Localising the onset of face sensitivity in EEG time-series

Peter Hebden

total word count: 5000 max

Contents

Abstract

How soon after stimulus does the brain start to decode object categories? Estimates from electroencephalography (EEG) range from 50 to 150ms. Given the extremely large number of

time points in multi-electrode EEG recordings, it can be difficult to detect the earliest onset time without multiple comparisons and incurring a large number of false positives. Cluster-based statistics can be used to correct for multiple comparisons and estimate onsets as the first statistically significant time-point in a cluster. Such methods pool spatial and temporal structure in the data, but are computationally intensive due to their reliance on 1000 or more permutations for significance testing. Change point detection (CPD) algorithms such as Binary Segmentation (Binseg) require far less computation, and yet we found that conventional CPD methods such as Binseg can be as accurate and reliable as the spatio-temporal cluster-based and temporal cluster-based methods when tested on real EEG data (Bieniek et al., 2015) recorded from participants while they performed a face recognition task. We found that the key to high performance was to preprocess the recordings in a way that extracted most relevant features from all electrodes and trials over time for each participant to construct a *virtual electrode*. Furthermore, when this preprocessing was applied all participants per session the estimated onset time for each session was 72ms. This is substantially earlier than some previous estimates of ∼90ms.

1 Introduction

How soon after object presentation does the brain start to decode object categories, i.e. what is the onset time for the event related potential (ERP) for face sensitivity?

Event-related potentials (ERPs) are voltages generated in the brain structures in response to specific events or stimuli (Blackwood and Muir, 1990). They are EEG changes time locked to sensory, motor or cognitive events that can be used to study psychophysiological correlates of mental processes. ERPs can be evoked by a wide variety of sensory, cognitive or motor events. They reflect the sum of the postsynaptic potentials produced when similarly oriented cortical pyramidal neurons (thousands or millions) fire in synchrony while processing information (Peterson et al., 1995).

Estimated onset times for face sensitive ERPs based electroencephalography (EEG) data range from 50ms to 150ms (Rousselet et al., 2008; Jonas et al., 2014). This range is relatively wide because EEG data is noisy and recorded by many electrodes over many time points. The multiple comparison problem arises when massive univariate tests conduct the same test at every time point and sensor, i.e. hundreds or thousands of individual tests with significance thresholds (e.g., $p < 0.05$) are applied, resulting in an error rate that is far greater than the nominal rate (Groppe et al., 2011). Correction for multiple comparisons can be applied to provide error rate controls, but many of these methods reduce power (Groppe et al., 2011), i.e. greatly reduce the probability that a true effect will be detected given that one exists (Button et al., 2013). Consequently the challenge is to accurately identify the most relevant electrode(s) and time points in this high dimen-

sional data for accurate and reliable onset detection, i.e. to develop a detection method that is sensitive enough for high power and specific enough for a low false positive rate.

A popular approach is to use cluster-based statistics to correct for multiple comparisons and estimate onsets as the first statistically significant time-point in a cluster (Delorme and Makeig, 2004; Maris and Oostenveld, 2007; Oostenveld et al., 2011; Gramfort, 2013). By considering spatial and/or temporal clusters of activity, these methods can detect distributed patterns of neural activity that may not be apparent in individual sensor or time point analyses. This increased sensitivity allows the detection of subtle effects that may be missed using conventional statistical methods (Rousselet, 2023).

However, cluster-based methods are computationally intensive and may not be as reliable as expected. Results from a simulation by (Sassenhagen and Draschkow, 2019) suggest that estimated onsets are both positively biased and underestimate real onsets in a large proportion of simulated experiments. Consequently one must be cautious when interpreting the accuracy of cluster-based onset detections (Sassenhagen and Draschkow, 2019).

As Figure 1 indicates, many predicted onsets are far too early or far too late. This seems to suggest that spatio-temporal clustering, at least as implemented by (Sassenhagen and Draschkow, 2019), failed to extract the most predictive features in the dataset and that using permutations to increase sensitivity and specificity is not effective for data of this type.

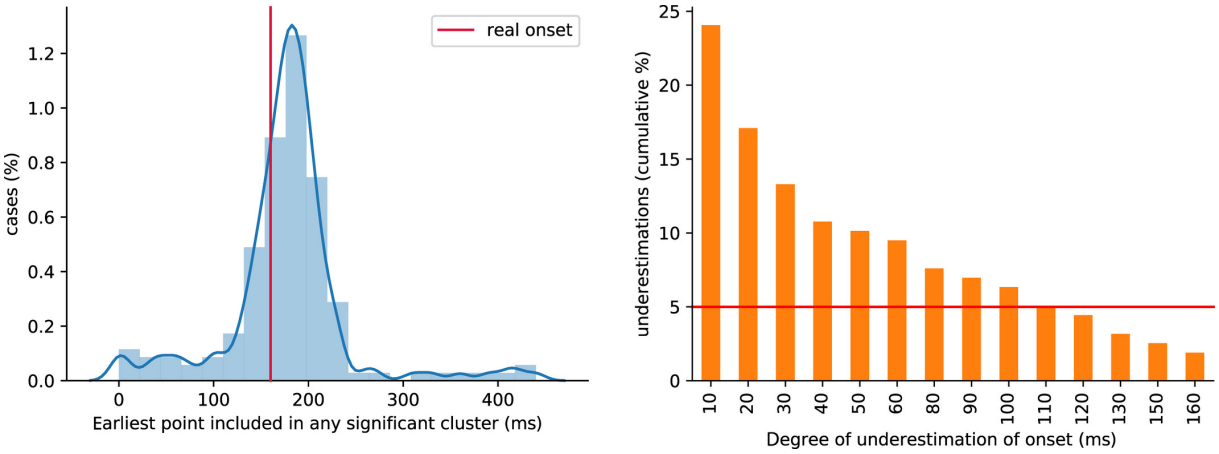


Figure 1: Results from their simulation show a wide distribution of predicted onset time, many far before or after the true onset time. Image: (Sassenhagen and Draschkow, 2019).

However, the study by (Sassenhagen and Draschkow, 2019) did not provide results from other methods for comparison (Rousselet, 2023). In light of their results and the need to compare the performance of cluster-based methods with other methods, this report compares the relative accuracy of ERP onsets predicted by the spatio-temporal cluster-based method (Bieniek et al., 2016), the temporal cluster-based method (Rousselet, 2023), and implementations of several well known change point detection algorithms, such as

Binary Segmentation Scott and Knott (1974) as suggested by (Rousselet, 2023), and the impact of data preprocessing on computational complexity and onset predictions.

We hypothesised that cluster-based methods will make earlier, more accurate and consistent onset predictions than non cluster-based methods due to their ability to exploit structure in data. To test this hypothesis we compared the performance of several non cluster-based algorithms, known as change point detection (CPD) algorithms, against each other and selected the best one for direct comparison with the spatio-temporal and temporal cluster-based methods. Given that there is no single best algorithm for all datasets (Wolpert and Macready, 1997), we also tested implementations of the other CPD algorithms.

We found that best conventional CPD algorithm performed as well as the temporal cluster-based method for predicting onset times for individual participants and for the group of participants in a session. Onset detection was greatly facilitated by data preprocessing technique that constructs a *virtual electrode* Rousselet et al. (2014).

Section 2 describes the datasets and our methods. Section 3 presents the results and Section 4 the discussion, while Section 5 presents some ideas for future work, and finally Section 6 our conclusion.

2 Data and Methods

2.1 Overview

First we generated four synthetic datasets using Python code to validate and optimise the conventional CPD algorithms implemented in the ruptures package (Truong et al., 2020). Then we compared the performance of the CPD methods with two cluster-based methods on a real EEG dataset (Bieniek et al., 2015).

We used the following software: Matlab 2024a, Spyder IDE 5.5.4 (Cordoba and et al., 2024), Python 3.12 and libraries, and the ruptures change point detection package (Truong et al., 2020), documentation Truong (2024).

The Python and Matlab code used to generate the figures in this report are available at `https://github.com/peter-426/MEG-EEG-onset`.

2.2 Synthetic data

We generated four synthetic datasets where the true onset time was known, as in Figure 2: (a) wavy time series with variable onset times and frequencies, and an abrupt change in mean amplitude, based on ruptures package, (Truong et al., 2020), (b) change of mean and variance with onset at 160, (c) flat baseline with sinusoidal onset at 160 plus a small Gaussian, (d) flat baseline with sinusoidal plus noise onset at 160 that was causal filtered.

