The Effects of Autocorrelation on Overlap Corrected rERPs

Felix Schröder

(s3086402)

31 August 2021

Major Thesis MSc Behavioural and Cognitive Neuroscience

Faculty of Science and Engineering

University of Groningen

1st Supervisor: Jun.-Prof Dr. Benedikt Ehinger

2nd Supervisor: Assistant Prof. Dr. Sebastiaan Mathôt

Daily Supervisor: René Skukies, MPhil

Abstract

Regression-based deconvolution allows for the separation of overlapping event related potentials. This carries great potential for naturalistic study designs but has not been validated for within subject analyses so far. As opposed to across subject analyses, tests on within subject estimates may not be valid since t-tests on individual β estimates rely on the assumption of independence to allow inference. EEG data violates this assumption. In this report we assessed the effects of this non-independence on type-1 error rates of β parameters estimated from EEG via continuous-time regression. To this end, we simulated null-signals that had different degrees of non-independence. Subsequently, a number of different experimental conditions were simulated, modelled, and their parameters tested. We found non-independence of the residuals increased type 1 errors dramatically once systematic overlap was added to the model. We further discuss possible solutions that could be tested. Since EEG data violate the assumption of independence to a much greater extent than traditionally dealt with in the time-series literature, none of these solutions should be trusted without thorough empirical testing beforehand. Thus, we discourage the application of overlap-correction for within-subject analyses for now.

Table of Contents

The Effects of Autocorrelation on Overlap Corrected rERPs

Applying regression to Electroencephalography (EEG) carries great potential. A myriad of different implementations have seen a recent spike in interest (Cornelissen et al., 2019; Dandekar et al., 2012; Gert et al., 2021; Kristensen et al., 2017). Obtaining Event Related Potentials (ERPs) via regression (rERPs) offers a convenient way to control for effects of confounding covariates and model non-linear effects with the help of splines (Smith & Kutas, 2015b). Most importantly for this report, it enables the deconvolution of EEG responses that may overlap due to the experimental design, like button presses, or uncontrollable factors like eye movements (Dimigen & Ehinger, 2021; Ehinger & Dimigen, 2019). This way regression-based deconvolution enables more naturalistic and dynamic research designs by removing previous constraints that came with traditionally averaged ERPs.

Within subject analyses face an issue with this approach that has not been explored so far in the literature around continuous-time regression of EEG (Ehinger & Dimigen, 2019; Smith, 2011; Smith & Kutas, 2015a). Inherent to regression techniques, like ordinary or generalised least squares, is the assumption that the residuals of the modelled data need to be independent from each other to assure valid inference (often shortened to the first “I” in i.i.d. and described in the Gauss-Markov theorem). Just as for fMRI and many other time series data (Bartlett, 1946; Hyndman & Athanasopoulos, 2018; van Rij et al., 2019; Woolrich et al., 2001). EEG signals will usually violate this assumption (Bae & Luck, 2019; Doyle & Evans, 2018; Linkenkaer-Hansen et al., 2001).

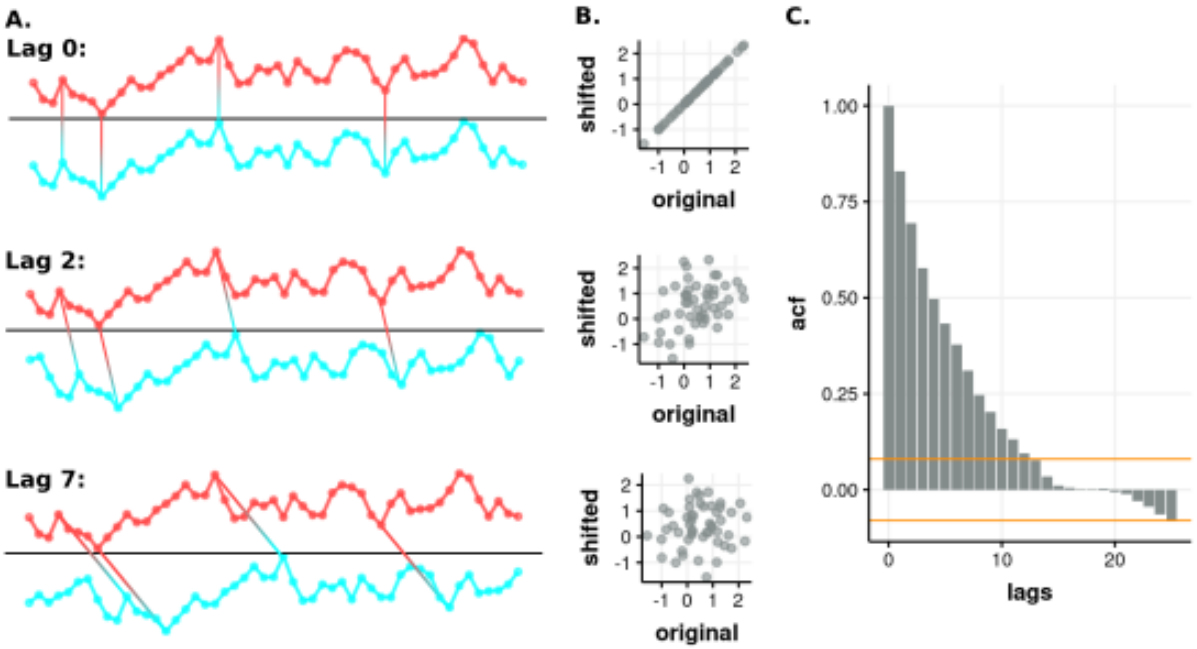
In what follows, we aimed at exploring the effect of serially correlated data / autocorrelation on the validity of regression-based deconvolution. To this end, we made use of artificial as well as resting-state EEG signals. With this, we hope to inform future studies that aim to to analyse overlap corrected rERPs within subjects. Such applications could be especially relevant to small clinical samples that would face power issues in a group analysis.

