2024 ■ Vol. 20 ■ no. 2

Local decorrelation for error bars in time series

Denis Cousineau [a], Anthony Proulx [b], Annabelle Potvin-Pilon [a] & Daniel Fiset [b]

[a]Université d'Ottawa
[b]Université du Québec en Outaouais

Abstract ■ Time series and electroencephalographic data are often noisy sources of data. In addition, the samples are often small or medium so that confidence intervals for a given time point taken in isolation may be large. Decorrelation techniques were shown to be adequate and exact for repeated-measure designs where correlation is assumed constant across pairs of measurements. This assumption cannot be assumed in time series and electroencephalographic data where correlations are most-likely vanishing with temporal distance between pairs of points. Herein, we present a decorrelation technique based on an assumption of local correlation. This technique is illustrated with fMRI data from 14 participants and from EEG data from 24 participants.

Keywords ■ Error bars, confidence intervals, time series, ERP, EEG, fMRI. **Tools** ■ R, Matlab, Mathematica.

denis.cousineau@uottawa.ca

10.20982/tqmp.20.2.p173

Introduction

Time series and electroencephalographic (EEG) data represent collections of information organized temporally. These data often involve repeated measures within a within-subject design. In such designs, measurements are taken multiple times under different conditions, which may be temporally organized, such as in clinical pre-treatment and post-treatment studies. However, repeated measurements are not always conducted in a time-related order. For instance, in experiments manipulating the difficulty of puzzle-solving tasks, the difficulty levels do not necessarily increase over time, i.e. the sequence of measurements can be counterbalanced across participants or randomized.

Within-subject designs are typically recommended due to their higher statistical power (e.g. Field, 2009; Howell, 2010; Tabachnick et al., 2013). However, traditional, also called standalone, confidence intervals do not account for this increased power, which translates to improved precision in the estimates. In seminal work, Loftus and Masson (1994) proposed an adjustment to the width of error bars to reflect this enhanced precision. The source of this additional precision is to be found in the correlation among repeated measures (Cousineau, 2019). Positive correlation indicates that participants' scores exhibit regularity (e.g., individuals with high scores on one measure tend to have high scores on other measures). This variation can be quantified using, for example, the between-subject variance in the context of ANOVAs. It can also be eliminated, as inter-participant differences are frequently of no theoretical interest (Cousineau, 2005; Morey, 2008). The remaining variance is then attributed to the factor(s) of interest (here the repeated-measure factor or factors). By discarding a portion of the uncertainty, the effect size is more narrowly estimated, resulting in greater precision (Jané et al., 2024). This approach is sometimes called decorrelation.

Another limitation of standalone confidence intervals is their focus on the precision of a single point in isolation. They are therefore unable to assess the relative position of that point to other points. Yet, researchers are most of the time interested in comparing conditions to other conditions, something that cannot be done with standalone confidence intervals. Goldstein and Healy (1995) proposed to adjust the width of the confidence intervals to reflect this objective of performing pairwise comparisons. Such adjustment was called a difference adjustment in subsequent works as it allows examining the difference between conditions (Baguley, 2012). Difference-adjusted confidence intervals are wider, by 41% (width increased by $\sqrt{2} \approx 1.41$) when homogeneity of variance is assumed.

To implement the gain in precision within confidence intervals, Cousineau (2017) proposed the concept of adjusted confidence intervals. These adjustments account for the impact of the experimental design (within-subject vs. between-groups) and the intended purpose of the confidence intervals (e.g., for pairwise comparisons or comparisons to hypothesized values determined a priori). Additionally, they consider the sampling process and the population size when it is not infinite. In this framework, standalone confidence intervals are relevant for between-group designs when performing comparisons to an a priori value from a sample obtained with simple randomized sampling out of a population of infinite size.

All these precepts that apply to the average of conditions were extended in superb (summary plots with adjusted error bars) to various other descriptive statistics (Cohen's d, proportions, frequencies, etc.; Cousineau et al., 2021; also see Laurencelle & Cousineau, 2023a, 2023b).

Because time series and EEG data can be seen as repeated measures, it would be sensible to use the decorrelation technique. As demonstrated by Cousineau (2019), the simplest form of decorrelation involves adjusting the confidence interval by multiplying its width by $\sqrt{1-r}$ where r is a measure of correlation between pairs of repeated measures. This approach implicitly assumes that the correlation is stationary over time, meaning that pairs of contiguous measurements have the same correlation as pairs of measurements separated by longer time intervals. When coupled with the assumption of homogeneity of variances, this structure is known as compound symmetry (Winer et al., 1991).

Sadly, for time series, the assumption of stationary correlation between pairs of measurements is often questionable. In the following section, we expand on compound symmetry and generalize the structure to one with correlations decaying with lag, also called autoregressive covariance structures. Under autoregressive covariance, we propose using a local decorrelation technique. This technique is described in detail and illustrated with simulated data. Subsequently, we demonstrate the generality of this technique using two real datasets.

Compound symmetry and beyond

In a repeated measures design with p measures, the scores are conceived as being sampled from a p-dimensional multivariate distribution. When normality is assumed, the distribution is a multinormal distribution. This theoretical distribution is characterized by a location vector μ and a variance-covariance matrix $\mathbf{\Sigma}$ (Winer et al., 1991; Kincaid, n.c.). When the scores are independent from each other, the covariance matrix has only variances along the main diagonal and zero elsewhere, a structure sometimes called variance components (VC). An example with 4 variables would be

$$\mathbf{\Sigma}_{VC} = \begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & 0 \\ 0 & 0 & 0 & \sigma_4^2 \end{bmatrix} \quad \text{(VC)}$$

where σ_i^2 is the variance of the ith measurement. If we assume that all the variances are identical, we additionally are in a situation of homogeneity of variances. The scenario of variance components plus homogeneity is actually the structure assumed in between-group analyses, and the multinormal distribution simplifies to normal distributions for each measurement.

The compound symmetry scenario assumes homogeneity of variance as well as homogeneity of covariance between the variables, which results in non-null entries in the off-diagonal cells.

$$\mathbf{\Sigma}_{CS} = \begin{bmatrix} \sigma^2 & \rho\sigma^2 & \rho\sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 & \rho\sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \rho\sigma^2 & \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \rho\sigma^2 & \rho\sigma^2 & \sigma^2 \end{bmatrix} \quad \text{(CS)}$$

In this scenario, the parameter ρ is the correlation in the population between pairs of variables. A simple estimator of ρ is obtain by calculating the mean pairwise correlation; Laurencelle and Cousineau (2023b) suggested a measure based on Cronbach's α; for two measures, the exact measure of correlation is the rectified Pearson correlation (Cousineau & Goulet-Pelletier, 2021).