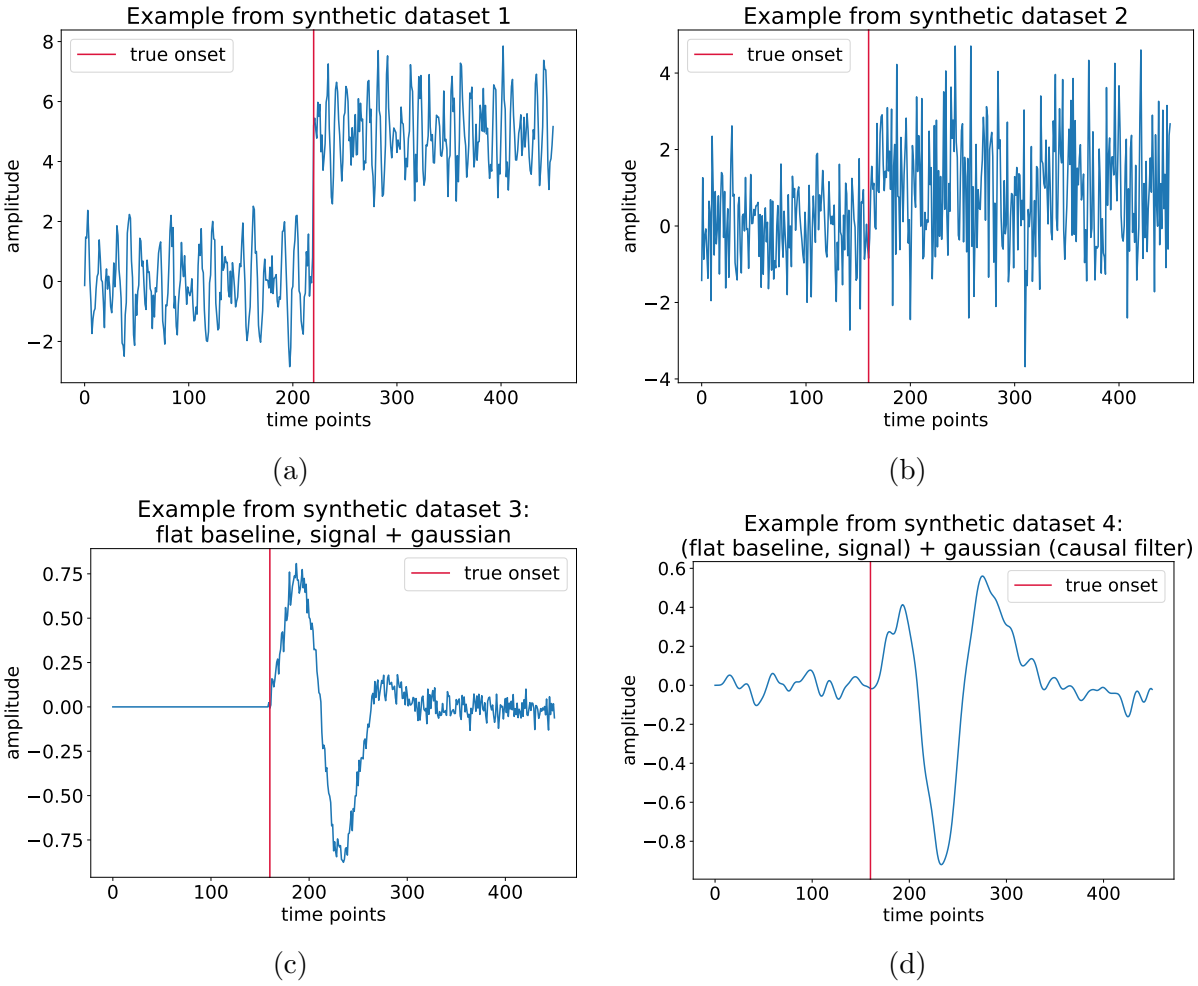
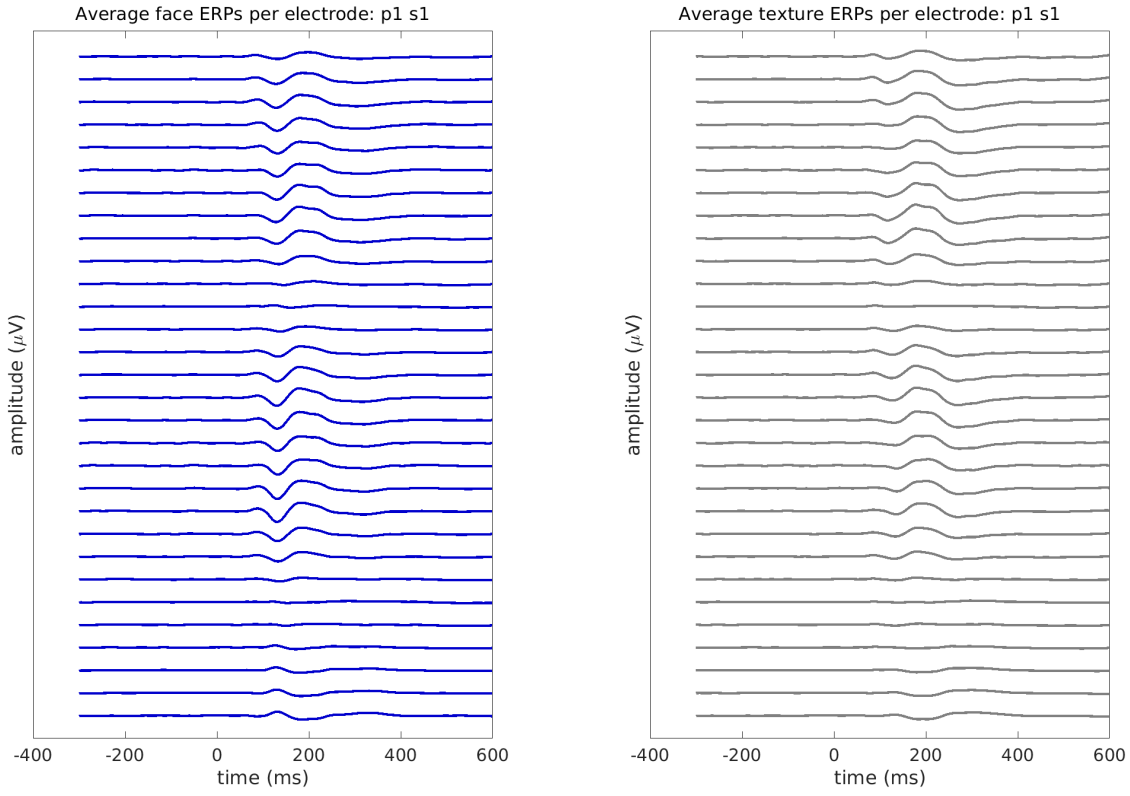


Figure 2: Synthetic dataset examples: (a) wavy time series with variable onset based on ruptures package, (Truong et al., 2020), (b) change of mean and variance with onset at 160, (c) flat baseline with sinusoidal onset at 160, (d) flat baseline with sinusoidal plus noise onset at 160 (causal filtered).

2.3 Real EEG data

Figure 3 shows the average ERPs face and texture ERPs recorded by 30 out of $\leq$ 128 electrodes from participant 1 during the first of two sessions. 75 participants did two sessions where they were shown face and texture images over $\sim$143 trials, generating a large high dmiensional dataset. This is a subset of the real EEG dataset from (Bieniek et al., 2015).



(a) (b)

Figure 3: Participant 1, session 1, face and texture (aka noise) ERPs: average ERPs from 30 out of $\leq$ 128 electrodes per participant are shown. Given up to 128 electrodes recording from 75 participants during two sessions, the aim is to accurately detect the onset of face sensitivity. This may be possible if the most informative electrodes are selected for each participant. Data from (Bieniek et al., 2015).

Stimuli used to generate the dataset (Bieniek et al., 2015) were grey-scale pictures of faces and textures, Figure 4. The ten faces are front view photographs that were cropped to remove hair and ears, then pasted on a grey background. The task was to discriminate between face and texture images (texture aka noise) presented for 104ms. Each participant was asked to categorize images of faces and textures as fast and accurately as possible by pressing 1 or 2 on a numerical keypad. The collected data subsequently used for onset detection was processed with a 2-Hz causal fourth-order Butterworth high-pass

filter (Bieniek et al., 2016).

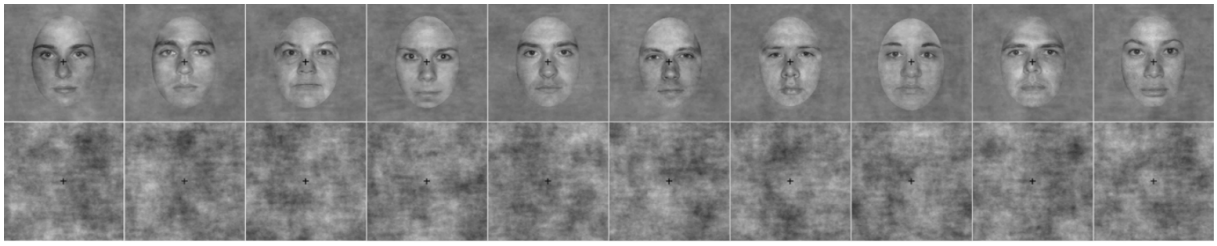


Figure 4: The face and texture images used during each trial. Image (Rousselet et al., 2014).

(Bieniek et al., 2016) compared mean with trimmed mean onsets using data without low-pass filtering (mean vs. tmean) and after application of a low-pass filter (mean lp vs. tmean lp). This was done to assess whether low-pass filtering, as commonly applied in ERP research, produces signal distortions leading to artificially earlier onsets. They found no significant differences between onset distributions in any of the comparisons, so their main analysis used mean lp data. Consequently, the analysis in this report is based on the onsets estimated from the same mean lp data (Bieniek et al., 2015).

2.4 Change point detection methods

The change point detection approach, such as exemplified by Binary segmentation algorithm (Scott and Knott, 1974), as suggested by (Rousselet, 2023), might provide better ERP onset estimation. For tis reason, we selected five change point detection (CPD) methods implemented by (Truong et al., 2020) for comparison with cluster-based method performance on real EEG data. Binary Segmentation (Scott and Knott, 1974) (Binseg), Pruned Exact Linear Time (Killick et al., 2012) (Pelt), Window-Based (Window), Dynamic Programming (Dynp), and BottomUp Segmentation (BottomUp) were optimised and tested.

2.5 Change point detection method optimisation

To prepare for the real EEG data competition, we evaluated the relative performance of five CPD algorithms implemented in the ruptures package (Truong et al., 2020) on the synthetic data and optimised their cost function parameters with gridsearch to identify their optimal cost functions for the synthetic data. Each algorithm was tested using three cost functions: `l1` (absolute distance), `l2` (squared distance), and `rbf` (radial basis function), except Binseg was also tested using the `normal` distribution cost function.

As show in Table 1, each algorithm was parameterised with minimum segment size `size`=2, `jump`=1, number of breakpoints `n_bkps`=1, and defaults otherwise except for Pelt where penalty `pen`=5.

Table 1: The CPD methods and their parameters: Python ruptures package (Truong et al., 2020)

	cost function	min segment size	jump	params	width
BinSeg	normal	2	1	None	NA
Pelt	rbf	2	1	None	NA
Window	l2	2	1	None	100
Dynp	rbf	2	1	None	NA
BottomUp	rbf	2	1	None	NA

2.6 Cluster-based methods

The spatio-temporal cluster-based code and results by (Sassenhagen and Draschkow, 2019) used MNE-python 0.17 and a collection of libraries with unknown version numbers; a container, e.g. Docker (Merkel, 2014) was not provided. The original Python code by (Sassenhagen and Draschkow, 2019) did not run correctly without the correct libraries when using MNE-python 0.17 or the current version MNE-python 1.70. Consequently it was not possible to reproduce their results reported in (Sassenhagen and Draschkow, 2019) for this report given the time constraint. Nonetheless, their study was the inspiration for this report .

The spatio-temporal cluster-based Bieniek et al. (2016) and temporal cluster-based methods (Rousselet, 2023) involve identifying clusters of spatially and/or temporally contiguous time points that show statistically significant effects and then assessing the significance of these clusters, e.g. whether or not the cluster-sum of each cluster is statistically significant.

1. **Compute t-values**: For each time point in the EEG data, compute a t-value to test the difference between conditions, face vs. texture (noise).
2. **Thresholding**: Apply a threshold to the t-values to identify significant time points. The threshold was set to $p < 0.05$, so only time points with t-values corresponding to $p < 0.05$ were considered significant.
3. **Form Clusters**: Identify contiguous time points that exceed the threshold and group them into clusters. Contiguity means consecutive time points.
4. **Compute Cluster-Sum**: For each cluster, calculate the sum of the t-values within the cluster. This is the cluster-level statistic.
5. **Permutation Testing**: To determine the significance of the observed clusters, perform a permutation test. This involves: (a) Randomly shuffling the data labels (e.g., condition labels) many times (typically thousands). (b) Recomputing the t-values and forming clusters for each permutation. (c) Calculating the cluster-sum for each permuted dataset to create a distribution of cluster-sums under the null hypothesis.