Autocorrelation

Whereas Person's r describes the magnitude and direction of similarity of two separate variables, autocorrelation describes the similarity of one variable with itself (usually through time). Any repeatedly measured variable of length n, has n – 1 autocorrelation coefficients. For example, a 100 points long time-series could have 99 autocorrelation coefficients. Each of these coefficients denotes a so called lag. At lag 0 we take the correlation of the time-series with itself. This will always yield a correlation of one. If we shift a copy of the time-series one datapoint into the past and line it up with the original time series their correlation will now denote the autocorrelation at lag 1 (see figure 1). By shifting the copied time-series n data points into the past we can calculate the autocorrelation at lag n. When any of the lags, apart from lag 0, is significantly different from zero, we speak of an autocorrelated time-series. Some may refer to such variables also as serially or temporally correlated.

Figure 1

Assessing Autocorrelation



Note. (a) A time series in red with itself shifted by n sampling points (lags) in blue underneath. (b) Scatter plots of the individual datapoints for each of the three lags. (c) The autocorrelation function (acf), meaning the autocorrelation as a function of lag. The orange horizontal line denotes the 95% confidence interval to support judging how far the signal significantly predicts itself (Hyndman & Athanasopoulo, 2018).

As alluded to earlier, such autocorrelational processes affect our analyses of time-series data and are ubiquitous across domains. For example, economists and ecologists use the phenomena to derive seasonalities and other trends in their data (Hyndman & Athanasopoulos, 2018). However, for neuroscientists, it has been more of a nuisance. While recent neuroimaging studies have used autocorrelation to derive connectivity statistics of different brain regions (Arbabshirani et al., 2019; Shinn et al., 2021), traditionally it has been regarded a violation of statistical assumptions that complicated statistical analysis. In the worst case, it goes unnoticed or uncorrected resulting in inflated the false-positive rates (Baayen et al., 2017; Burock & Dale, 2000; Lund et al., 2006; Purdon & Weisskoff, 1998; Woolrich et al., 2001). Even classical response variables from psychology like reaction times aren't free of the interdependencies that are described by autocorrelation (Baayen et al., 2017; Thul et al., 2021).

Effects of Violating the Assumption of Independence

In an ideal scenario all our inference gleaned from our data is valid. For a statistical test to be valid its false positive rate needs to be under the nominal specificity (often $\alpha = 0.05$). Certain assumptions may be violated without the test becoming invalid (we usually refer to these tests as robust against certain violations). It is often the combination of violations that invalidate statistical inference (Friston et al., 2000). However, violating the independence assumption has been shown to impact the false-positive rate greatly across domains (Astrophysics: Carter & Winn, 2009; Ecology: Kissling & Carl, 2008; Neuroscience: Woolrich et al., 2001) .

In the case of non-independence (measured via autocorrelation), the usual rules for calculating degrees of freedom no longer apply and overestimate the true, or effective, degrees of freedom (Afyouni et al., 2019; Friston et al., 2000; Worsley & Friston, 1995). In other words, the effective sample size of independent datapoints used to compute an estimate becomes smaller than the actual dataset length under autocorrelation. Thus, adapting the hypotheses tests by estimating the effective degrees of freedom could already be the solution to this problem. However, previous studies on the topic in fMRI have noted that this quickly becomes an impractical problem to solve

due to the complex covariance structures involved (Afyouni et al., 2019; Bollmann et al., 2018). Furthermore higher sampling rates seem to exacerbate the complexity (Bollmann et al., 2018). We will return to this issue towards the end of the paper. Another solution proposed is to whiten your data by removing the autoregressive components from your data with a two-step procedure. While this approach is still widely used in the fMRI community (Olszowy et al., 2019) it has also seen its fair share of criticism across domains for reducing size and power properties of hypotheses tests (Cliff et al., 2021; Sul et al., 2005).

Before any such considerations can be made for regression-based deconvolution of EEG data, we first need to assess its robustness against the independence violation.

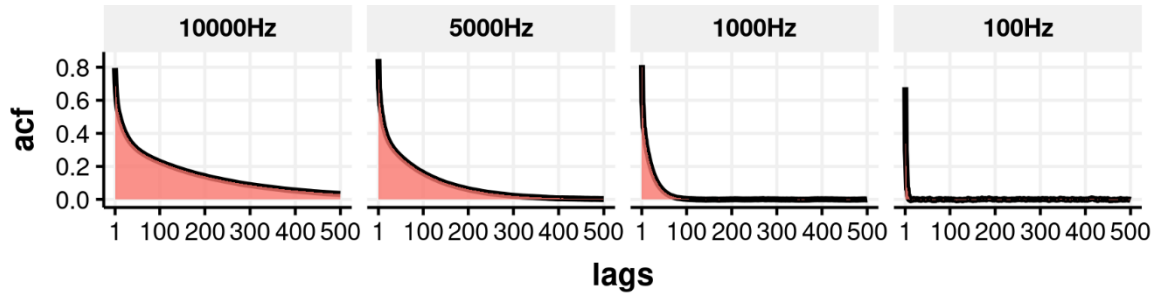
Autocorrelation in EEG

In the spectral domain EEG data will usually follow a 1/f structure (Cohen, 2014; Zarahn et al., 1997), in which power of frequency components declines the higher the given frequency. Such high power / low frequency components are the strongest source of autocorrelation (Fadili & Bullmore, 2002; Smith, 2011). While these 1/f structures are present all throughout nature (Halley, 1996), we assumed the strength of autocorrelation to be especially strong far into high lags in EEG data because of the common usage of high sampling rates between 250 and 1000 Hz (see figure 2 a). This leads to highly correlated residuals across long spans in the data. In comparison, the autocorrelation of the hemodynamic response function observed in the functional magnetic resonance literature, while present, is probably less intense (Bollmann et al., 2018) given the typically slower sampling rates of 3 to 1 Hz (see figure 2). Whether common pre-processing steps of EEG like high-pass filtering with cut offs at 0.1 Hz / 0.01 Hz or down sampling may already attenuate false positive rates to acceptable levels will be assessed here.