The autoregressive structure, formally called the autoregressive covariance of order 1 (sometimes noted AR(1)), is characterized by an amount of correlation that declines exponentially as we move away from the main diagonal. One way to achieve this is to raise the correlation to increasingly larger power, ρ^1, ρ^2, ρ^3, etc. as we move away from the main diagonal, resulting in

$$\mathbf{\Sigma}_{AR(1)} = \begin{bmatrix} \sigma^2 & \rho^1\sigma^2 & \rho^2\sigma^2 & \rho^3\sigma^2 \\ \rho^1\sigma^2 & \sigma^2 & \rho^1\sigma^2 & \rho^2\sigma^2 \\ \rho^2\sigma^2 & \rho^1\sigma^2 & \sigma^2 & \rho^1\sigma^2 \\ \rho^3\sigma^2 & \rho^2\sigma^2 & \rho^1\sigma^2 & \sigma^2 \end{bmatrix} \quad \text{(AR1)}$$

There exists other covariance structures including the spatial power structure where the correlation has a power related to the exact duration separating the measurements. This is sensible when the lag between time points can be expressed quantitatively (in days, for example).

To determine which covariance structure fits best, formal tests can be performed (e.g., Keselman et al., 1998). Choosing a covariance structure is a common requirement form mixed models (e.g., Howell, 2010). However, for very large covariance matrix (as can be the case with long EEG stream composed of hundreds or thousands of voltage

Table 1 ■ ANOVA results for the data of the first illustration shown in Figure 2

Effect	df	SS	MS	F	p	η_p^2
Within group						
Time (T)	25	$122,780.6$	$4,911.23$	25.353	<.001	0.459
Error ($S \times T$)	600	$116,182.0$	193.63			
Between group						
Error (e)	24	$28,270.1$	$1,177.92$			

Note. As expected, the Winer test of compound symmetry is highly significant ($W = 3495$, $\chi^2(349) = 2203$, $p < .001$) as well as the Mauchly test of sphericity ($M = 1902.8$, $\chi^2(324) = 1227.8$, $p < .001$; Huynh-Feldt $\varepsilon = .225$).

potential measurements), the covariances away from the main diagonal rapidly become dominated by noise. Also, the above structures predict zero correlation for distant measurements which is doubtful as the participants may change over time, but will certainly keep some characteristics. Thus, distant covariances do not bring much information and consequently, formally testing for covariance structure is likely to be little informative.

Herein, we adopt a more pragmatic view where it is assumed that close-by measurements have informative covariance but that their relevance declines rapidly. Consequently, we consider the average correlation of the other measurements to a given measurement, weighted by their distance to this measurement.

To weight the correlations, we chose a Gaussian function, that is, weights that are given by the Gaussian equation:

$$W_\sigma(d) = \frac{e^{-d^2/(2\sigma^2)}}{\sqrt{2\pi}\sigma}; \quad W_\sigma(0) = 0 \tag{1}$$

in which σ is the radius factor of the weights and d is the distance, i.e., the time lag between two measurements. The top of the Gaussian is replaced by 0 as the correlation of a variable with itself does not reflect the correlations in the population. For example, using a radius of 1, the weights, for distances varying from -5 to +5 are

$$0.0000,\ 0.0001,\ 0.0044,\ 0.0540,\ 0.2420,\ 0.000,\ 0.2420,\ 0.0540,\ 0.0044,\ 0.001,\ 0.000$$

The weighted average correlation to the ith time point, noted r_{LD} herein, is obtained with

$$r_{LD} = \frac{\sum_{j=1}^{p} W_\sigma(i-j) r_{i,j}}{\sum_{j=1}^{p} W_\sigma(i-j)} \tag{2}$$

where p is the number of time point, $r_{i,j}$ is the correlation between the ith and the jth measurements, and $i - j$ is the lag between the two time points.

The resulting weighted average correlation r_{LD} is then used to compute the correlation-adjusted confidence intervals for the ith condition,

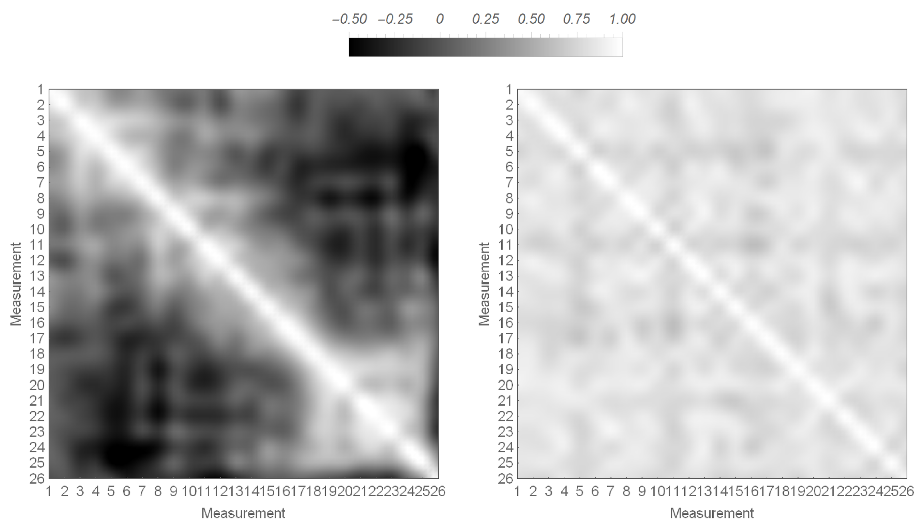
$$CI = M \pm SE\,\sqrt{2}\,\sqrt{1 - r_{LD}} \times t_{n-1}(\gamma) \tag{3}$$

where M is the mean of measurements at a given time point, SE, the standard error, is given by the standard deviation of the same measurements divided by $\sqrt{n}$ with n the number of subjects, and $t_{n-1}(\gamma)$ is the t critical value at the confidence level γ (typically 95%) with degree of freedom $n - 1$.

A first illustration

To illustrate the procedure, we generated simulated data. The data were meant to represent annual variation in a certain measure performed biweekly (there are therefore 26 time measurements per participant) from 24 simulated participants. To that end, the population grand mean was 100 units modulated by a sine wave whose amplitude was one standard deviation above and below the grand mean. We examine two simulated datasets, the first having an autoregressive covariance structure, the second a compound symmetry structure. While the first scenario is more plausible for time series data, comparing it with the second scenario will help highlight the benefits of the local decorrelation technique.

