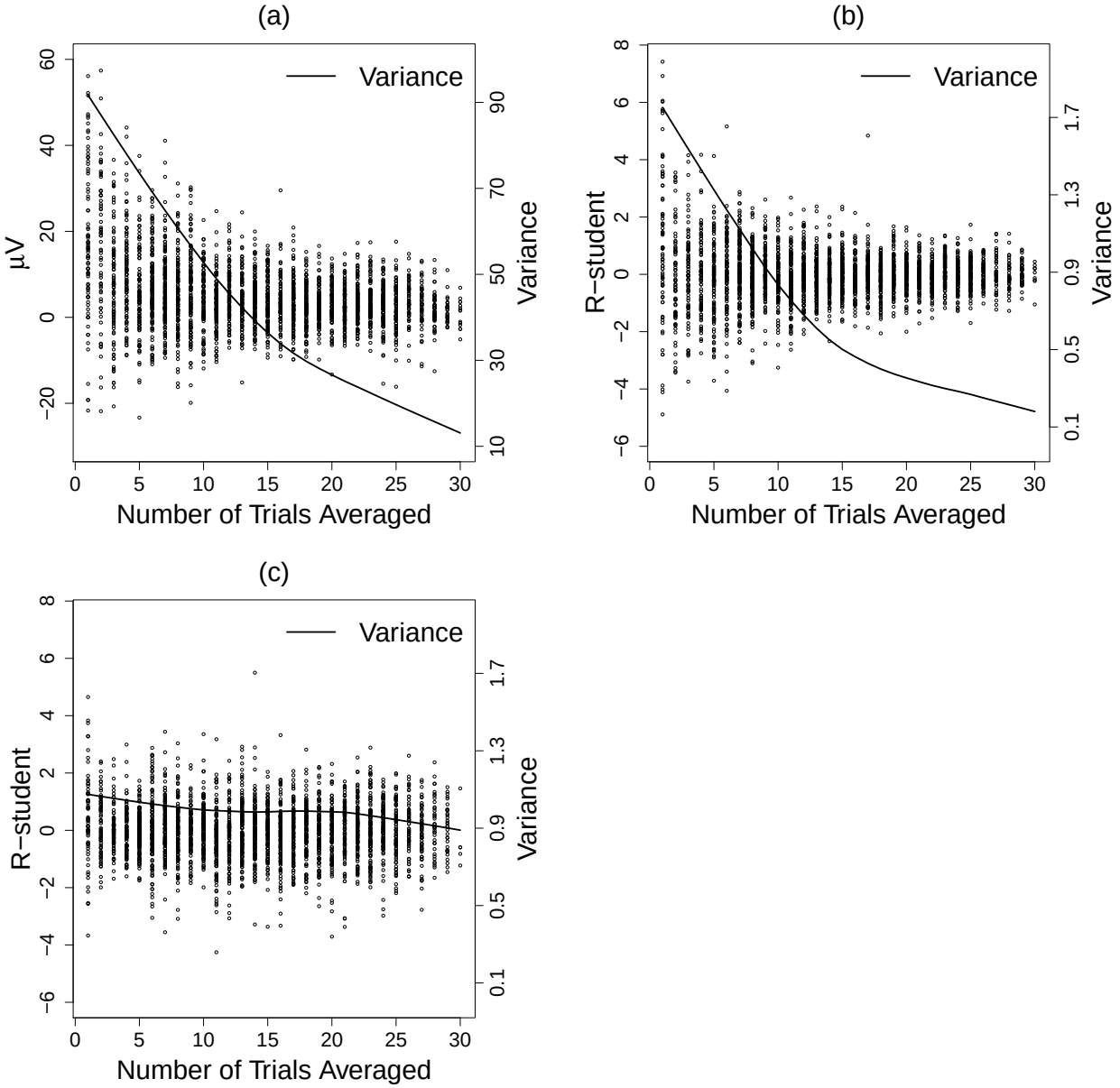

Figure 3: (a) P3 amplitudes as a function of number of trials averaged, $c_{ijk\ell}$, in the analysis of the ERP data from the implicit learning paradigm, (b) studentized residuals obtained from the mixed effects model without weights as a function of $c_{ijk\ell}$, (c) studentized residuals from the proposed weighted mixed effects model as a function of $c_{ijk\ell}$. Variance of the P3 amplitudes and the residuals are shown (solid line) corresponding to the y-axes given on the right hand side.