Apart from low-frequency induced autocorrelation, we could also imagine a scenario where a dominant oscillatory frequency could induce autocorrelation only for certain conditions (theta) or, in blocked designs, only towards the end of the experiment due to fatigue (alpha). Take, for example, a n-back task where theta band power increases with working memory load. In the case

Figure 2

Downsampling 1/f Noise



Note. (a) ACFs of 1/f noise sampled at decreasingly lower sampling rates.

that autocorrelation increases false-positive rates we might see effects in the high memory conditions that are absent / nonsignificant in the low memory conditions. Whilst such a finding could easily lead to a myriad of post-hoc explanations, it could possibly also be attributed to ill-defined degrees of freedom in the autocorrelation laden condition.

The Current Study

This study empirically tested the validity of standard parametric significance tests of β parameters of overlap corrected rERPs and is divided into three parts. First, we generated artificial signals and compared their relevant statistics to those of real resting-state EEG data. From this, a suitable stand-in for further simulations was chosen. This was done because of EEG's hard to control for complexities overcomplicating the interpretations of our simulations. Next, we generated three different types of event related signals using their onsets as parameters for the regression model. One void of overlap, one in three different, but heavily controlled, intensities of overlap, and one type with purely randomly placed onsets. In a last step, we modelled the artificial null data with the Unfold toolbox (Ehinger & Dimigen, 2019) and measured the false positive rates. By varying the model parameters and filter/resampling settings this yielded insights into the robustness of the regression-based deconvolution approach under autocorrelation.

Methods

To begin with, we needed to decide how we would approximate EEG time-series. We focussed on comparable power spectral density (PSD) functions, the most important aspect being

the general 1/f shape PSD functions take with EEG data (Doyle & Evans, 2018). This was important since high power low frequency oscillations should be the strongest cause for long lasting autocorrelation. We needed to find such a stand-in since real data would most likely have too many complexities / idiosyncrasies depending on the specific data set used. These would over complicate any conclusions drawn from our simulations. To find this stand-in, we decided to compare not only different kinds of artificial noise but also different manners of generating them.

Generating Noise

In the literature, 1/f noise is usually described as pink noise, one of many types of coloured noise that all differ in their PSD slopes. Pink noise has been referred to as the best null model of neural activation (Doyle & Evans, 2018). It turned out that generating pink noise was not an easy endeavour and different implementations all differed in their low frequency component power. We ultimately decided for the implementation of the SignalAnalisys.jl Julia package (version 0.4.0), which filtered white noise with a manually defined digital filter (see appendix A).

White noise was generated by repeatedly sampling from a normal distribution (N(0,1)). This generated white noise's characteristic flat PSD function with equal power throughout all present frequency components.

To get insights into how 1/f noise would differ from a simple autoregressive sequence, we also included an AR(1) signal in some our analyses. AR(1) noise was generated as seen in equation 1. Where p = 1 , ϕ = 0.8, and ω ~ N(0,1).

$$\text{Equation 1: } X_t = \sum_{j=1}^{p} \phi_j X_{t-j} + \omega_t$$

Our last tools to manipulate the intensity of autocorrelation were high-pass filtering and resampling the signal to lower sampling rates. For our filters we made use of MNE's default FIR filter settings since they ought to be commonly used by the field. They make use of a hamming window and have a transition bandwidth of at least 2Hz or 25% of the lower passband edge (Gramfort, 2013; version: 0.23). When downsampling, MNE performs appropriate low pass filter to avoid aliasing.

The resting state EEG data used to compare the artificial signals to were previously recorded as part of an unpublished master's thesis (Skukies, 2020). They used 62 passive electrodes with multi-channel amplifiers (BrainAmp, BrainProducts GmBH, Gilching Germany). The electrodes were laid out according to the international 10-10 system (EasyCap, BrainProducts GmBH, Gilching Germany), with their reference and ground electrodes having been placed just above the nasion. Additionally, ocular movements were recorded with two electrodes placed at the outer canti. The EEG was recorded at 5000hz with an online low-pass filter set to 1000hz with the BrainVision recorder software (BrainProducts GmBH, Gilching Germany).

Generating Overlapping Signals

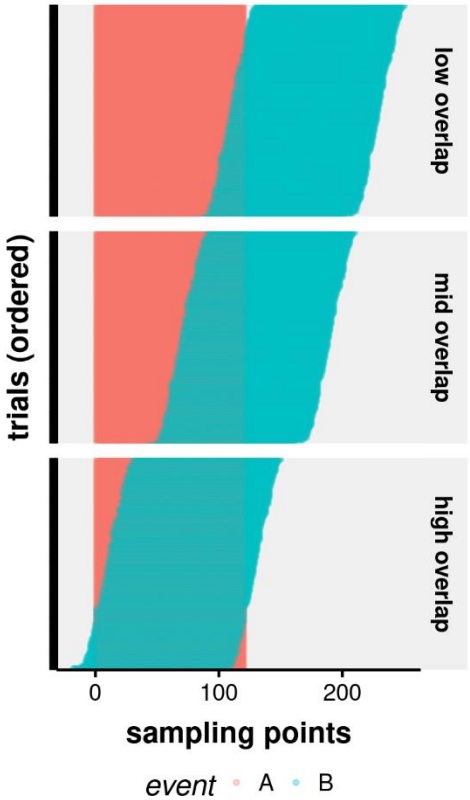
There was no event related activity in our modelled signals. Since we were interested in false positives, we merely generated differing event onsets to test how robust regression-based deconvolution would perform the face of autocorrelation under different experimental set ups.