Figure 1 ■ Heat map showing the correlations from data based on two different covariance structures. Left: based on an autoregressive covariance structure AR(1), right: based on a compound symmetry covariance structure. The correlations along the main diagonal are 1.00 but are not used in computing the mean correlation.



Autoregressive covariance

The population covariance matrix had AR(1) structure with $\sigma = 15$ and $\rho = .75$. The generated dataset from these specifications can be found as supplementary material, file `illustrationWithAR1.tsv` in the folder `FirstIllustration`. With such parameters, the ANOVA, not surprisingly, shows a very significant effect of Time, as seen in Table 1. The effect size η_p^2 is large (0.46).

The sample correlations show a decreasing trend with increasing lag between the measurement times. The heat map in Figure 1, left panel, shows the correlations for each pair of measurements. The correlations along the main diagonal are 1.00 (correlation of a measurement with itself). The correlations become weaker as we move further away from the main diagonal. As the sample is small (25 simulated participants), correlations are rarely zero but end up fluctuating around this value (in the extreme, correlation between the first and the last measurement is expected to be $\rho^{25} = .0008$ with a large expected 95% confidence interval or [-.395, +.396] as the sample is fairly small).

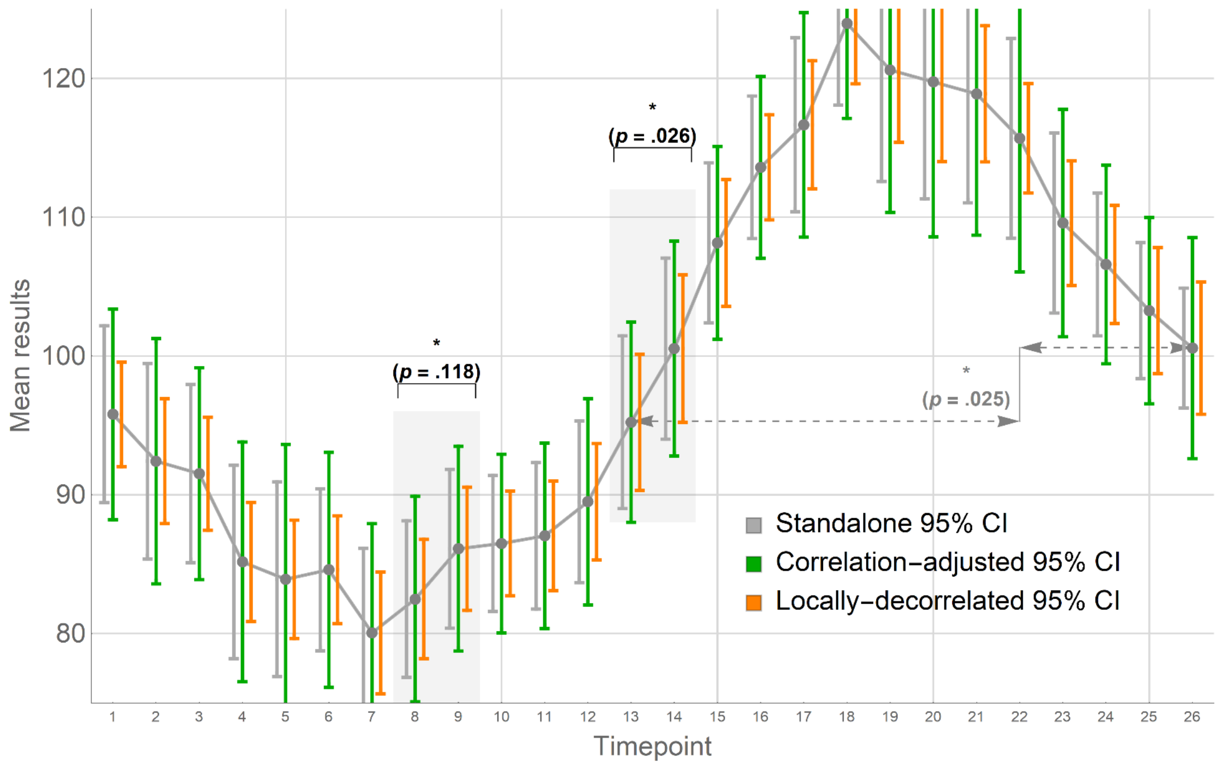
Figure 2 shows the means along with confidence intervals (CI) adjusted in various ways. The gray error bars denote the (unadjusted) standalone 95% confidence intervals. As argued elsewhere, they are useless for pairwise comparisons. They can only be used if the purpose is to compare one measurement to an a priori value. For example, does the average measurement at time 2 differ from 100 (answer is yes, $M = 92.4$, $t(24) = 2.22$, $p = .036$). The gray error bar on that measurement just barely excludes the value 100, explaining why the p value of the difference is not stronger.

The green error bars show the Cousineau-Morey adjusted 95% confidence intervals. It is first adjusted for pairwise differences (Baguley, 2012): their widths are $\sqrt{2}$ = 1.41 times longer than the standalone confidence intervals. It is further adjusted for the average correlation across the measurements (here the average correlation $\bar{r}$ is .163). This is a moderately weak correlation so that the correlation adjustment is fairly modest, reducing the error bar widths by a factor of approximately $\sqrt{1-\bar{r}} = .91$ (that is, they are shortened by close to 10%). Taking the two adjustments together, the error bars are 28% longer than the standalone error bars ($1.41 \times 0.91 = 1.28$). These error bars would be adequate if the correlation was stationary across measurements (a CS structure). However, in an AR(1) structure, they are too long when comparing nearby measurements.

The orange error bars show the locally-decorrelated 95% confidence intervals. Using weighted average with a radius of 1, the mean correlation r_{LD} is –on average over the 26 measurements– .729 (ranging from .85 for the 22th measurement to .39 for the 26th measurement). The locally-decorrelated confidence intervals are the shortest with a few exceptions.

To illustrate the utility of the locally-decorrelated confidence intervals, some pairs have been highlighted in Figure 2 (how the p values in the Figure were computed is discussed later). The measurements 13 and 14 for example

Figure 2 ■ Plot of the mean measurements as a function of time, along with standalone 95% confidence intervals (CI; gray), Correlation an difference-adjusted 95% CI (green), and locally-decorrelated 95% CI (orange).



can be tested using an a priori contrast (Howell & Lacroix, 2012). Because they are contiguous, correlation between these two measurements is strong ($r_{13,14} = .697$), and the contrast suggests a significant difference ($F(1,24) = 4.865$, $p = .037$). It could alternatively be tested with an LSD test with no difference in the conclusion ($t(24) = 2.206$, $p = .037$). These formal results are perfectly corroborated by the locally decorrelated confidence interval as the mean for time point 13 is close but not included in the interval of time point 14 (and vice-versa).