6. **Assess Significance**: Compare the observed cluster-sums to the permutation distribution. The p-value for a cluster is the proportion of permuted cluster-sums that are greater than or equal to the observed cluster-sum. Clusters with p-values below the chosen alpha level ($p < 0.05$) are considered statistically significant.

2.7 Virtual electrodes for feature extraction

The real EEG dataset was used to create one "virtual electrode" from ≤ 128 electrodes per participant, as in (Rousselet et al., 2014; Rousselet, 2023) where a virtual electrodes were constructed from maximum t^2 values.

The steps for constructing each participant's virtual electrode were (1) compute t-tests across trials at every electrode and time point, (2) save the max t^2 values across electrodes to form the virtual electrode. The virtual electrode for participant 1, session 1 is shown in Figure 5(d).

For the cluster-based (cluster-sum) methods, a permutation test was applied: shuffle trials between conditions and compute the max t^2 values as above. Use the permutation distribution to derive a cluster-forming threshold, i.e. the 95th percentile, which was then applied to the permutation distributions and the original data. A cluster of max t2 values is statistically significant if the sum of its t^2 values greater than threshold. The first time point in the first such cluster is the estimated ERP onset for a participant in a session.

Figure 6 illustrates the steps for constructing one max t^2 virtual electrode for one session for all participants were (1) compute the average max t^2 from the max t^2 time series from all participants in a session, (2) for each participant compute the average face and texture ERPs from their max t^2 electrode only, (3) use permutations of the 75 pairs of face and texture ERPs to test the statistical significance of each t^2 value and clusters of statistically significant adjacent t^2 values in the session average max t^2 time series (the virtual electrode for one session). The first time point in the first such cluster is the estimated ERP onset for the session.

2.8 Methods for assessing performance

2.8.1 Assessment of CPD methods on synthetic data

CPD algorithms were assessed and optimised using synthetic data with known true onset times. Mean absolute error (MAE), Equation 1, was the primary metric for assessing performance.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|\hat{Y}_i - Y_i| \tag{1}$$

where $\hat{Y}$ is the predicted onset time and Y is the true onset time.

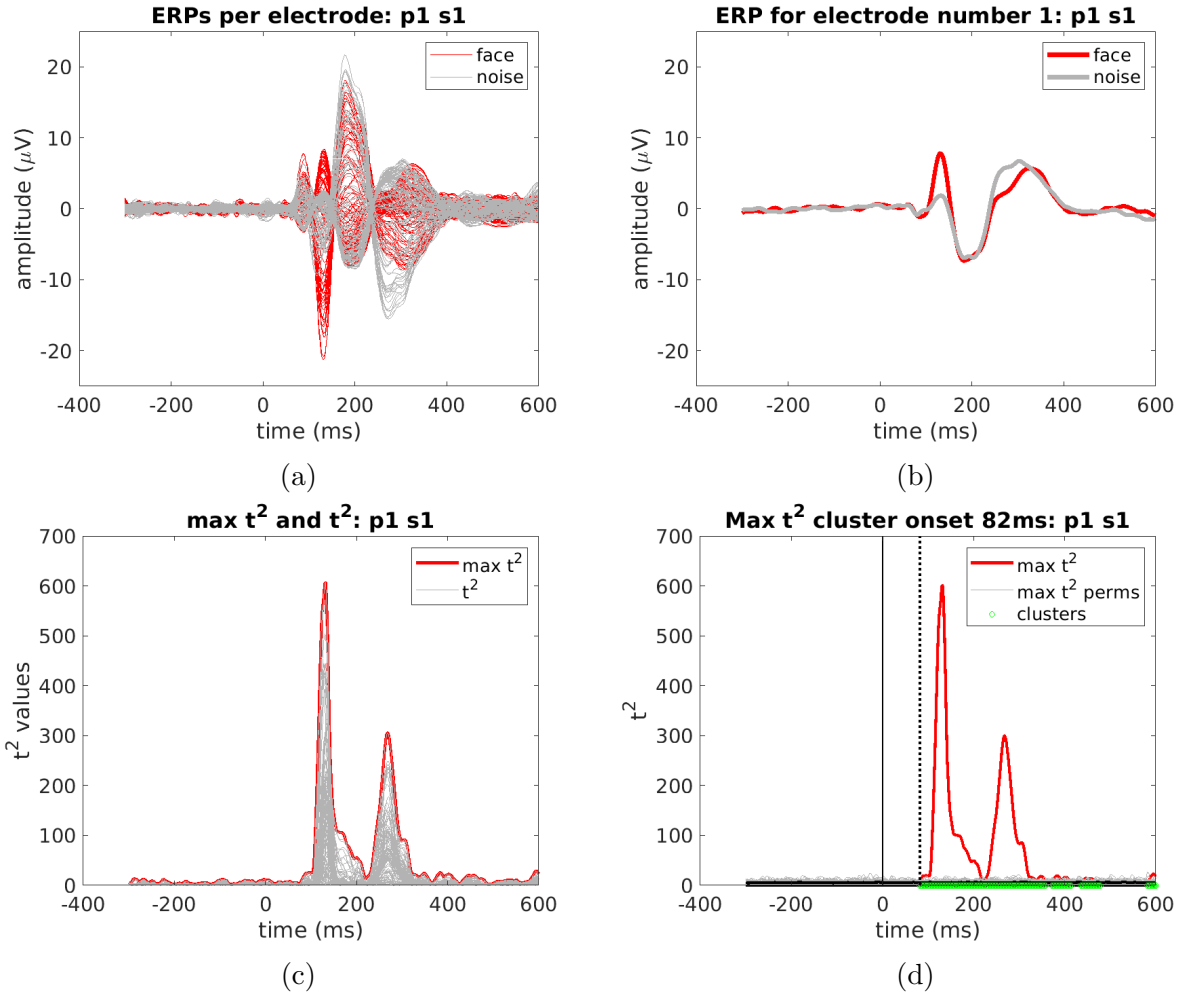


Figure 5: Participant 1 session 1: (a) average ERP for all electrodes, (b) average ERP for electrode number 1 (name A1) has a relatively small amplitude, the onset and positive peak are followed by a negative peak (N170) as expected, (c) max t^2 for each time point (red), and t^2 time series for all electrodes in grey, (d) max t^2 for each time point (red), permutations of max t^2 (grey/black), and clusters of significant time points (green). Temporal cluster-based methods predicted onset time (vertical dashed line) is at the earliest time point in the first statistically significant cluster. Data from (Bieniek et al., 2015).

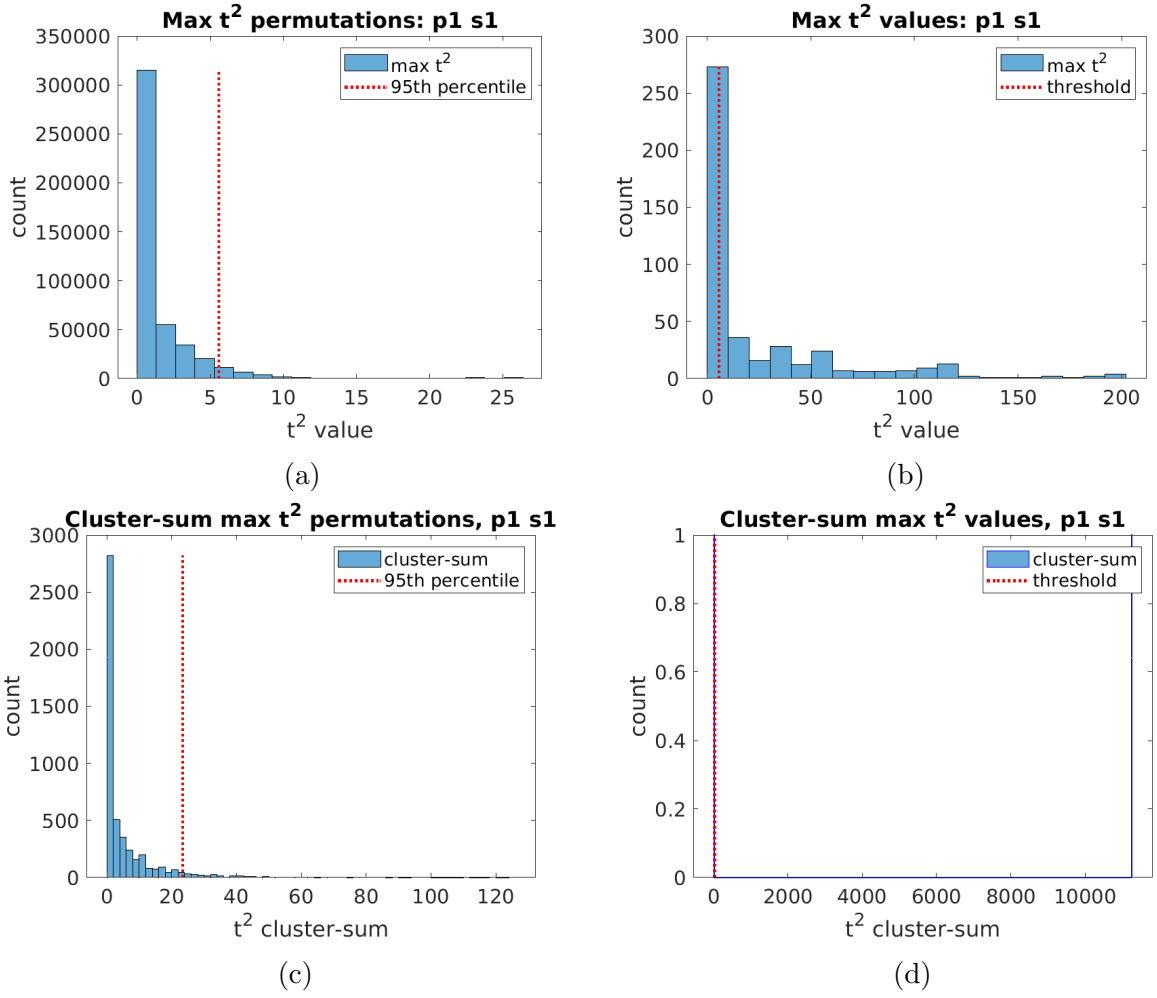


Figure 6: The cluster-sum method: (a) distribution of max t^2 values after 1000 permutations of face and texture data for participant 1, sessions 1, (b) the max t^2 values for this participant, many t^2 values are greater than the threshold, (c) distribution of cluster-sums for the permuted data, (d) only two clusters of significant t^2 values, the cluster-sum on the right is extremely large, far exceeding the threshold of ∼102.5, first time point in that cluster is 78ms. Data from (Bieniek et al., 2015).