First, we wanted to see if autocorrelation alone would bias the false-positive rate. We did so by creating an intercept only model of a single event type. Its onsets were evenly spaced throughout the signal with sufficient distance (2.5 seconds) between the onsets to avoid any overlap. This will be referred to as the "no overlap" condition.

Next, we added a second event type onto this model, placed in one of three different distances to the first event. This resulted in a low-,

Figure 3

Visualizing the Three Different Overlap Conditions



Note. In red this figure shows the duration and position of the 'A' event in a trial. The 'B' event is shown in green. Each row plots all modelled trials of this example seed, ordered by overlap intensity.

mid-, and high-overlap condition respectively (see figure 3). In the low overlap condition, 9.3 % (~11 points on average) of the second event's sampling points overlapped with those of the first. For the mid condition this proportion increased to 42% (~50 points on average) and in the high overlap condition to 90.4% (~108 points on average). These conditions should roughly mimic classical EEG study designs in which, for example, a button press would follow a certain event in a predictable manner.

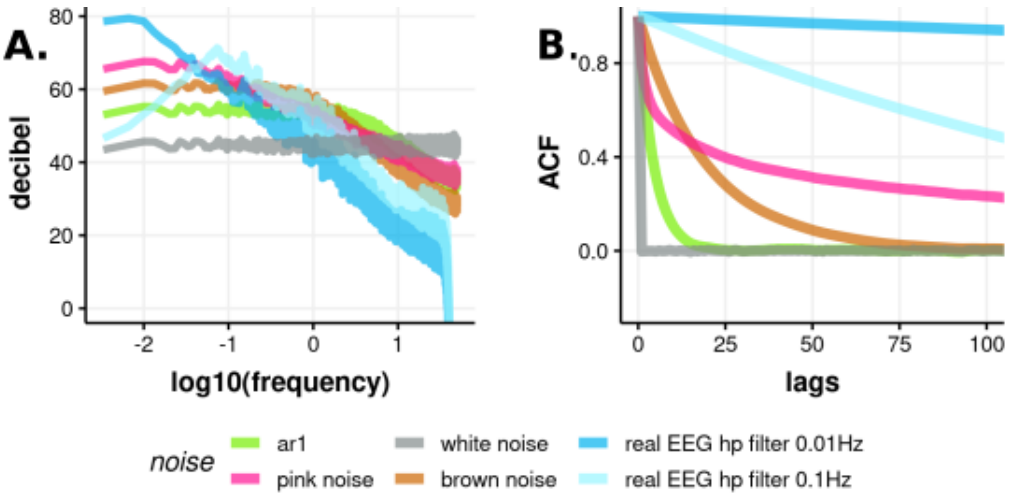
Lastly, since previous work on autocorrelation induced biases championed randomness as the easiest solution (Thul et al., 2021), we also included a condition that contained no controlled relation between the two events. They were both placed randomly throughout the signal causing variable overlap intensities depending on the seed. Random or at least variable overlap would be best compared with unpredictable event onsets in free viewing and other naturalistic experimental designs. All conditions had a total of 95 events per event type. Both events were modelled with a set of FIR basisfunctions spanning 1.2 seconds from 400 ms before to 800ms after onset. Defined like this, such a basis set is also used in the Unfold toolbox's tutorial, likely making them the default for many researchers starting out with this method.

Testing Validity

In order to compute the false positive rates, we kept all factors but one fixed in our simulations (made easier with DrWatson.jl; Datseris et al., 2020). This way we could test claims about the impact of autocorrelation and model parameters separately. At 200 repetitions per model condition this means that each data point in the results presented below is an average from 200 separately fitted models that only differed in the seed used for the onset jitter. When testing the effects of autocorrelation, all signals used were modelled with the exact same model specifications and parameters. By inspecting the change in false-positive rates due to model parameters and the change to this change over autocorrelation conditions, we can make statements about when regression-based deconvolution stops yielding valid statistical inference at a first stage analysis.

Figure 4

Compare EEG to Artificial Signals



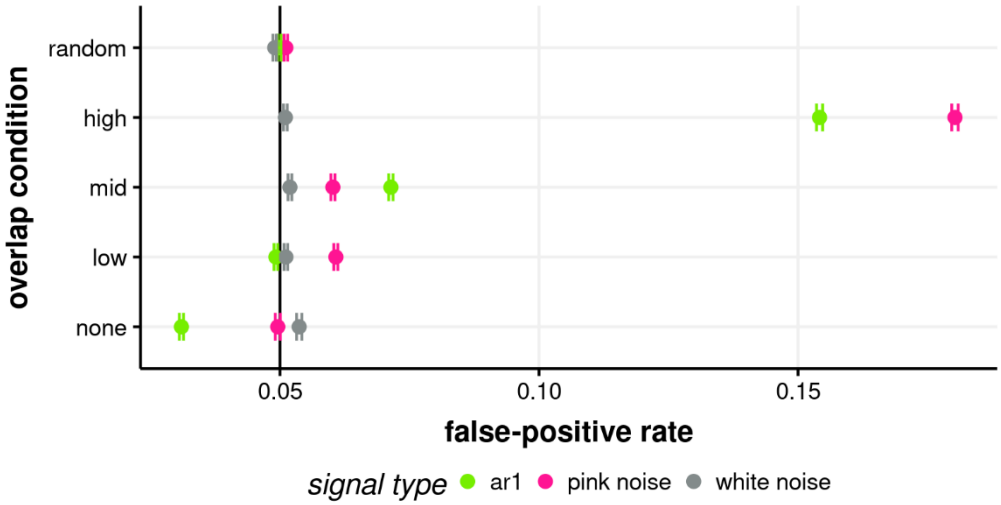
Note. (A) Power spectral density function computed via the Welch method of 4 different artificial noise types and 2 EEG signals that were filtered at either 0.01Hz or 0.1 Hz with a high-pass filter. (B) Autocorrelation functions of the same signals.