The most time-separated means, measurements 1 and 26 are not significantly different according to the above mean comparison tests (with the contrast method, we get $F(1,24) = 1.58, p = .221$). The reason is that their correlation is about null ($r_{1,26} = -.037$). The difference-adjusted confidence interval (the standalone CI width times $\sqrt{2}$) captures this lack of difference well. However, discounting the effect of lag, we can test the differences using the total mean squared error, as suggested next using an adjusted t-test.

A t-test discounting the lag effect

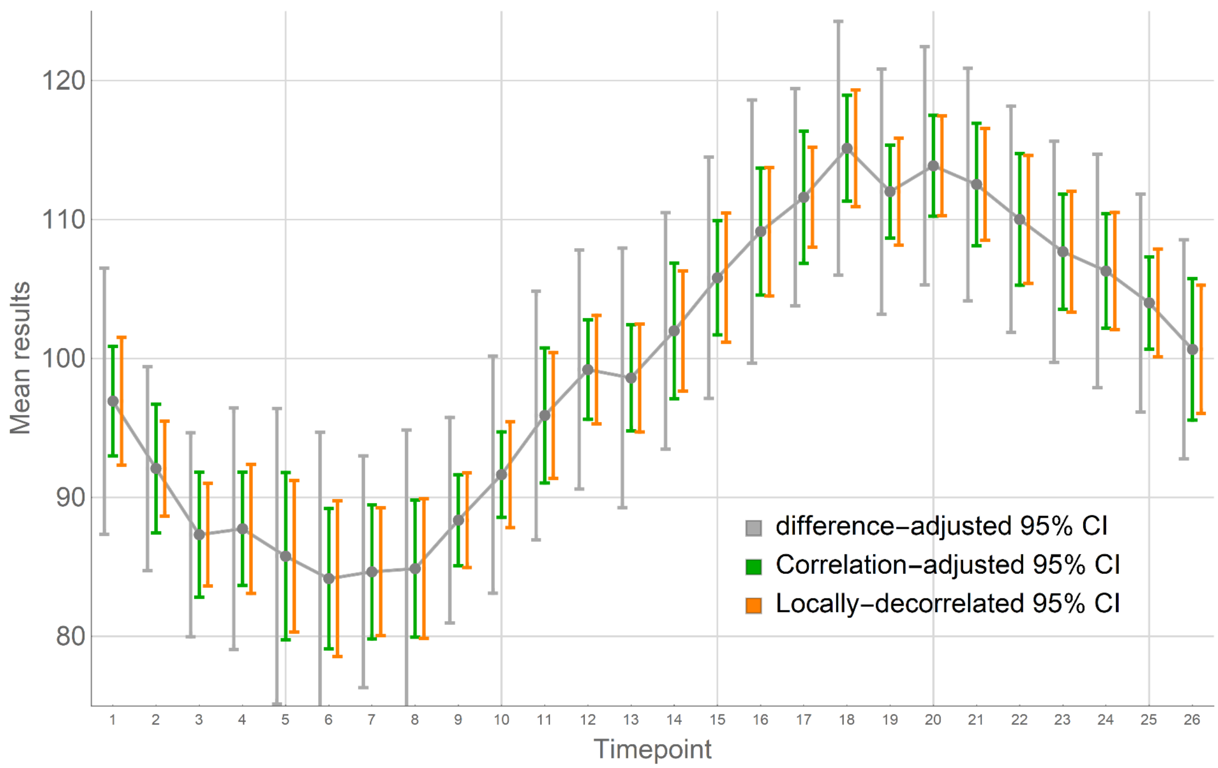
The total mean squared error is a quantity which indicates the amount of variance that cannot be attributed to the factor examined (the time effect). In Table 1, the total sum of squared error (e) and sum of squared within-group error ($S \times E$) yields $116,182.0 + 28,270.1 = 144,452.1$. When this total is divided by the degrees of freedom of the error terms, 624, we get a mean squared error of 231.49. We get the same estimate when we analyze the data using a between-subject design in which there is a single error term. It is therefore the total mean squared error. Using this error term in a test of mean difference, we have

$$t_{LD} = \frac{|M_i - M_j|}{\sqrt{\frac{2}{n}}\sqrt{1-\overline{r}_{LD}}\sqrt{CM_e}} \tag{4}$$

where M_i and M_j are the two means to compare, $\overline{r}_{LD}$ is the overall mean correlation obtained locally with the weighted average, and CMe is the error term. Doing so for the first and the last measurements ($M_1 = 95.798$, $M_{26} = 100.567$, $\overline{r}_{LD} = 729$), we obtain $t(24) = 2.129$, $p = .044$. This borderline yet significant result is mirrored by the orange CI of time 26 which barely exclude time 1 (and vice versa).

The purpose of the term $\sqrt{1-\overline{r}_{LD}}\sqrt{CM_e}$ is to partition the error term into a component which is neither at-

Figure 3 ■ Plot of the mean measurements in the same format as Figure 2. The data's covariance are compound symmetric.



tributed to the lag effect nor to the time effect. It does so by discounting the proportion of variance which is not explained by local correlation.

If we examine all the 325 possible pairs of measurements, we find 28 cases where the adjusted t-test reports a significant mean difference but where the LSD is not significant. One such case is the pairs of measurements from the 13th time and the 26th time. A LSD test reports no significant difference ($t(24) = 1.312$, $p = .202$); the adjusted t-test suggests a significant result ($t(24) = 2.387$, $p = .025$). We can see that the mean on the 26th measurement is about identical to the mean on the 14th measurement. Above, we saw that the 13th and 14th measurements were significantly different. The locally decorrelated confidence intervals do suggests that the measurements on the 26th time point are different on average from the measurements of the 13th time point.