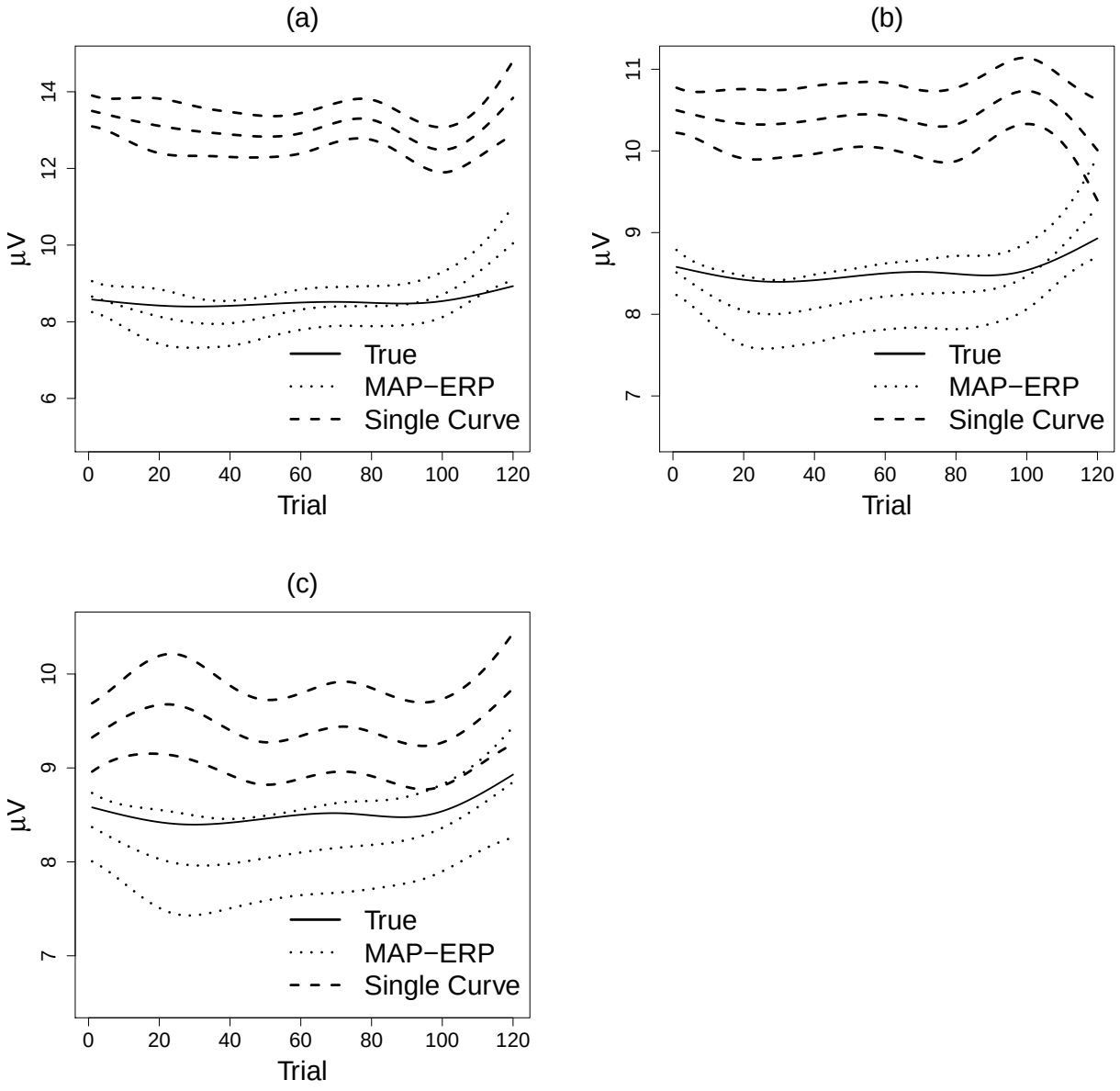

Figure 4: Estimated fixed effects mean trajectories along with 95% confidence intervals corresponding to the median of the ME in simulations based on the single curve (trial) approach and MAP-ERP for (a) SNR=0.4, (b) 0.6 and (c) 0.8.