Results

To decide on a suitable stand-in for real EEG data we first compared the ACF and PSD of our artificial signals with those of an exemplar EEG subject (see Figure 4). Pink noise, while not a perfect match, approximated both of these measures the best.

Figure 5

False-Positives by Overlap



Note. The y-axis shows the different overlap conditions we modelled. Each plotted point represents the average false -positive rate over 400 models with the error bars denoting their standard errors. The black vertical line represents the nominal alpha value, values beyond it indicate an invalid test.

Next, we assessed to what extent hypotheses tests carried out on single estimates were robust to autocorrelation. We found that uncorrelated white noise showed no signs of increased false positive rates under any of the tested conditions (see figure 5). Both correlated noise types, the AR1 and pink noise (i.e., the EEG null model) did show increases in the false positive rate in the mid and high overlap condition. Furthermore, pink noise also displayed increased false positives at the low overlap condition already. All noise types, however, showed no signs of increased false positives in the random event onset condition, even though this condition also included overlapping events.

As a last step, we wanted to see how continuous changes to autocorrelation and overlap intensity change these results (see figure 6). The pink noise condition showed continually increasing false positives with higher autocorrelation (i.e., lower high-pass filter cut-off values). This effect became more apparent with higher degrees of systematic overlap. The same trend with somewhat higher false positive rates could be observed for the AR(1) signal. By looking at the partial autocorrelation functions, which partial out the influence of earlier lags when calculating later lags, we could also see that the pink noise signal falls short of the AR(1) signal's first lag autocorrelation. Lag two and three still showed significant autocorrelation in the pink noise signal but as expected not in the AR(1) one. No false positive rate increases were observed in the white noise signal. However, with high overlap the average false-positive rate seems to shift further below the nominal alpha. We found that not only is autocorrelation directly increasing the false-positive rate, but the degree of overlap also seemingly mediated this effect.

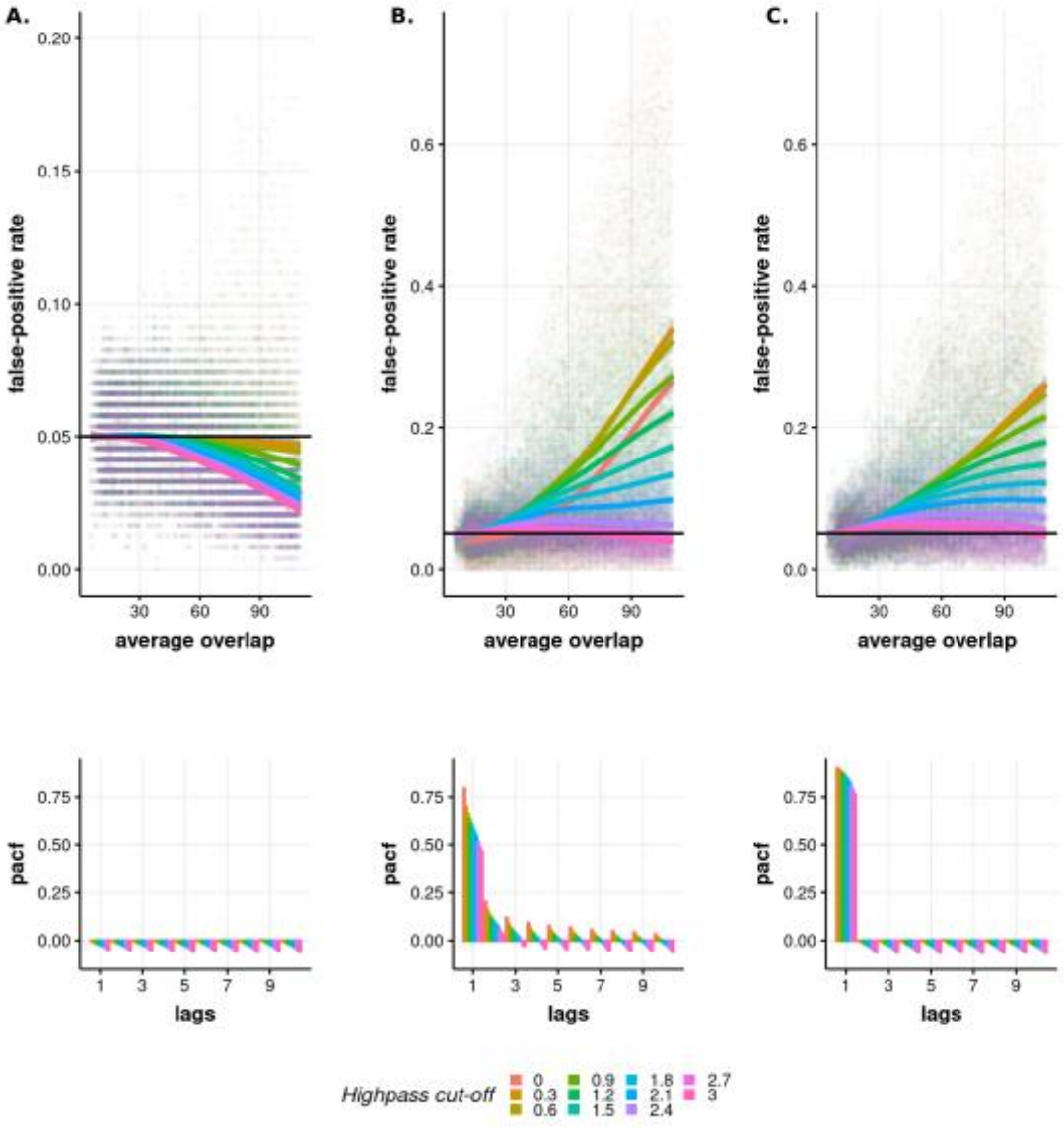
Discussion

In this report we assessed the validity of parametric statistical hypothesis tests run on the β estimates of overlap-corrected rERPs. We found the tests to be robust against violations of the independence assumption when no overlap was present or when the events' onsets were fully randomised. In the case of systematic overlap, as would be the case with fixed trial structures causing the overlap of evoked potentials, the tests quickly became invalid under autocorrelation. Furthermore, by varying the degree of overlap and autocorrelation on a more continuous scale, we

observed that the strength of the first autocorrelated lag especially could already be highly informative about the increases to the false positive rate.