Compound symmetry covariance

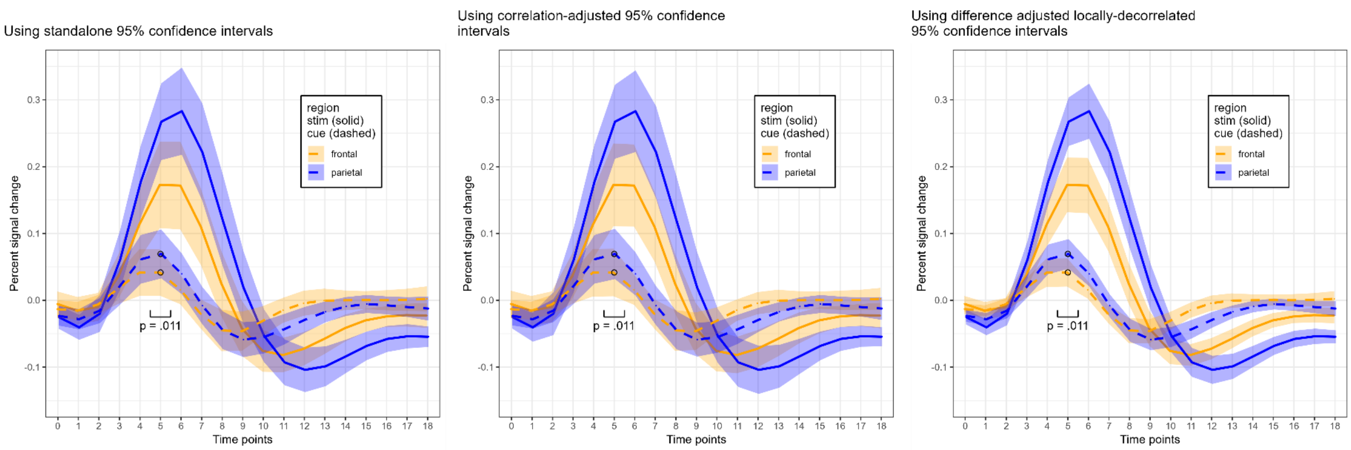
In a second scenario, we generated simulated data as before except that the covariance matrix at the level of the population is compound symmetric (the parameter ρ being set to 0.75). The generated data are available as supplementary material with this manuscript, folder `FirstIllustration`, file `IllustrationWithCS.tsv`. Figure 1, right panel shows a heat map of the correlation matrix and Figure 3 shows the means with various error bars.

The mean pairwise correlation is .756 whereas the mean weighted correlation is .754. In other words, the correlations are similar for short lags as they are overall, clearly indicating that we are not in an AR(1) scenario. That result was hinted by Figure 1, right panel, where the correlations are mildly fluctuating (except along the main diagonal). In such a scenario, there should be no benefit of using a local decorrelation technique over a regular decorrelation technique.

Inspection of the confidence intervals in Figure 3 confirms this: the correlation-adjusted and the locally-decorrelated 95% confidence intervals are about the same. Both are about $\sqrt{1-\bar{r}} \approx 50\%$ shorter than the confidence intervals that are only difference-adjusted (i.e., ignoring correlation).

As seen, the difference-adjusted intervals are the longest. They are meant to compare independent groups which is not the case in the present simulated dataset. The correlation-adjusted (which could be labeled the globally-decorrelated intervals) and the locally-decorrelated inter-

Figure 4 ■ Waskom et al. (2017) data for parietal (lighter curves) and frontal (darker curves) in the "Stimulus" conditions (top curves) and the "cue" conditions (bottom curves). In the "cue" conditions, the usual confidence intervals suggest no difference but the LD intervals indicates a difference between the parietal/frontal curves.



vals are about the same. This is expected as the structure being compound symmetric, correlation is stationary for any pair of measurements.

The first illustration was composed of a single time series. Comparisons involve different time points along the curve. In the subsequent illustration, the technique is applied to a dataset with distinct curves from the same participants. In this setting, the intervals are convenient to compare contiguous points but also different curves at a given time point.

Illustration with fMRI data from Waskom et al. (2017)

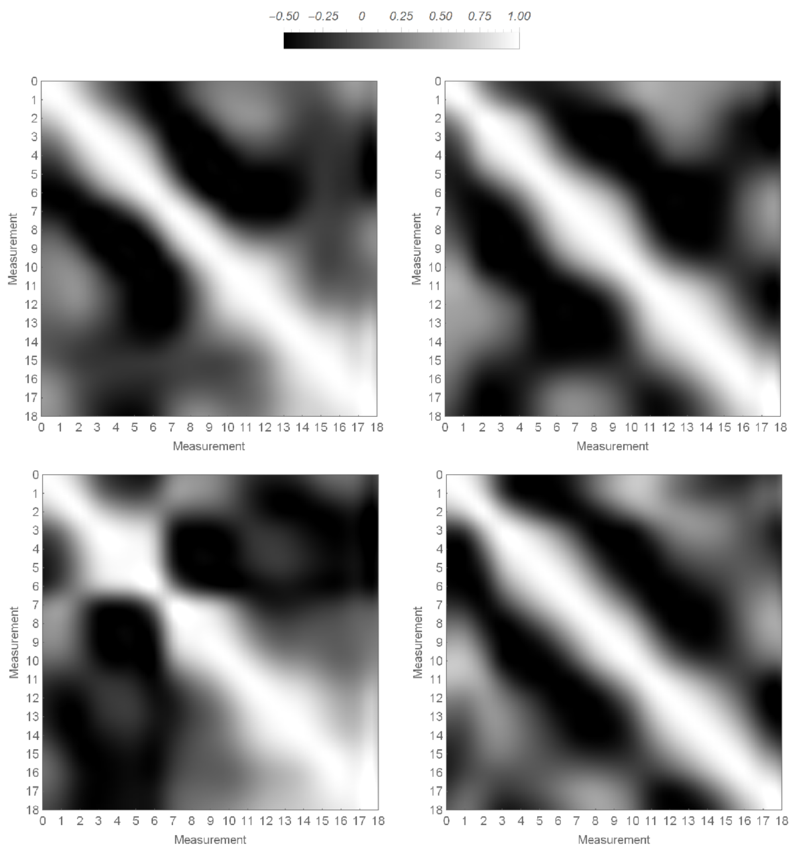
Waskom et al. (2017) published data from an fMRI experiment. The data showed the evolution of the BOLD signal over time in two regions of interest, the parietal and the frontal. The data are available from the first author's Git repository (github.com/mwaskom/Waskom_CerebCortex_2017). It contains data from 14 participants broken down by region of interest (Intra-Parietal Sulcus IPS or Inferior Frontal Sulcus FPS) and by stimulus condition (Stimulus+Cue or Cue-alone). The recordings were converted to finite impulse responses (FIR). In the original article, they report the means in their Figure 6, the two top left panels. They mentioned that the error bars were obtained with bootstrap from the simulation of a multilevel model (Politzer-Ahles, 2017) but it is not specified if they represent standard errors or confidence intervals.