We also considered root mean squared error (RMSE), Equation 2, bias (early or late), Equation 3, and percent too early.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i - Y_i)^2} \tag{2}$$

$$\text{bias} = \frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i - Y_i) \tag{3}$$

2.8.2 Assessment of CPD and cluster-based methods on real EEG data

Cluster-based and CPD methods were applied to the real EEG time-series dataset (Bieniek et al., 2015) with unknown onset times. The estimated onset times for participants that did two sessions were analysed. Metrics for evaluation were mean absolute difference (MAD), mean difference (bias) and its 95% confidence interval.

We evaluated the reliability of each methods in terms of its accuracy and the consistency of its onset predictions across the two sessions, i.e. the predicted onset for each participant was expected to be near the first abrupt change in the max t^2 time series and very similar for both sessions. We used the following metrics.

Mean absolute difference, Equation 4, is the average of the absolute differences between the measurements in the two lists.

$$\text{MAD} = \frac{1}{n}\sum_{i=1}^{n}|X_i - Y_i| \tag{4}$$

Mean difference (bias), Equation 5, measures the average difference between two lists.

$$\text{bias} = \frac{1}{n}\sum_{i=1}^{n}(X_i - Y_i) \tag{5}$$

Confidence interval for bias, Equation 6, provides a range likely to contain the true mean difference

$$\text{CI} = \text{bias} \pm t_{\alpha/2,\,n-1} \cdot \frac{s}{\sqrt{n}} \tag{6}$$

where s is the standard deviation of the differences D_i, n is the number of measurements, and $t_{\alpha/2,\,n-1}$ is the critical value from the t-distribution with $n-1$ degrees of freedom for the desired confidence level (e.g., 1.96 for 95% confidence).

3 Results

3.1 Synthetic data study

Five CPD methods as implemented by (Truong et al., 2020) were validated and optimised using synthetic data. We found that Binseg with a normal cost function had the best performance overall and would be the competition for the two cluster-based methods.

We noted that Pelt is relatively slow. Running time for 100 waveforms with 451 time points each was about 0.15 minutes for Binseg, BottomUp, Window, and Dynp compared to 5.40 minutes for Pelt when each algorithm was parameterised by three cost functions during gridsearch: `l1`, `l2`, and `rbf`. Pelt's time complexity may make it unsuitable for applications where a very large number of epochs must be processed within a hard time constraint.

As indicated by Tables 2 to 5, Binseg with a normal cost function usually had the best performance according to the MAE metric and best overall. So Binseg was selected to compete directly with the cluster-based methods when they are applied to the real EEG dataset (Bieniek et al., 2015). Essentially, this optimisation step was intended to validate the CPD methods and get some indication of what their relative performance on real EEG data might be while bearing in mind that there is no single "best" algorithm for all datasets (Wolpert and Macready, 1997).

Figure 7 shows the distributions of onset predictions made by the five CPD methods on the four synthetic datasets. Binseg onset predictions tend to be the most consistent, with Pelt taking second place.

Table 2: Synthetic dataset 1: CPD algorithms optimised for mean absolute error (MAE)

	cost function	MAE	RMSE	bias	pct too early
BinSeg	rbf	3.94	5.27	0.92	43.0
Pelt	rbf	5.55	7.0	4.53	16.0
Window	l1	4.42	5.79	1.06	41.0
Dynp	l2	4.21	5.09	0.19	46.0
BottomUp	rbf	4.22	5.11	-1.78	58.0

Table 3: Synthetic dataset 2: CPD algorithms optimised for mean absolute error (MAE)

	cost function	MAE	RMSE	bias	pct too early
BinSeg	normal	2.9	4.66	0.96	28.0
Pelt	rbf	4.4	6.55	0.6	51.0
Window	l2	6.17	9.17	1.13	47.0
Dynp	rbf	4.4	6.55	0.6	51.0
BottomUp	rbf	7.2	15.96	2.72	59.0

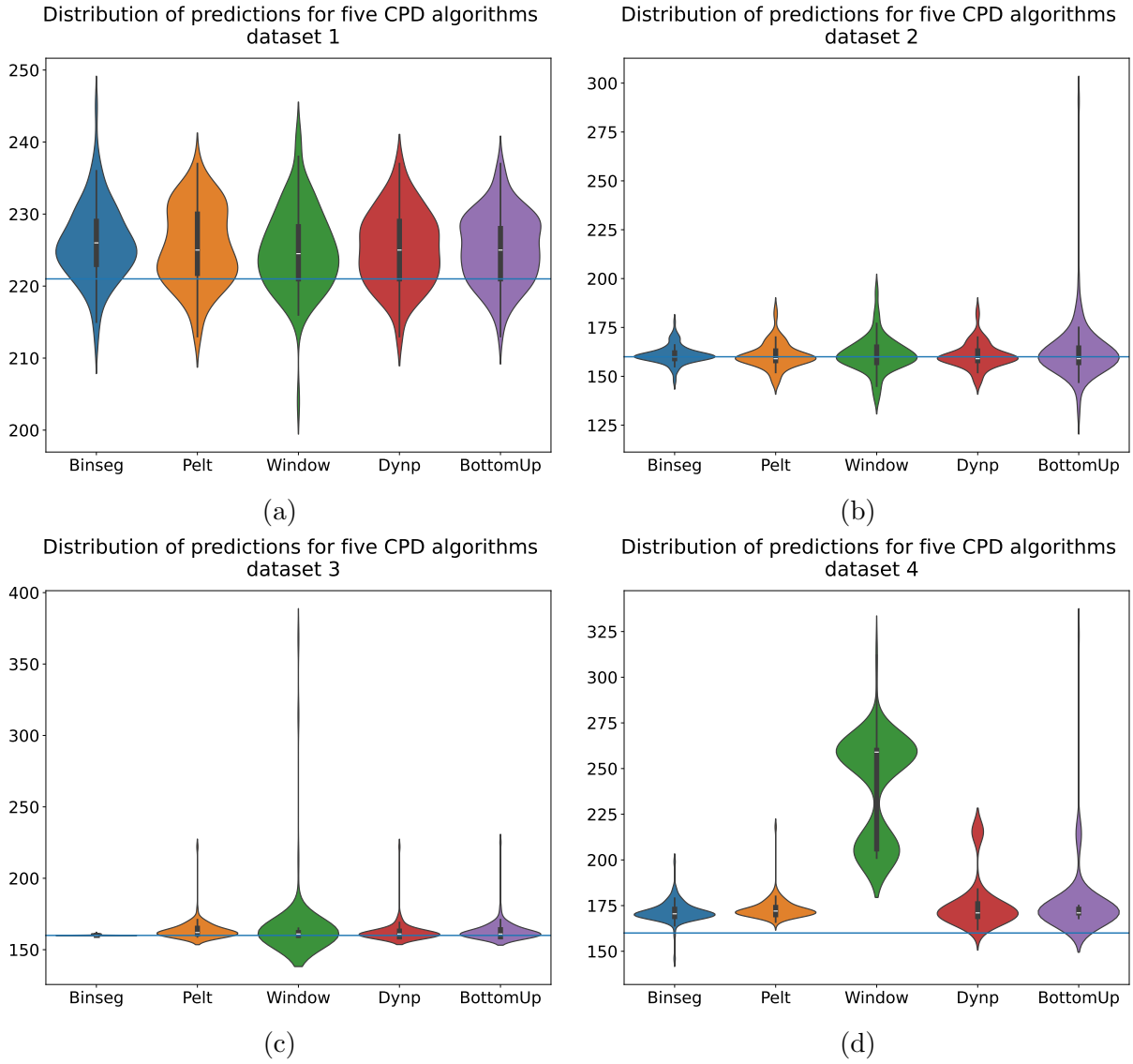


Figure 7: The distribution of onset detection by the five CPD methods on four synthetic datasets.

Table 4: Synthetic dataset 3: CPD algorithms optimised for mean absolute error (MAE)

	cost function	MAE	RMSE	bias	pct too early
BinSeg	normal	0.13	0.39	0.13	0.0
Pelt	rbf	3.77	7.72	3.51	13.0
Window	rbf	6.1	26.64	5.7	20.0
Dynp	rbf	2.77	7.14	2.25	26.0
BottomUp	rbf	3.52	7.9	3.0	26.0

Table 5: Synthetic dataset 4: CPD algorithms optimised for mean absolute error (MAE)

	cost function	MAE	RMSE	bias	pct too early
BinSeg	normal	11.9	12.83	11.6	2.0
Pelt	rbf	13.12	14.28	13.12	0.0
Window	l2	80.45	84.85	80.45	0.0
Dynp	rbf	16.32	21.67	16.32	0.0
BottomUp	rbf	15.81	24.11	15.81	0.0

3.2 Real EEG data study

The CPD methods were validated and optimised using synthetic datasets. The best performing CPD method was then directly compared with the cluster-based methods using one real EEG dataset, i.e. a large dataset of face and texture ERPs from (Bieniek et al., 2015). This dataset includes recordings from 74 (Bieniek et al., 2016) or 75 (Rousselet, 2023) participants who were tested during two sessions. This data was used to quantify test-retest reliability of predicted face sensitivity onsets using different onset detection methods.