Figure 6

False-Positives with Continuous Overlap and Autocorrelation



Note. The x-axis shows how many data points on average overlapped in a given model. The y-axis shows the false-positive rate. Each colour represents a different high-pass filter cut-off that practically lowered the autocorrelation present in the signal by removing low frequency components. The coloured lines were fitted to their respective data points, each representing the average of at least 242 estimates. To fit the lines, we used the default geom_smooth function's "gam" method available through ggplot2 (Wickham, 2016). Beneath, are the first ten lags of the partial autocorrelation function. Column (A) Shows the results for white noise, (B) for pink noise, and (c) for AR(1).

Both conditions with nominal type 1 errors differed from the systematic overlap conditions by having orthogonal (no overlap) and closer to orthogonal (randomised) predictors. The degree to which the modelled predictors are orthogonal seemingly mediates the increase of the false-positive rate due to autocorrelation in figure 6. By looking at the white noise simulations, one can observe how less orthogonal designs without autocorrelation actually resulted in more conservative false-positive rates. Less orthogonal predictors mean that the same datapoints are reused for different predictors making them (partial) collinear. This causes increased standard errors, as the variance present in the data cannot be distributed across predictors as clearly as I it would be without collinearity. This resulted in less false positives overall.

Interestingly, the decrease in false positives seemed to have strengthened with stronger negative partial autocorrelation lags caused by the high-pass filters (see figure 6a). While we cannot explain this effect, it lends itself well to advice for caution when generalizing the results of this report to real EEG data. Real EEG data will often contain such negatively autocorrelated lags which were absent from our artificial signals. Furthermore, our artificial signals differed from real EEG data by having been generated from stationarity stochastic processes. Thus, autocorrelation measured at an early interval in our pink, white or AR(1) noise will equal the autocorrelation measured at an interval towards the end of the signal. Real EEG's voltage deflections arise from summed post-synaptic potentials, muscle artifacts, and may wax and wane in erratic or even task-dependent manners. Ignoring the bursting nature of EEG activity has caused erroneous conclusions with other methods before (Jones, 2016). Whether such non-stationarity decreases or further increases the false-positive rates is unclear at this point. Running our simulations on the EEG data used in figure 4, we did observe the same pattern as we observed for the artificial data (see figure 5), but with large variations from subject to subject (Appendix B).

How to Proceed?

The next question becomes what the best method would be to reduce or fully remove any non-independence of the residuals to test hypotheses on a single subject level.

As showcased in figure 6, high-pass filters with cut offs as high as 2.7Hz did not sufficiently lower the partial autocorrelation function to make the tests valid. The same held true for downsampling as evident by the large correlation of early lags even after downsampling the signal to a 100th of its original sampling frequency (see figure 2). Since EEG research usually relies on high-pass filters with much lower cut offs (Luck, 2005) both of these approaches do not suffice to render the residuals independent and with that the tests remain invalid.

Traditional pre-whitening methods used in the fMRI community, and briefly mentioned in the introduction here, may also not be applicable. With higher sampling rates (designated by smaller TRs) these pre-whitening methods have been shown to be insufficient since the complexity of the covariance structure increases with higher sampling rates (Bollmann et al., 2018; Smith, 2011). Similar problems have been reported for the estimation of effective degrees of freedom (Afyouni et al., 2019). Whether the proposed algorithms to solve these issues for other kinds of data would also work for rERPs would have to be investigated in future work.

A promising sounding solution iteratively fits models in which not the data but the residuals get filtered with an optimal band-pass filter (this was also explored in Friston et al., 2000) to whiten them (Smith, 2011). This and a wavelet transforms to effectively remove any autocorrelation were suggested by Smith in their dissertation (Smith, 2011). Up until one week before finalising this report we had only considered their published work in which these considerations are missing. Had we known about these promising sounding solutions before, we would have tested them for this project. Now, this remains to be done in future work.

Conclusion

We can conclude that autocorrelation indeed poses a problem for single subject analyses and that inference drawn from overlap corrected models that deal with systematic overlap quickly become invalid. For now, overlap-correction should only be applied when statistical tests will be carried out across subjects. However, we have also shown that inference in the case of low overlap or less systematic overlap may not be as heavily affected by autocorrelation. Before within-subject

analyses on overlap-corrected rERPs can be considered, the solutions proposed by Smith should be thoroughly tested for their feasibility and flexibility.

References

Afyouni, S., Smith, S. M., & Nichols, T. E. (2019). Effective degrees of freedom of the Pearson's correlation coefficient under autocorrelation. *NeuroImage*, *199*, 609–625. https://doi.org/10.1016/j.neuroimage.2019.05.011

Arbabshirani, M. R., Preda, A., Vaidya, J. G., Potkin, S. G., Pearlson, G., Voyvodic, J., Mathalon, D., van Erp, T., Michael, A., Kiehl, K. A., Turner, J. A., & Calhoun, V. D. (2019). Autoconnectivity: A new perspective on human brain function. *Journal of Neuroscience Methods*, *323*, 68–76. https://doi.org/10.1016/j.jneumeth.2019.03.015

Baayen, H., Vasishth, S., Kliegl, R., & Bates, D. (2017). The cave of shadows: Addressing the human factor with generalized additive mixed models. *Journal of Memory and Language*, *94*, 206–234. https://doi.org/10.1016/j.jml.2016.11.006