The mean FIR are shown in Figure 4. The left panel shows the standalone 95% confidence intervals. The figure resembles what Python's seaborn package produces (Waskom, 2021; see seaborn.pydata.org/examples/errorband_lineplots.html) although this package uses bootstrap confidence intervals.

If the reader wants to compare the curves at various time points, a correlation-adjusted confidence interval would be preferable. The central panel shows the 95% correlation-adjusted confidence intervals. There is very little difference between the left and the central panels because correlations are on average close to null (mean correlations of -.01). Yet, the correlation heat maps shown in Figure 5 indicate that they are not stationary with high correlations for nearby measurements. The correlations weighted with a Gaussian whose radius is 1 indicate mean correlation $\bar{r}_{LD}$ of .765.

Thus, we plotted the means along with locally-decorrelated 95% confidence intervals (and the difference adjustment). The intervals are approximately 33% shorter (multiplied by $\sqrt{1-.765} \times 1.41 \approx 0.683$). In the previous panels, no hint of differences between the dashed lines are visible in the first time points (from 0 to 9, all the points of one curve are included in the confidence interval of the second curve). The locally-decorrelated confidence intervals however suggest significant differences between the dashed lines for time points 4 to 7. As an instance, we highlighted time point 5 in the figure. A simple paired t-test indicates a significant difference ($t(26) = 4.12, p = .001$). A locally-decorrelated t-test (Eq. 4) returns $t_{LD}(26) = 2.72$, $p = .011$. The reason that the paired t-test's significance is so strong is that the correlation between these two specific measurements is quite high ($r = .97$) whereas the mean local average correlation using the Gaussian weights

Figure 5 ■ Heat map showing the correlations in the Waskom et al. (2017) fMRI dataset. Top panels are for the IPS region of interest and bottom panels are for the FPS region of interest; left panels are for the Stimulus+Cue condition, and right panels are for the Cue-alone condition.



is .765. Thus, the locally-decorrelated 95% confidence intervals return more adequate depictions of the significant differences than the (globally) decorrelated and standalone confidence intervals.

This example is an example of a situation where the covariance matrix does not have a simple structure. It is neither CS nor AR(1) with correlations which fluctuate depending on the time points. For example, the falling in FIR at times 7, 8, 9 and 10 correlates negatively with the rising in FIR at times 1, 2, 3, and 4. This is seen in Figure 5 by black areas. The mean correlation over these 16 pairs of time points is -.323. The technique presented herein is not dependent on any particular covariance structure; its only requirement is that correlations close in time (with a small lag between them) be relevant for your objectives.

A Matlab function that performs local decorrelation

Here, we present the `Local_Decorrelation` function implemented in Matlab, which can be found in the supplementary materials folder ThirdIllustration. For the demonstration, data from Potvin-Pilon (2024) are being used. These data are composed of EEG signals sampled at a high rate, so that the time series is long (330 time points). In that study, we recorded participants' brain electrophysiological activity in a Same-Different task where they must indicate if two strings of consonants of various lengths, presented in close succession, are identical or not. We recorded events-related potentials (ERP) across 62 electrodes as we manipulated the length of the strings and the number of differences between them. For brevity, only two conditions (Same and 1 difference for strings composed of a single letter) and only one electrode (PO8) is examined and visualized here (see Potvin-Pilon, 2024, for preliminary analyses).

The variables `EEG_SignalSame` and `EEG_SignalDiff` are 24×330 matrices containing data from 24 participants, across 330 time points for the two conditions (Same and 1 difference). The variable timepoints contains the 330 sample moments, ranging from -0.1 to 1.0 second (i. e., there is one time point every 3.33 ms) and is used to label the horizontal axis. The variable radius represents the weights' standard deviation applied to the correlation' coefficients, that is, the radius. A small radius gives more weight to coefficients of close time points. In the present data, contiguous measurements are more strongly correlated than distant measurements (average correlations in the Same and Different conditions $\bar{r}$ are .293 and .379 respectively; locally-weighted correlations $\bar{r}_{LD}$ are .818 and .849 using a radius of 5 and .707 and .756 using a radius of 10). Consequently, using a small radius leads to greater adjustment and narrower intervals. As the purpose is to compare the points, an adjustment for pairwise difference is also requested with the argument 'Adjustment' set to true (resulting in intervals being $\sqrt{2}$ wider).

The Matlab lines of code in Listing 1 at the end show how to utilize the `Local_Decorrelation` function. Plots produced from these are shown in Figure 6.

Figure 6, top row, shows the figures obtained with two different radius values (5 and 10 in the left and right panels respectively). In comparison, the bottom panel shows the procedure without the local decorrelation adjustment. As seen, confidence intervals with locally-decorrelated adjustments are narrower. Considering the elevated amount of noise in EEG datasets, the locally decorrelated adjustment may give a better representation of the differences between time points and between conditions.

Discussion

We described a method which computes confidence intervals that capitalizes on correlations of nearby measurements. The method, herein called the local decorrelation technique, is adequate when the correlations across measurements are not stationary (not a compound symmetric covariance structure) but instead becoming less and less relevant as the lag between measurements increases. When used in plots, its benefit is to favor comparisons of nearby points either on the same curve or on adjacent curves. It can likewise be integrated into a paired t-test which uses a locally estimated correlation across the whole dataset instead of the correlation of the two measurements to be compared.

The degrees of freedom are given in Eq. 3 as $n - 1$. This is used for simplicity. However, the pooled degrees of freedom $(p - 1)(n - 1)$ where p is the number of time point in the time series is more adequate (Cousineau et al., 2021; Zitzmann et al., in press). Keep in mind that for non-negligible sample sizes, the difference is immaterial.

This technique requires a radius parameter. In Figures 2, 3 and 4, a radius of 1 was used as there were relatively few time points (26, 26 and 19 respectively); in Figure 6, larger radii of 5 and 10 were used, as there were 330 time points in the sample. When correlation is influenced by lag, this choice makes a difference, and in the extreme, if a very large radius is used, the weights are nearly uniform and the locally-decorrelated confidence intervals then becomes identical to the (global) correlation-adjusted confidence intervals. In the first illustration, when the radius takes the following values, 1, 2, 5, 10, 100, 1000, the average r_{LD} take the values 0.729, 0.624, 0.432, 0.279, 0.164, 0.163, the last being equal to the mean correlation in the correlation matrix. At this time, we do not have any motivated rule to provide on how to set the radius parameter; we suggest as a preliminary rule of thumb to use a radius 20 to 50 times smaller than the number of time points.