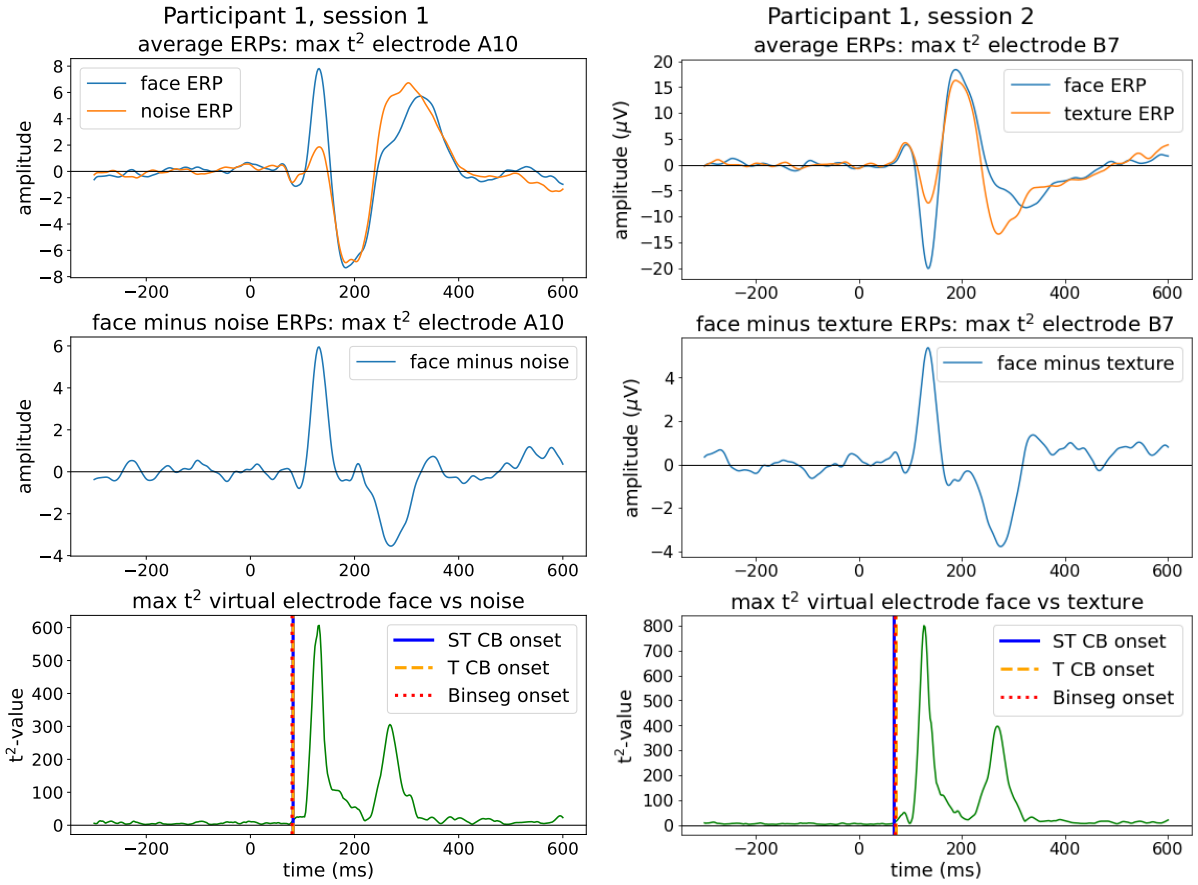
A spatio-temporal cluster-based method was applied in (Bieniek et al., 2016) to the face and texture dataset (Bieniek et al., 2015). The temporal cluster-based method combined with the virtual electrode technique (one per participant) was applied to the same dataset in (Rousselet, 2023)

For the real EEG study the virtual electrode technique was used at two levels. First at the level of individual participants per sessions, e.g. 75x2 virtual electrodes were constructed. Second at the level of sessions where the average ERP from the max t^2 electrode for each participant was combined to construct one virtual electrode per session.

3.2.1 One virtual electrode per participant

As show in Figure 8 (row 3), the spatio-temporal cluster-based (Bieniek et al., 2016), temporal cluster-based (Rousselet, 2023), and Binseg method (Truong et al., 2020) predicted the same onset time for participant 1 for both sessions. In contrast, Figure 9 shows results for another participant where that data is more irregular and the predictions (row 3) for the three methods are not identical. There is substantial variability in the average ERPs across participants and sessions, probably due to noise and artefacts, which could

be reduced or averaged out when, for example, a single virtual electrode is constructed for all participants per sessions.



(a) session 1 (b) session 2

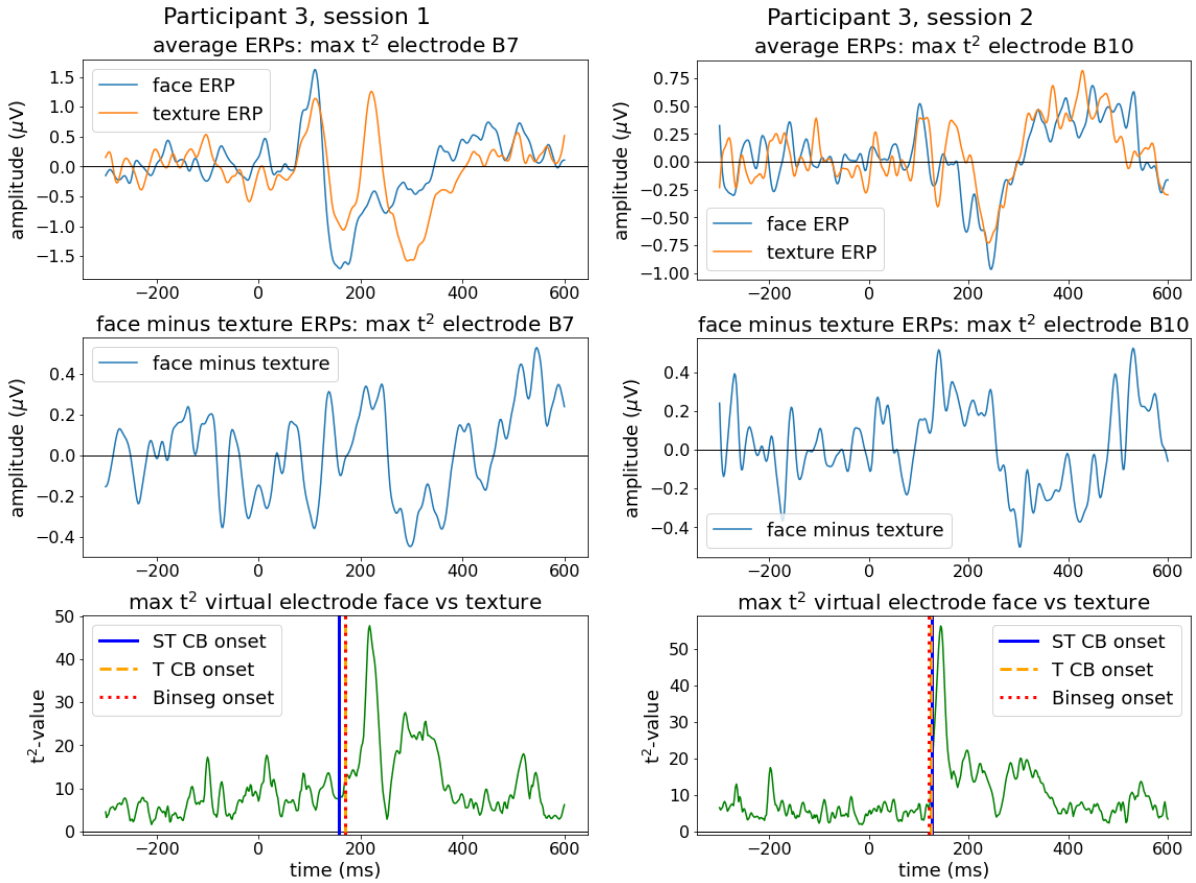
Figure 8: Participant 1, sessions 1 and 2: (row 1) face and texture ERPs averaged over all trials per electrode, (row 2) face minus texture ERPs, (row 3) onsets predicted by the spatio-temporal cluster-based (ST CB) and temporal cluster-based (T CB) methods, and the Binseg method from the t^2 max time series.

Table 6 indicates that the temporal cluster-based method had the best test-retest reliability in terms of mean squared difference (MSD), bias (Equation 5), and width of confidence interval, with Binseg taking second place. However, the confidence interval includes zero in both cases.

Table 7 indicates that Binseg had the lowest variance, with temporal cluster-based taking second place. However, the difference is obviously very small.

Table 6: Test-retest reliability across two sessions: cluster-based and Binseg CPD

	MSD	bias	95% CI lower	95% CI upper
spatio-temporal CB	364.65	3.46	-0.89	7.81
temporal CB	261.28	1.20	-2.53	4.93
Binseg	306.40	2.00	-2.03	6.03



(a) session 1 (b) session 2

Figure 9: Participant 3, sessions 1 and 2: (row 1) face and texture ERPs averaged over all trials per electrode, (row 2) face minus texture ERPs, (row 3) onsets predicted by the spatio-temporal cluster-based (ST CB) and temporal cluster-based (T CB) methods, and Binseg from the t^2 max time series.

Table 7: Sample variance across two sessions and confidence interval

	variance	95% CI lower	95% CI upper
spatio-temporal CB	466.267	375.653	594.325
temporal CB	357.342	288.293	454.696
Binseg	351.511	283.589	447.276

Figure 10 shows the case where one virtual electrode was constructed for each participant per session. Scatter plots (a)-(c) show the onsets predicted by the three methods for sessions 1 and 2, while scatter plot (d) shows a dense cluster of overlapping predicted onsets near 80ms. The distribution of onsets appears to be similar and does not show an obvious bias or pattern between the sessions or the methods.

Figure 11 shows kernel density plots of predicted onsets for both sessions where the spatio-temporal CB method peaks at 76.7ms, temporal CB at 78.2ms, and Binseg at 80.0ms. This was based on one virtual electrode per participant, 74 or 75 virtual electrodes per session.

3.2.2 One virtual electrode per session

Figure 12 shows the case where one max t^2 virtual electrode was constructed per session from the average ERP of the max t^2 electrode for each participant. The max t^2 electrodes are assumed to be the most relevant. Row 1 shows the average max t^2 ERPs for all participants in each session. Row 2 shows the difference of these face and texture ERPs with one standard deviation error band. Row 3 shows the Binseg onset predictions where Binseg was applied to the resulting max t^2 time series. Figures 12(a) and (b) are very similar but not identical.

Figure 13(a) shows a scatter plot of predicted onset times for session 1 versus session 2 based on one virtual electrode per session. Figure 13(b) shows that Binseg predicted an onset at 72ms for each session. The other CPD methods predicted onset times in the 78 to 82ms range, except for Window which predicted 90 and 98ms for sessions 1 and 2 respectively. The temporal CB method predicted 76ms for both session when allowed 1000 permutations. This earlier detection time by Binseg may be partly explained by the significance threshold α=0.05 used by the cluster-based methods, no such limit applies to Binseg predicted onsets.

It should be kept in mind that different onsets may be predicted with small changes to the CPD method parameters, and the random number generator (and seed value) used by the cluster-based methods.

3.2.3 Max t^2 electrode locations

Figure 14 shows the max t^2 counts for each electrode with a count of ≥ 1. Electrode B8 has by far the highest count with 33 out of 150 possible.

Figure 15 shows a Biosemi 128 electrode map. PO7 and PO8 electrodes, located over the parieto-occipital areas, often record robust ERPs in response to face stimuli (Bentin and Deouell, 2000). In this study, the max t^2 electrode with the highest count, B8, is highlighted in red. This location is consistent with the literature, further validating the virtual electrode technique.