Bae, G.-Y., & Luck, S. J. (2019). *Appropriate Correction for Multiple Comparisons in Decoding of ERP Data: A Re-Analysis of Bae & Luck (2018)* [Preprint]. Neuroscience. https://doi.org/10.1101/672741

Bartlett, M. S. (1946). On the Theoretical Specification and Sampling Properties of Autocorrelated Time-Series. *Supplement to the Journal of the Royal Statistical Society*, *8*(1), 27–41. https://doi.org/10.2307/2983611

Bollmann, S., Puckett, A. M., Cunnington, R., & Barth, M. (2018). Serial correlations in single-subject fMRI with sub-second TR. *NeuroImage*, *166*, 152–166. https://doi.org/10.1016/j.neuroimage.2017.10.043

Burock, M. A., & Dale, A. M. (2000). Estimation and detection of event-related fMRI signals with temporally correlated noise: A statistically efficient and unbiased approach. *Human Brain Mapping*, *11*(4), 249–260. https://doi.org/10.1002/1097-0193(200012)11:4<249::aid-hbm20>3.0.co;2-5

Carter, J. A., & Winn, J. N. (2009). PARAMETER ESTIMATION FROM TIME-SERIES DATA WITH CORRELATED ERRORS: A WAVELET-BASED METHOD AND ITS APPLICATION TO TRANSIT

LIGHT CURVES. *The Astrophysical Journal*, *704*(1), 51–67. https://doi.org/10.1088/0004-637X/704/1/51

Cliff, O. M., Novelli, L., Fulcher, B. D., Shine, J. M., & Lizier, J. T. (2021). Assessing the significance of directed and multivariate measures of linear dependence between time series. *Physical Review Research*, *3*(1), 013145. https://doi.org/10.1103/PhysRevResearch.3.013145

Cohen, M. X. (2014). *Analyzing neural time series data: Theory and practice*. The MIT Press.

Cornelissen, T., Sassenhagen, J., & Võ, M. L.-H. (2019). Improving free-viewing fixation-related EEG potentials with continuous-time regression. *Journal of Neuroscience Methods*, *313*, 77–94. https://doi.org/10.1016/j.jneumeth.2018.12.010

Dandekar, S., Ding, J., Privitera, C., Carney, T., & Klein, S. A. (2012). The Fixation and Saccade P3. *PLoS ONE*, *7*(11), e48761. https://doi.org/10.1371/journal.pone.0048761

Datseris, G., Isensee, J., Pech, S., & Gál, T. (2020). DrWatson: The perfect sidekick for your scientific inquiries. *Journal of Open Source Software*, *5*(54), 2673. https://doi.org/10.21105/joss.02673

Dimigen, O., & Ehinger, B. V. (2021). Regression-based analysis of combined EEG and eye-tracking data: Theory and applications. *Journal of Vision*, *21*(1), 3. https://doi.org/10.1167/jov.21.1.3

Doyle, J. A., & Evans, A. C. (2018). What Colour is Neural Noise? *ArXiv:1806.03704 [q-Bio]*. http://arxiv.org/abs/1806.03704

Ehinger, B. V., & Dimigen, O. (2019). Unfold: An integrated toolbox for overlap correction, non-linear modeling, and regression-based EEG analysis. *PeerJ*, *7*, e7838. https://doi.org/10.7717/peerj.7838

Fadili, M. J., & Bullmore, E. T. (2002). Wavelet-Generalized Least Squares: A New BLU Estimator of Linear Regression Models with 1/f Errors. *NeuroImage*, *15*(1), 217–232. https://doi.org/10.1006/nimg.2001.0955

Friston, K. J., Josephs, O., Zarahn, E., Holmes, A. P., Rouquette, S., & Poline, J.-B. (2000). To Smooth or Not to Smooth? *NeuroImage*, *12*(2), 196–208. https://doi.org/10.1006/nimg.2000.0609

Gert, A. L., Ehinger, B. V., Timm, S., Kietzmann, T. C., & König, P. (2021). Wild lab: A naturalistic free viewing experiment reveals previously unknown EEG signatures of face processing. *BioRxiv*, 2021.07.02.450779. https://doi.org/10.1101/2021.07.02.450779

Gramfort, A. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, *7*. https://doi.org/10.3389/fnins.2013.00267

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice* (2nd edition). Otexts, online, open-access textbook.

Jones, S. R. (2016). When brain rhythms aren't "rhythmic": Implication for their mechanisms and meaning. *Current Opinion in Neurobiology*, *40*, 72–80. https://doi.org/10.1016/j.conb.2016.06.010

Kissling, W. D., & Carl, G. (2008). Spatial autocorrelation and the selection of simultaneous autoregressive models. *Global Ecology and Biogeography*, *17*(1), 59–71. https://doi.org/10.1111/j.1466-8238.2007.00334.x

Kristensen, E., Rivet, B., & Guérin-Dugué, A. (2017). Estimation of overlapped Eye Fixation Related Potentials: The General Linear Model, a more flexible framework than the ADJAR algorithm. *Journal of Eye Movement Research*, *10*(1). https://doi.org/1614425591

Linkenkaer-Hansen, K., Nikouline, V. V., Palva, J. M., & Ilmoniemi, R. J. (2001). Long-Range Temporal Correlations and Scaling Behavior in Human Brain Oscillations. *Journal of Neuroscience*, *21*(4), 1370–1377. https://doi.org/10.1523/JNEUROSCI.21-04-01370.2001

Luck, S. J. (2005). *An introduction to the event-related potential technique*. MIT Press.