One way to conceptualize this technique is in term of window averaging. Window averaging is commonly used in time series to smooth the mean scores by weighting partially the scores of the neighboring time points in the average of the current score. Herein, we do the same not at the level of the means, but at the level of the standard errors of the difference.

An analogy to understand this method is to consider that longer lags "erase" any dependencies there may be between measurements. Thus, we partition the error variance into error deduced from nearby points and error variance resulting from this "forgetting" in the relations between measurements.

In the supplementary material, we provide code to perform the plots with three different software, Mathematica (first illustrations), R (second illustration), and Matlab (third illustration).

Authors' note

This research was funded in part by the *Conseil pour la recherche en sciences naturelles et en génie du Canada*.

In a nutshell...

Locally-decorrelated confidence intervals are to be used when nearby time points are expected to be positively correlated in a significant way and where distant time points may be uncorrelated because the passage of time washes away correlation. The technique performs a smoothing of the standard errors using a Gaussian kernel. Such confidence intervals allow comparing nearby time points along a single curve or located on distinct curves. A t-test is proposed in which the influence of lag is discounted. When correlation is assumed constant across time points, prefer correlation-adjusted confidence intervals.

CrossMark click for updates

References

Baguley, T. (2012). Calculating and graphing within-subject confidence intervals for ANOVA. *Behavior Research Methods*, *44*, 158–175. doi: 10.3758/s13428-011-0123-7.

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*, *1*, 42–45. doi: 10.20982/tqmp.01.1.p042.

Cousineau, D. (2017). Varieties of confidence intervals. *Advances in Cognitive Psychology*, *13*, 140–155. doi: 10.5709/acp-0214-z.

Cousineau, D. (2019). Correlation-adjusted standard errors and confidence intervals for within-subject designs: A simple multiplicative approach. *The Quantitative Methods for Psychology*, *15*, 226–241. doi: 10.20982/tqmp.15.3.p226.

Cousineau, D., Goulet, M.-A., & Harding, B. (2021). Summary plots with adjusted error bars: The superb framework with an implementation in R. *Advances in Methods and Practices in Psychological Sciences*, *4*(3), 1–18. doi: 10.1177/25152459211035109.

Cousineau, D., & Goulet-Pelletier, J.-C. (2021). A study of confidence intervals for Cohen's dp in within-subject designs with new proposals. *The Quantitative Methods for Psychology*, *17*, 51–75. doi: 10.20982/tqmp.17.1.p051.

Field, A. (2009). *Discovering statistics using SPSS* (3rd). Sage.

Goldstein, H., & Healy, M. J. R. (1995). The graphical presentation of a collection of means. *Journal of the Royal Statistical Society A*, *158*, 175–177. doi: 10.2307/2983411.

Howell, D. C. (2010). *Statistical methods for psychology* (7th). Wadsworth.

Howell, G. T., & Lacroix, G. L. (2012). Decomposing interactions using GLM in combination with the COMPARE, LMATRIX and MMATRIX subcommands in SPSS. *Tutorials in Quantitative Methods for Psychology*, *8*(1), 1–22. doi: 10.20982/tqmp.08.1.p001.

Jané, M. B., Xiao, Q., Yeung, S. K., Ben-Shachar, M. S., Caldwell, A. R., Cousineau, D., Dunleavy, D. J., Elsherif, M., Johnson, B. T., Moreau, D., et al. (2024). Guide to effect sizes and confidence intervals. *OSF preprints*, *1*, 1–55. doi: 10.17605/osf.io/d8c4g.

Keselman, H. J., Algina, J., Kowalchuk, R. K., & Wolfinger, R. D. (1998). A comparison of two approaches for selecting covariance structures in the analysis of repeated measurements. *Communications in Statistics - Simulation and Computation*, *27*(3), 591–604. doi: 10.1080/03610919808813497.

Kincaid, C. (n.c.). *Guidelines for selecting the covariance structure in mixed model analysis* (Technical report No. 198-30). COMSYS Information Technology Services. Portage. https://support.sas.com/resources/papers/proceedings/proceedings/sugi30/198-30.pdf

Laurencelle, L., & Cousineau, D. (2023a). Analysis of frequency data: The ANOFA framework. *The Quantitative Methods for Psychology*, *19*, 173–193. doi: 10.20982/tqmp.19.2.p173.

Laurencelle, L., & Cousineau, D. (2023b). Analysis of proportions using arcsine transform with any experimental design. *Frontiers in Psychology*, *13*, 1045436–1045436. doi: 10.3389/fpsyg.2022.1045436.

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, *1*, 476–490. doi: 10.3758/BF03210951.

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology*, *4*, 61–64. doi: 10.20982/tqmp.04.2.p061.

Politzer-Ahles, S. (2017). An extension of within-subject confidence intervals to models with crossed random effects. *The Quantitative Methods for Psychology*, *13*(1), 75–94. doi: 10.20982/tqmp.13.1.p075.

Potvin-Pilon, A. (2024). *La familiarité de stimuli présentés successivement et le rôle de leurs différences dans l'activité enregistrée avec un EEG* [Undergraduate dissertation, Université d'Ottawa].

Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2013). *Using multivariate statistics* (6th). Pearson.