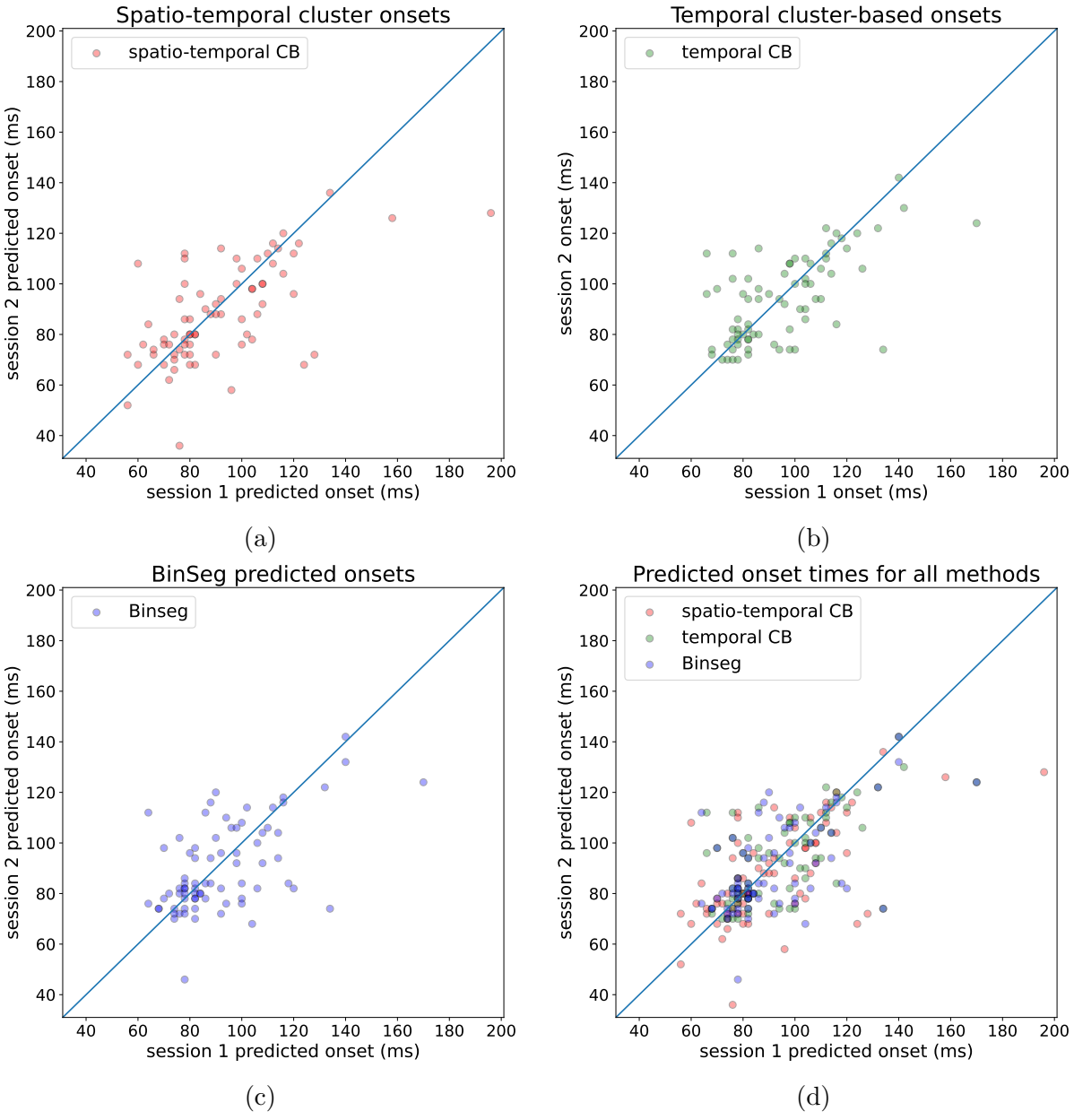


Figure 10: t^2 analysis, sessions 1 & 2: (a) spatio-temporal cluster-based predicted onsets, (b) temporal cluster-based predicted onsets, (c) Binseg predicted onsets, (d) predicted onsets superimposed. Data from (Bieniek et al., 2015).

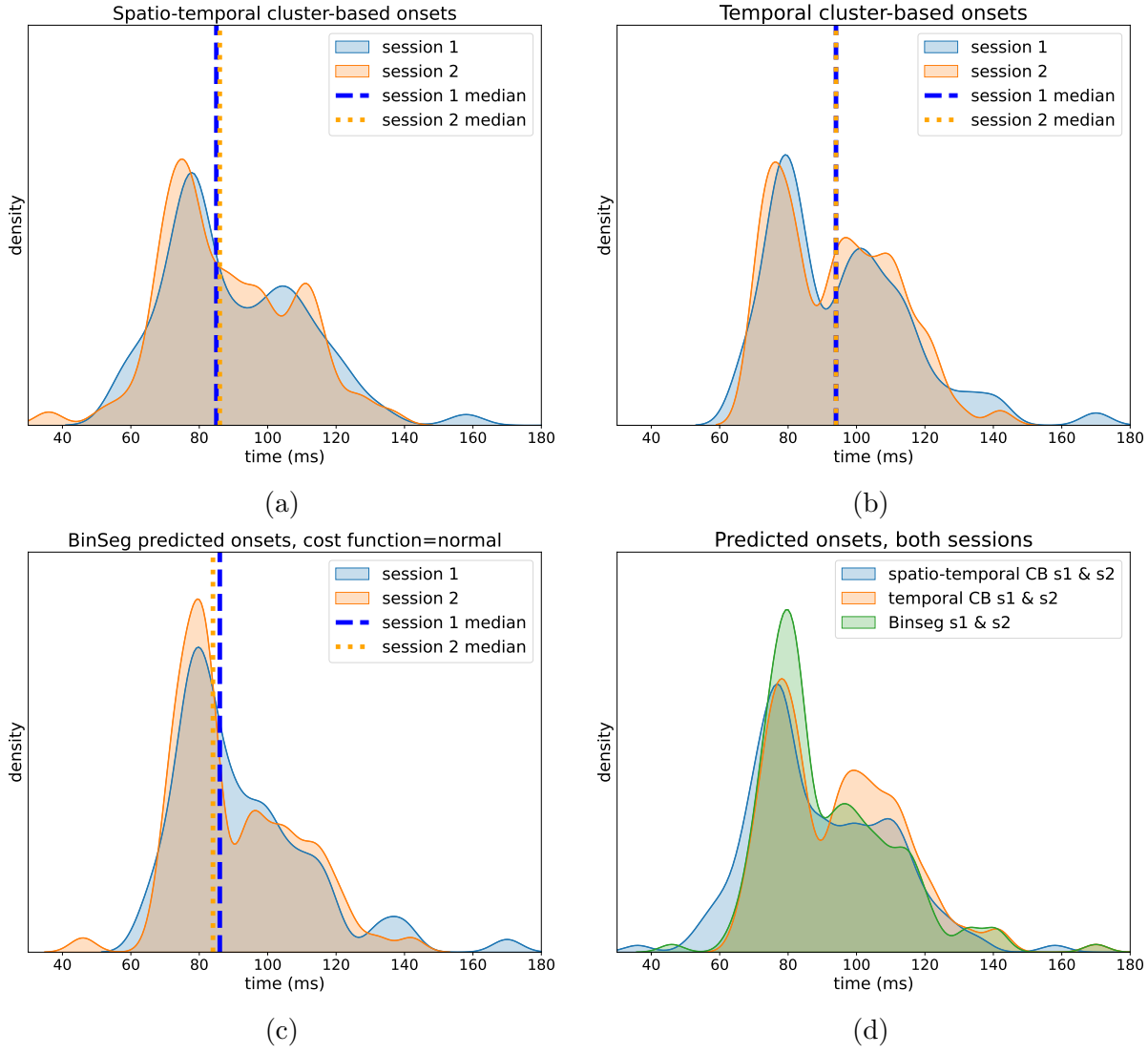
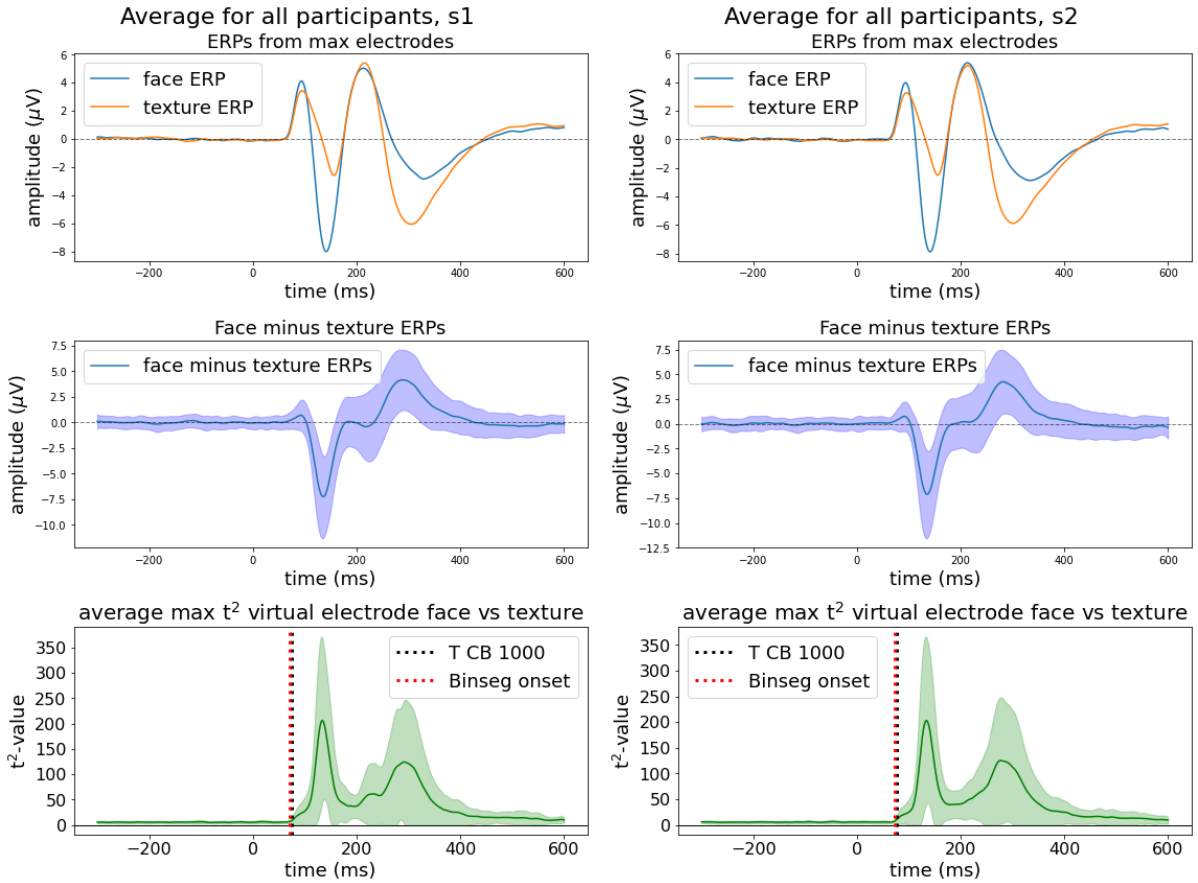