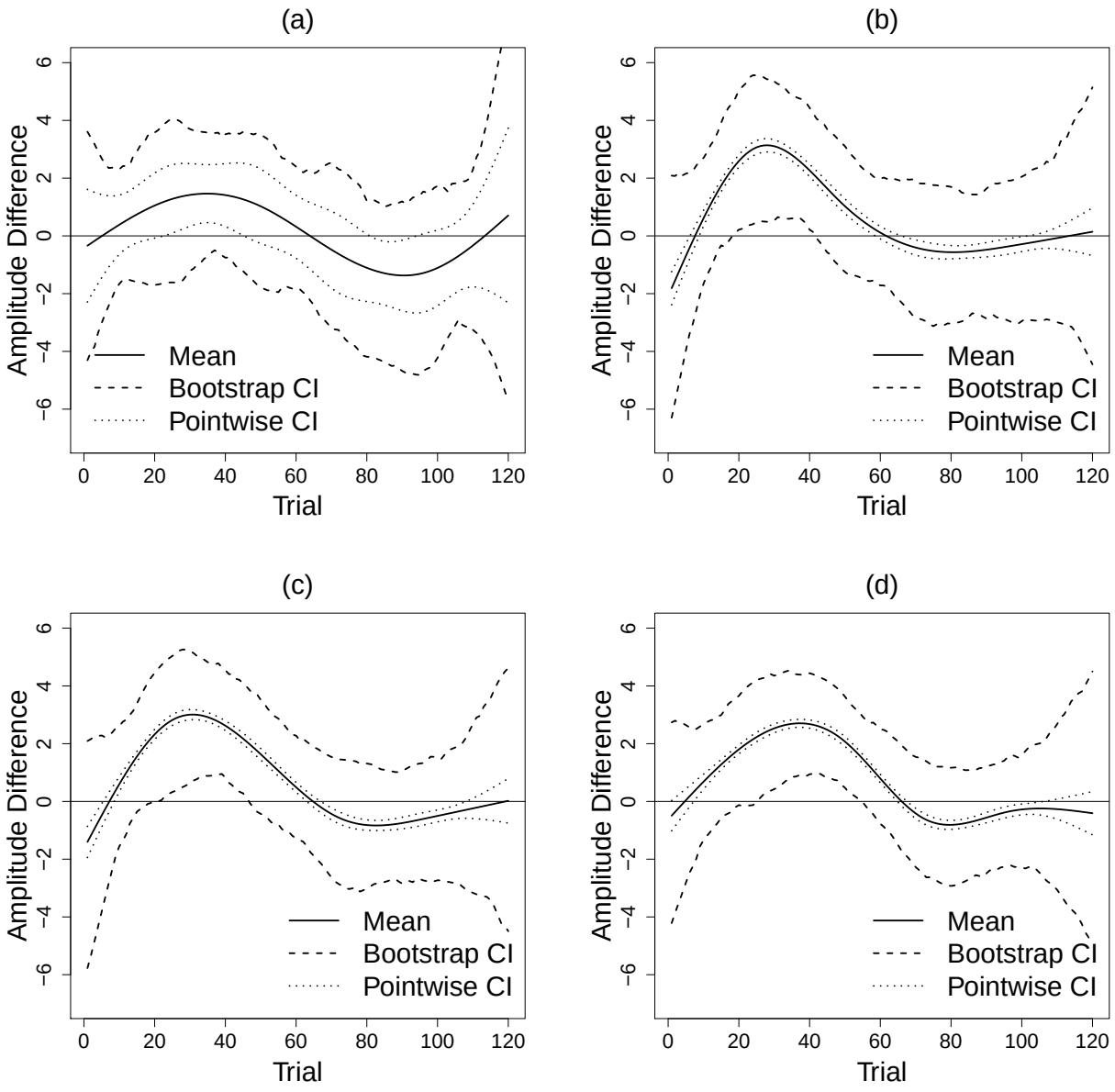

Figure 5: Estimated mean group and condition difference ((ASD expected - unexpected) - (TD expected - unexpected)) trajectories based on the (a) single-trial approach and MAP-ERP with window sizes (b) b=20, (c) b=30 and (d) b=40. 90% bootstrap bands from 200 runs are also provided (dashed lines).