Lund, T. E., Madsen, K. H., Sidaros, K., Luo, W.-L., & Nichols, T. E. (2006). Non-white noise in fMRI: Does modelling have an impact? *NeuroImage*, *29*(1), 54–66. https://doi.org/10.1016/j.neuroimage.2005.07.005

Olszowy, W., Aston, J., Rua, C., & Williams, G. B. (2019). Accurate autocorrelation modeling substantially improves fMRI reliability. *Nature Communications*, *10*(1), 1220. https://doi.org/10.1038/s41467-019-09230-w

Purdon, P. L., & Weisskoff, R. M. (1998). Effect of temporal autocorrelation due to physiological noise and stimulus paradigm on voxel-level false-positive rates in fMRI. *Human Brain Mapping*, *6*(4), 239–249. https://doi.org/10.1002/(sici)1097-0193(1998)6:4<239::aid-hbm4>3.0.co;2-4

Shinn, M., Hu, A., Turner, L., Noble, S., Achard, S., Anticevic, A., Scheinost, D., Constable, R. T., Lee, D., Bullmore, E. T., & Murray, J. D. (2021). *Spatial and temporal autocorrelation weave human brain networks* [Preprint]. Neuroscience. https://doi.org/10.1101/2021.06.01.446561

Skukies, R. S. (2020). *Validation of Measures of Consciousness using Propofol Anesthesia* [University of Oslo]. DUO Research Archive. http://hdl.handle.net/10852/79300

Smith, N. J. (2011). *Scaling up psycholinguistics* [UC San Diego]. https://escholarship.org/uc/item/9hh0x7tq

Smith, N. J., & Kutas, M. (2015a). Regression-based estimation of ERP waveforms: I. The rERP framework. *Psychophysiology*, *52*(2), 157–168. https://doi.org/10.1111/psyp.12317

Smith, N. J., & Kutas, M. (2015b). Regression-based estimation of ERP waveforms: II. Nonlinear effects, overlap correction, and practical considerations. *Psychophysiology*, *52*(2), 169–181. https://doi.org/10.1111/psyp.12320

Sul, D., Phillips, P. C. B., & Choi, C.-Y. (2005). Prewhitening Bias in HAC Estimation*. *Oxford Bulletin of Economics and Statistics*, *67*(4), 517–546. https://doi.org/10.1111/j.1468-0084.2005.00130.x

Thul, R., Conklin, K., & Barr, D. J. (2021). Using GAMMs to model trial-by-trial fluctuations in experimental data: More risks but hardly any benefit. *Journal of Memory and Language*, *120*, 104247. https://doi.org/10.1016/j.jml.2021.104247

van Rij, J., Hendriks, P., van Rijn, H., Baayen, R. H., & Wood, S. N. (2019). Analyzing the Time Course of Pupillometric Data. *Trends in Hearing*, *23*, 2331216519832483. https://doi.org/10.1177/2331216519832483

Woolrich, M. W., Ripley, B. D., Brady, M., & Smith, S. M. (2001). Temporal Autocorrelation in Univariate Linear Modeling of FMRI Data. *NeuroImage*, *14*(6), 1370–1386. https://doi.org/10.1006/nimg.2001.0931

Worsley, K. J., & Friston, K. J. (1995). Analysis of fMRI Time-Series Revisited—Again. *NeuroImage*, *2*(3), 173–181. https://doi.org/10.1006/nimg.1995.1023

Zarahn, E., Aguirre, G. K., & D'Esposito, M. (1997). Empirical analyses of BOLD fMRI statistics. I. Spatially unsmoothed data collected under null-hypothesis conditions. *NeuroImage*, *5*(3), 179–197. https://doi.org/10.1006/nimg.1997.0263

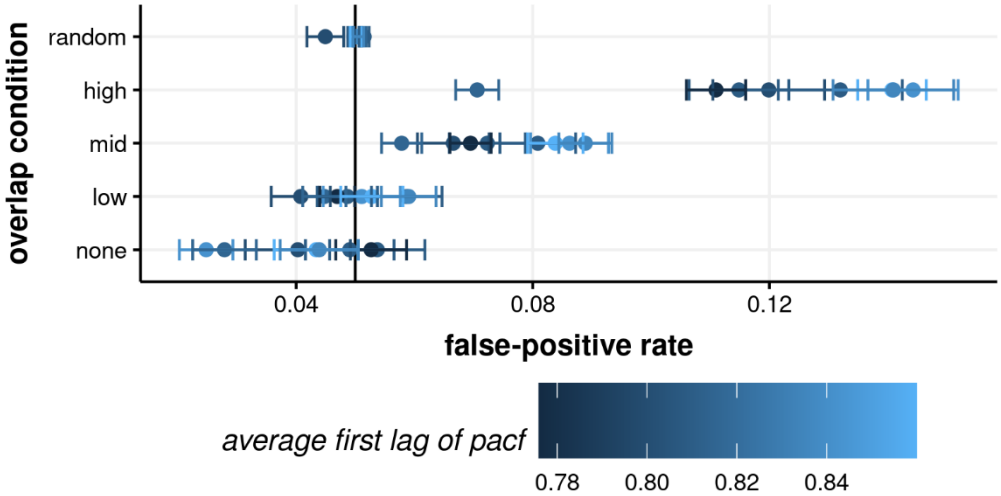
Appendix

A.

```
%Synthesise pink noise:
Nx = 2^16;  % number of samples to synthesize
B = [0.049922035 -0.095993537 0.050612699 -0.004408786];
A = [1 -2.494956002   2.017265875  -0.522189400];
nT60 = round(log(1000)/(1-max(abs(roots(A))))); % T60 est.
v = randn(1,Nx+nT60); % Gaussian white noise: N(0,1)
x = filter(B,A,v);    % Apply 1/F roll-off to PSD
x = x(nT60+1:end);    % Skip transient response
```

B

False Positives by Overlap



Note. The y-axis shows the different overlap conditions we modelled. Each plotted point represents one Subject (7 in total). We calculated the average false -positive rate over two data sets per subject and 16 different permutations of randomly selected channels (to get signals of sufficient length). 400 models were regressed onto these signals with the error bars denoting the standard error across the 16 generated signals. The black vertical line represents the nominal alpha value, values beyond it indicate an invalid test.