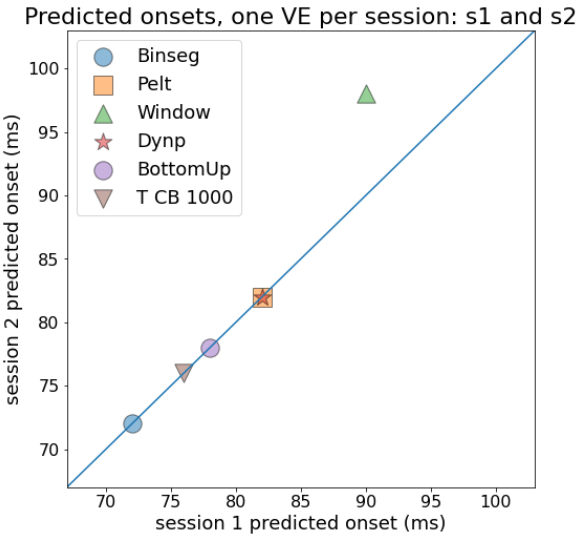
Figure 11: t^2 analysis, sessions 1 & 2: (a) spatio-temporal cluster-based onsets, (b) temporal cluster-based onsets, (c) Binseg onsets.



(a) session 1

(b) session 2

Figure 12: All participants, sessions 1 and 2: (row 1) average of all max electrode face and texture ERPs, (row 2) face minus texture, (row 3) onsets predicted by the Binseg algorithm from face (F) minus texture (N) and the average max t^2 time series. Binseg predicted a 70ms onset for face ERP sensitivity face from the face minus texture difference (row 2), and from the average max t^2 time series (row 3) for both sessions.



(a)

method	cost fx	s1 (ms)	s2 (ms)
BinSeg	normal	72	72
Pelt	rbf	82	82
Window	l2	90	98
Dynp	rbf	82	82
BottomUp	rbf	78	78
T CB 2		70	70
T CB 100		76	74
T CB 1000		76	76

(b)

Figure 13: One virtual electrode (VE) for all participants per session. (a) Onsets were predicted based on the two max t^2 time series derived from the two virtual electrodes. CPD algorithms used the parameters that were identified during validation with synthetic data. The temporal cluster-based cluster-sum (T CB) method was applied to the max t^2 waveforms, using 2, 100, and 1000 permutations to detect the face sensitivity onset time. (b) Binseg predicted the earliest onset times, while T CB predicted similar times but required many permutations to predict the same onset time for both sessions.

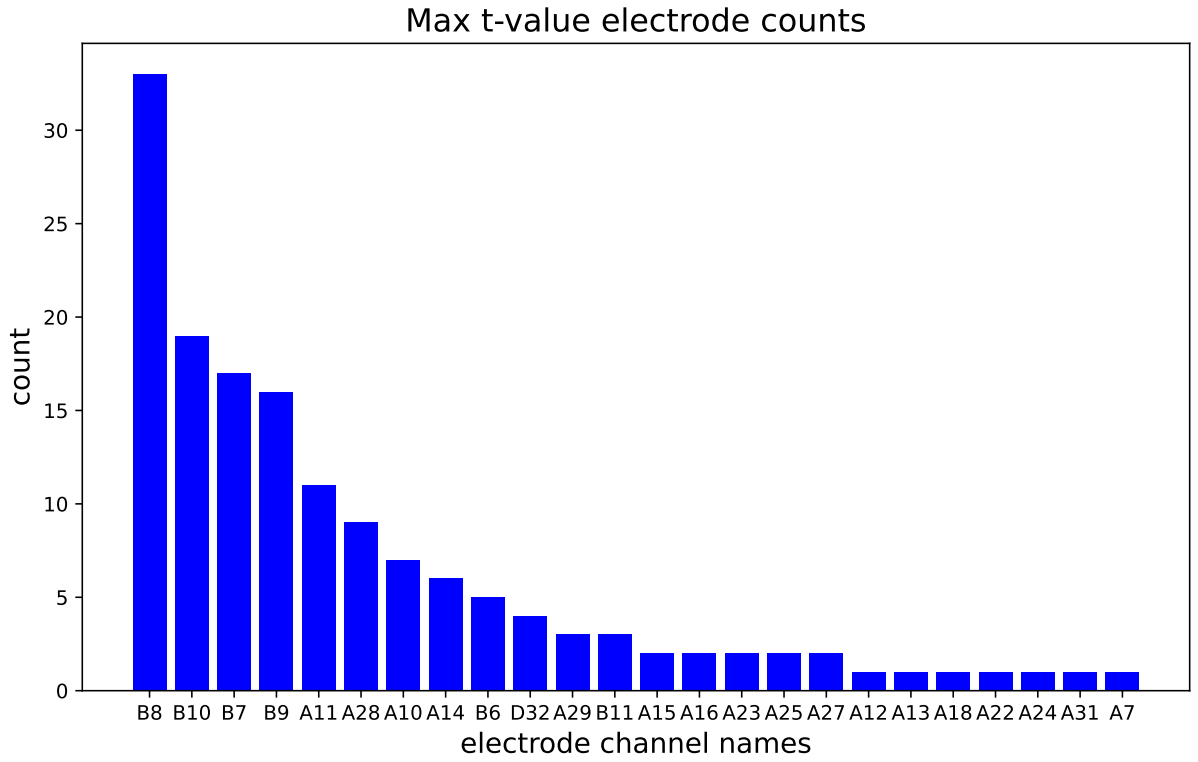


Figure 14: Max t-val electrode channel counts. 75 participants, 2 sessions each, total count 150.

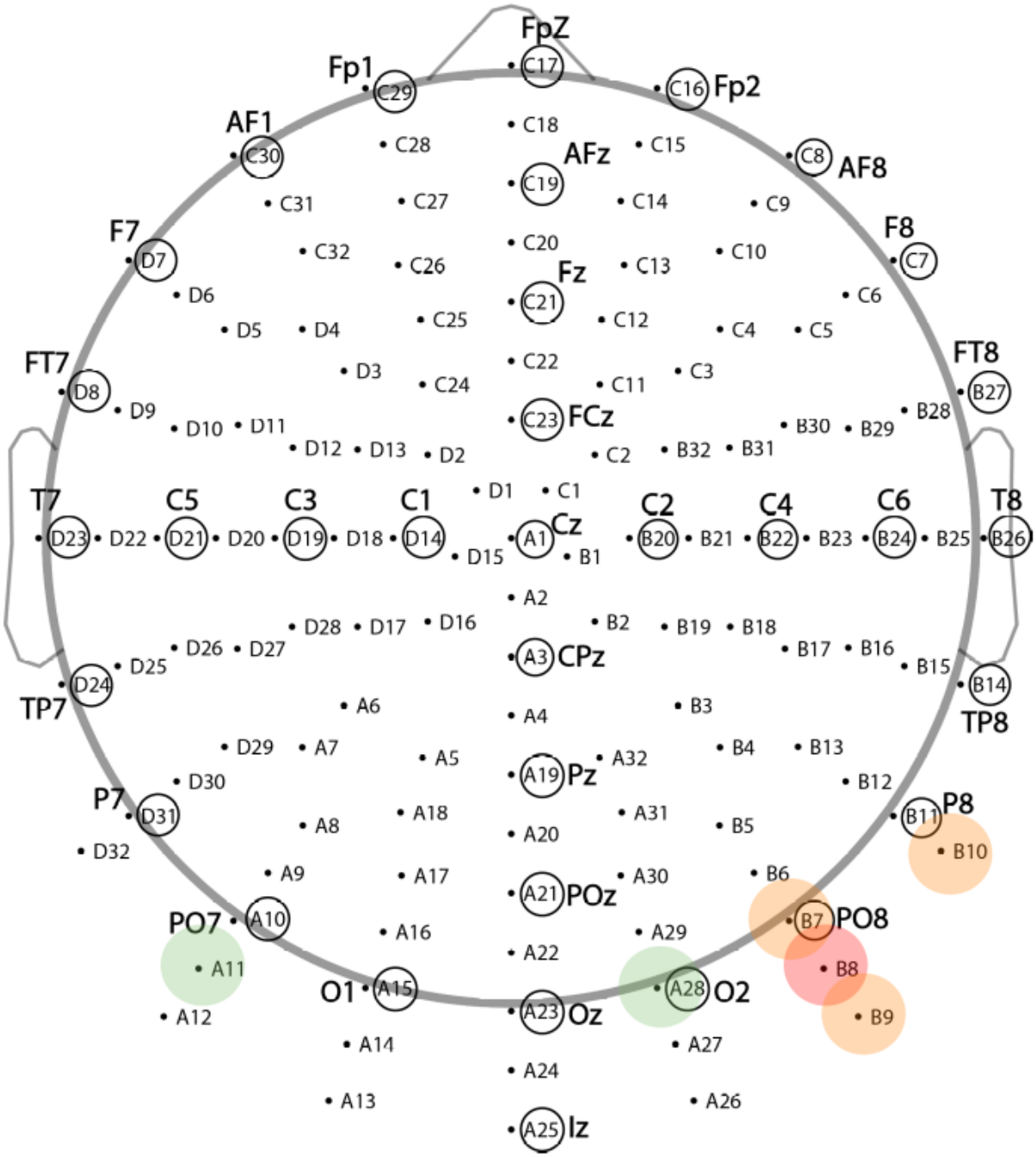


Figure 15: Biosemi 128 EEG electrode map: the six most frequent max t-value electrode channels are coloured. Colours map to counts: red=(33), orange=(16,17,19), green=(9,11).

4 Discussion

4.1 Overview

Our original question was: How soon after stimulus does the brain start to decode object categories? In particular, what is the ERP onset time for face sensitivity? While 70ms may be plausible (Bieniek et al., 2016), a rare distribution of onsets reported by (Liu et al., 2009) near 30ms after stimulus presentation is not.