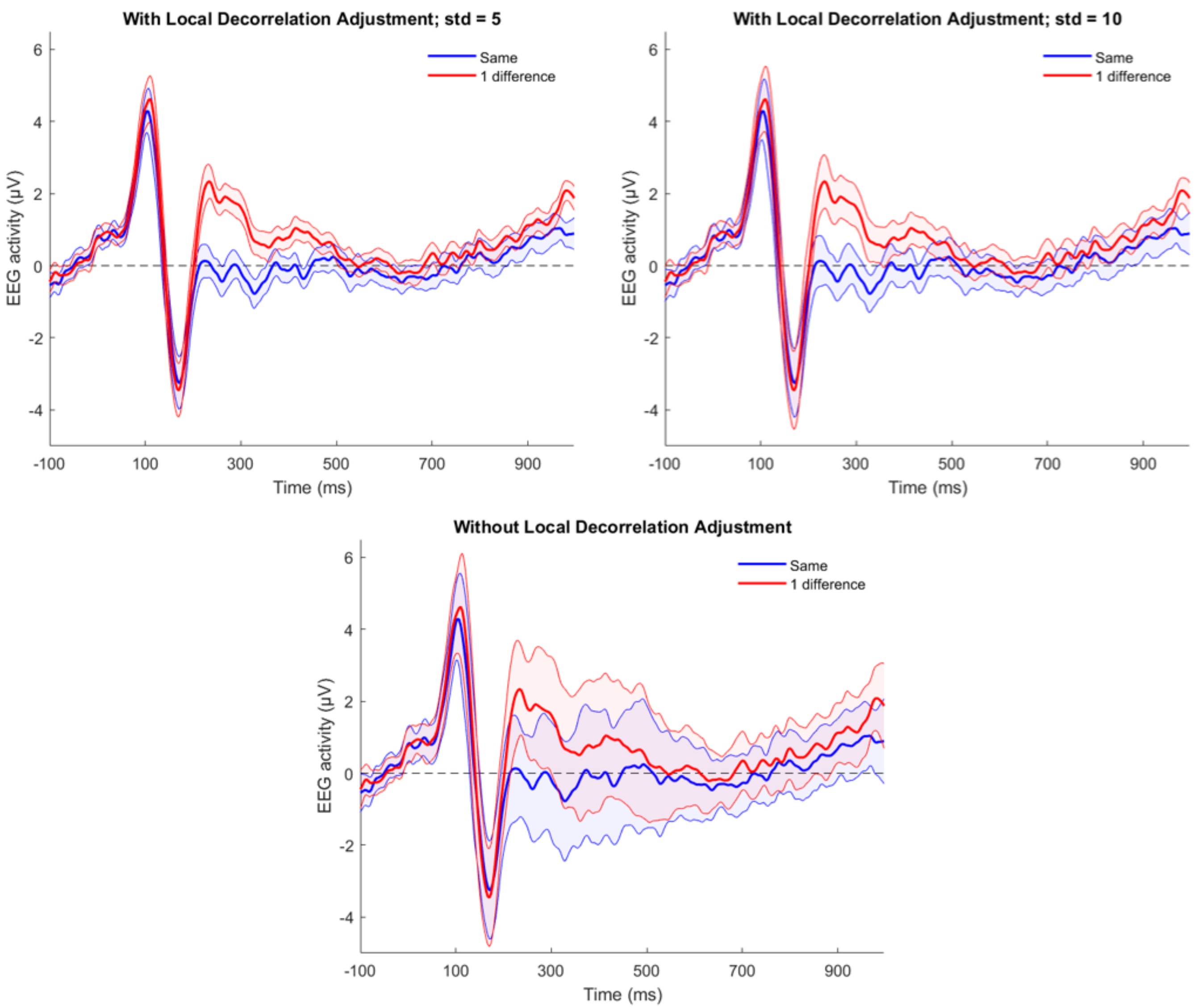
Waskom, M. L. (2021). seaborn: Statistical data visualization. *Journal of Open Source Software*, *6*(60), 3021. doi: 10.21105/joss.03021.

Waskom, M. L., Frank, M. C., & Wagner, A. D. (2017). Adaptive engagement of cognitive control in context-dependent decision-making. *Cerebral Cortex*, *27*(2), 1270–1284. doi: 10.1093/cercor/bhv333.

Winer, B. J., Brown, D. R., & Michels, K. M. (1991). *Statistical principles in experimental design*. McGraw-Hill.

Zitzmann, S., Lindner, C., & Hecht, M. (in press). A straightforward and valid correction to Nathoo et al.'s bayesian within-subject credible interval. *Journal of Mathematical Psychology*, 1–19.

Figure 6 ■ Plots showing EEG signals with confidence intervals from the Potvin-Pilon (2024) dataset. Top row received the Local Decorrelation adjustment with different radius values (5 and 10). Bottom row shows the confidence intervals without Local Decorrelation adjustment.



Listing 1 ■ Matlab code to perform the plot of Figure 6

```
% the time points on the x axis
timepoints = round([1:330]*10/3-100);

% read the two data files , one for Same, one for Different , when the total number of letter is 1
EEG_SignalSame = table2array(readtable("EEG_nTotal1_Same.tsv",'FileType','text'));
EEG_SignalDiff = table2array(readtable("EEG_nTotal1_Diff.tsv",'FileType','text'));

% desired radius
radius = 5;

% Calculate CI and mean for Same condition
[CI_same,mean_same] = Local_Decorrelation(EEG_SignalSame, radius, 'Adjustment',true
    );

% Calculate CI and mean for 1-difference condition
[CI_diff,mean_diff] = Local_Decorrelation(EEG_SignalDiff, radius, 'Adjustment',true
    );

% Only for filling between CI
fillX = [timepoints,fliplr(timepoints)];
fillY_same = [CI_same(1,:),fliplr(CI_same(2,:))];
fillY_diff = [CI_diff(1,:),fliplr(CI_diff(2,:))];

% Put a dashed line at 0
figure(1); hold on
plot(timepoints,zeros(1,length(timepoints)),'--black','HandleVisibility','off');

% Plot the Same condition
plot(timepoints,CI_same,'-b','LineWidth',0.5,'HandleVisibility','off'); % Plot the CI
plot(timepoints,mean_same,'-b','LineWidth',1.5,'HandleVisibility','on'); % Plot the
    mean
fill(fillX,fillY_same,'-b','FaceAlpha',0.05,'EdgeColor','none','HandleVisibility','
    off'); % Fill the space

% Plot the 1-difference condition
plot(timepoints,CI_diff,'-r','LineWidth',0.5,'HandleVisibility','off'); % Plot the CI
plot(timepoints,mean_diff,'-r','LineWidth',1.5,'HandleVisibility','on'); % Plot the
    mean
fill(fillX,fillY_diff,'-r','FaceAlpha',0.05,'EdgeColor','none','HandleVisibility','
    off'); % Fill the space

% From here, it is only for visualization preference
axis([timepoints(1) timepoints(end) -5.0 6.5]); % Adjust axis
xticks(timepoints(1):0.2:timepoints(end)); xticklabels(timepoints(1)*1000:200:
    timepoints(end)*1000);
box off; set(gcf,'color','w'); % Remove the box and make the background white
legend({'Same';'1 difference'},'box','off'); % Put the legend
xlabel('Time (ms)'); ylabel('EEG activity (μV)'); % Name axis labels
```

Open practices

The *Open Data* badge was earned because the data of the experiment(s) are available on the journal's web site.
The *Open Material* badge was earned because supplementary material(s) are available on the journal's web site.

Citation

Cousineau, D., Proulx, A., Potvin-Pilon, A., & Fiset, D. (2024). Local decorrelation for error bars in time series. *The Quantitative Methods for Psychology*, *20*(2), 173–185. doi: 10.20982/tqmp.20.2.p173.

Received: 06/07/2024