For this report we hypothesised that cluster-based methods can make earlier, more accurate and consistent onset predictions than non cluster-based methods due to their ability to exploit structure in data. However, a simulation study by (Sassenhagen and Draschkow, 2019) caused us to have doubts. In fact (Rousselet, 2023) had already suggested the Binseg CPD method make better ERP onset predictions on some ERP datasets.

So, to test this hypothesis we compared the performance of several non cluster-based methods, known as change point detection (CPD) methods, with the spatio-temporal and temporal cluster-based methods. But CPD methods are general methods that do not exploit structure in time series data to detect abrupt changes in, e.g. mean, variance, or frequency. It seemed that the cluster-based methods should be more accurate. However, data preprocessing levelled the playing field such that their performance was similar.

We found that the temporal cluster-based and Binseg in agreement and both made fairly consistent predictions, suggesting that the most likely face sensitivity ERP onset time is in the 72ms to 76ms range.

4.2 Virtual electrodes make onset detection more accurate

The virtual electrode technique leverages the strongest, most reliable signals for each participant, both spatially and temporally, to significantly enhance the accuracy and reliability of ERP onset detection, allowing for more precise and meaningful analysis of neural responses to face and texture image stimuli. Onset predictions based on one virtual electrode for all participants per session were earlier and more consistent that predictions based on one virtual electrode per participant per session. The implication is that the virtual electrode technique increases the signal to noise ratio, making the ERP onset easier to detect. In other words, sensitivity and selectivity is increased.

Indeed, as expected, constructing one virtual electrode for all participants per session further reduced noise and the impact of artefacts incurred during EEG recordings, as confirmed by Figures 12(a) and (b) for sessions 1 and 2, which are almost identical. The time series is consistently very flat all the way until about 70ms where the onset is obvious.

Briefly, we inferred that the virtual electrode technique is effective for the following reasons: enhanced signal-to-noise ratio, adaptive selection of electrodes, t^2 tests at every time point allows temporal precision, reduction of variability in experimental factors,

focus on task-relevant differences (face minus texture).

4.3 Relative performance of the methods

The spatio-temporal cluster-based method and the temporal cluster-based method, in particular, made reasonable and consistent onset predictions across the two sessions. This can be partly explained by their ability leverage structural patterns in data that occur when EEG recorded values adjacent in time and/or space are not statistically independent.

Cluster-based methods can therefore mitigate the multiple comparisons problem by taking into account clusters of correlated data points, thereby also increasing sensitivity and specificity for detecting true ERP onsets. Rather than focussing on individual time points, temporal cluster-based method focusses on clusters of time points, which effectively controls for multiple comparisons, and reduces the likelihood of false positives. As a result, this method is more sensitive to detecting true effects that may be temporally diffuse because it aggregates evidence across contiguous time points.

The permutation test ensures that the identified clusters are unlikely to occur by chance, enhancing the reliability of the detections (Maris and Oostenveld, 2007). Indeed, the cluster-sum method provides a robust and statistically valid approach to detecting ERP onsets in EEG data (Rousselet, 2023).

However, the number of permutations necessary for accurate and reliable onset detection makes this method computationally intensive relative to conventional CPD methods when all electrodes and all trials per participant are used to predict the ERP onset for each individual participant. The virtual electrode technique may make cluster-based methods unnecessary for some datasets. However, when one virtual electrode is constructed per session, the temporal cluster-based methods is not computationally intensive even for 1000 permutations due to the dimensionality reduced provide by the virtual electrode technique.

We found that a data preprocessing technique, that constructs a *virtual electrode*, can extract the most relevant patterns from data such that best conventional CPD algorithm performed as well or better than the cluster-based methods for predicting onset times for individual participants, especially when data preprocessing was applied to all participants in a session for predicting the ERP onset time for the group of participants in a session.

5 Future Work

It may be that cluster-based methods when combined with multiple testing correction techniques will be more useful when applied to more complex data sets where the number and approximate location of onsets is not known. In that case conventional CPD methods could perform poorly even when combined with virtual electrodes. The real EEG data

used in this report had only one onset per epoch. So it may be worthwhile to test cluster-based methods and supporting techniques, including virtual electrodes, on other EEG datasets, especially offline where time constraints are not strict.

6 Conclusion

We validated and optimised the CPD algorithms implemented in the Python package `ruptures` using synthetic data and a gridsearch over a lists of cost functions. The other parameters were not optimised. We found that the Binseg method performed better overall than the other CPD methods on the synthetic datasets.

We applied cluster-based and CPD methods to the real EEG dataset and evaluated the ERP onsets predicted by spatio-temporal and temporal cluster-based methods relative to those predicted by the CPD methods. The two cluster-based methods made consistent onset predictions across the two sessions, with the temporal cluster-based onsets being slightly more consistent. However, the test-retest reliability Binseg was similar to that of the temporal cluster-based method while requiring far less computation in the one virtual electrode per participant per session case. But in the one virtual electrode per session case the computational requirements of these methods were similar. The key to accurate and reliable onset predictions for this dataset appears to depend more on data preprocessing and engineering, i.e. the construction of virtual electrodes, than on complex clustering methods and computationally intensive permutations for detecting the earliest statistically significant cluster and time point within that cluster.

References

Shlomo Bentin and Leon Y. Deouell. STRUCTURAL ENCODING AND IDENTIFICATION IN FACE PROCESSING: ERP EVIDENCE FOR SEPARATE MECHANISMS. *Cognitive Neuropsychology*, 17(1-3):35–55, February 2000. ISSN 0264-3294, 1464-0627. doi: 10.1080/026432900380472.

Magdalena M. Bieniek, Patrick J. Bennett, Allison B. Sekuler, and Guillaume A. Rousselet. Data from: A robust and representative lower bound on object processing speed in humans, November 2015.

Magdalena M. Bieniek, Patrick J. Bennett, Allison B. Sekuler, and Guillaume A. Rousselet. A robust and representative lower bound on object processing speed in humans. *European Journal of Neuroscience*, 44(2):1804–1814, July 2016. ISSN 0953-816X, 1460-9568. doi: 10.1111/ejn.13100.

D. H. R. Blackwood and W. J. Muir. Cognitive Brain Potentials and their Application. *British Journal of Psychiatry*, 157(S9):96–101, December 1990. ISSN 0007-1250, 1472-1465. doi: 10.1192/S0007125000291897.

Katherine S. Button, John P. A. Ioannidis, Claire Mokrysz, Brian A. Nosek, Jonathan Flint, Emma S. J. Robinson, and Marcus R. Munafò. Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, 14(5): 365–376, May 2013. ISSN 1471-003X, 1471-0048. doi: 10.1038/nrn3475.

Carlos Cordoba and et al. Spyder: The Scientific Python Development Environment, 2024.

Arnaud Delorme and Scott Makeig. EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1):9–21, March 2004. ISSN 01650270. doi: 10.1016/j.jneumeth.2003.10.009.

Alexandre Gramfort. MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, 7, 2013. ISSN 1662453X. doi: 10.3389/fnins.2013.00267.

David M. Groppe, Thomas P. Urbach, and Marta Kutas. Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review. *Psychophysiology*, 48 (12):1711–1725, December 2011. ISSN 0048-5772, 1469-8986. doi: 10.1111/j.1469-8986.2011.01273.x.

Jacques Jonas, Bruno Rossion, Julien Krieg, Laurent Koessler, Sophie Colnat-Coulbois, Hervé Vespignani, Corentin Jacques, Jean-Pierre Vignal, Hélène Brissart, and Louis Maillard. Intracerebral electrical stimulation of a face-selective area in the right inferior occipital cortex impairs individual face discrimination. *NeuroImage*, 99:487–497, October 2014. ISSN 10538119. doi: 10.1016/j.neuroimage.2014.06.017.

R. Killick, P. Fearnhead, and I. A. Eckley. Optimal Detection of Changepoints With a Linear Computational Cost. *Journal of the American Statistical Association*, 107(500): 1590–1598, December 2012. ISSN 0162-1459, 1537-274X. doi: 10.1080/01621459.2012.737745.

Hesheng Liu, Yigal Agam, Joseph R. Madsen, and Gabriel Kreiman. Timing, Timing, Timing: Fast Decoding of Object Information from Intracranial Field Potentials in Human Visual Cortex. *Neuron*, 62(2):281–290, April 2009. ISSN 08966273. doi: 10.1016/j.neuron.2009.02.025.

Eric Maris and Robert Oostenveld. Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164(1):177–190, August 2007. ISSN 01650270. doi: 10.1016/j.jneumeth.2007.03.024.

Dirk Merkel. Docker: Lightweight linux containers for consistent development and deployment. *Linux Journal*, 2014(238):2, 2014.

Robert Oostenveld, Pascal Fries, Eric Maris, and Jan-Mathijs Schoffelen. FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience*, 2011:1–9, 2011. ISSN 1687-5265, 1687-5273. doi: 10.1155/2011/156869.

Nancy Nicholson Peterson, Charles E. Schroeder, and Joseph C. Arezzo. Neural generators of early cortical somatosensory evoked potentials in the awake monkey. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, 96(3):248–260, May 1995. ISSN 01685597. doi: 10.1016/0168-5597(95)00006-E.

G. A. Rousselet, R. A. A. Ince, N. J. Van Rijsbergen, and P. G. Schyns. Eye coding mechanisms in early human face event-related potentials. *Journal of Vision*, 14(13):7–7, November 2014. ISSN 1534-7362. doi: 10.1167/14.13.7.

Guillaume A. Rousselet. Using cluster-based permutation tests to estimate MEG/EEG onsets: How bad is it?, November 2023.

Guillaume A Rousselet, Cyril R Pernet, Patrick J Bennett, and Allison B Sekuler. Parametric study of EEG sensitivity to phase noise during face processing. *BMC Neuroscience*, 9(1):98, December 2008. ISSN 1471-2202. doi: 10.1186/1471-2202-9-98.

Jona Sassenhagen and Dejan Draschkow. Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or location. *Psychophysiology*, 56(6):e13335, June 2019. ISSN 0048-5772, 1469-8986. doi: 10.1111/psyp.13335.

A. J. Scott and M. Knott. A Cluster Analysis Method for Grouping Means in the Analysis of Variance. *Biometrics*, 30(3):507, September 1974. ISSN 0006341X. doi: 10.2307/2529204.

Charles Truong. Documentation for the ruptures change point package, 2024.

Charles Truong, Laurent Oudre, and Nicolas Vayatis. Selective review of offline change point detection methods. *Signal Processing*, 167:107299, February 2020. ISSN 01651684. doi: 10.1016/j.sigpro.2019.107299.

D.H. Wolpert and W.G. Macready. No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, 1(1):67–82, April 1997. ISSN 1089778X. doi: 10.1109/4235.585893.