The Journal of Open Source Software

NeuroCluster: A Python toolbox for nonparametric cluster-based statistical testing of neurophysiological data with respect to continuous predictors.

Alexandra Fink Skular [1,2,3*], **Christina Maher** [1,2,3*], **Salman Qasim** [2,3], **and Ignacio Saez** [1,2,3,4,5¶]

1 The Nash Family Department of Neuroscience, Icahn School of Medicine at Mount Sinai, NY, NY **2** The Nash Family Center for Advanced Circuit Therapeutics, The Mount Sinai Hospital, NY, NY **3** The Center for Computational Psychiatry, Icahn School of Medicine at Mount Sinai, NY, NY **4** Department of Neurosurgery, The Mount Sinai Hospital, NY, NY **5** Department of Neurology, The Mount Sinai Hospital, NY, NY ¶ Corresponding author * These authors contributed equally.

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor: Open Journals

Reviewers:

- @openjournals

Submitted: 01 January 1970
Published: unpublished

Summary

Cognitive neurophysiology offers a unique framework for studying cognitive brain-behavior relationships by relating electrophysiological signals to complex behaviors. With the advent of new technical and behavioral paradigms, researchers can design cognitive experiments that leverage both the spatiotemporal resolution of electrophysiological data and the complexity of continuous behavioral variables. Analyzing these data requires sophisticated statistical methods that can interpret multidimensional neurophysiological data and dynamic, continuous behavioral variables. Often used statistical frameworks for nonparametric, cluster-based statistical tests are specifically focused on the contrast between discrete behavioral conditions but are not suitable for assessing how continuous variables predict the occurrence of clusters in neurophysiological data. NeuroCluster is an open-source Python toolbox for analysis of two-dimensional electrophysiological data (e.g. time-frequency representations) related to multivariate and continuous behavioral variables. NeuroCluster introduces a statistical approach which uses nonparametric cluster-based permutation testing in tandem with linear regression to identify two-dimensional clusters of neurophysiological activity that significantly encodes time-varying, continuous behavioral variables. Uniquely, it also supports multivariate analyses by allowing for multiple behavioral predictors to model neural activity. NeuroCluster addresses a methodological gap in statistical approaches to relate continuous, cognitive predictors to underlying electrophysiological activity with time and frequency resolution, to determine the neurocomputational processes giving rise to complex behaviors.

Statement of need

`NeuroCluster` addresses a methodological gap in cognitive and behavioral neuroscience, by providing a Python-based statistical toolbox to relate continuous predictors to two-dimensional neurophysiological activity. Continuous predictors vary over an experimental session, reflecting dynamic behaviors, underlying cognitive processes, complex movements, trial-varying experimental conditions, perceptual signals, or value-based trial outcomes (Collins & Shenhav, 2022; Hoy et al., 2021; Mathis & Mathis, 2020; **?**). Standard analytical approaches for relating complex behavioral variables to neuronal activity sacrifice the complexity of neurophysiological signals by reducing the dimensionality of neuronal timeseries data (e.g., averaging across temporal, spectral, or spatial domains, or dimensionality reduction) (Crosse et al., 2016; Rey et al., 2015; Saez et al., 2018; Stokes & Spaak, 2016; **?**; **?**). Conversely, analysis methods

that preserve the complexity of neurophysiological data (i.e., two-dimensional timeseries) constrain behavioral predictors to discrete conditions (Domenech et al., 2020; Kosciessa et al., 2020; Maris & Oostenveld, 2007; **?**; **?**). Directly linking continuous experimental variables to two-dimensional physiological timeseries data offers a rigorous way to study brain-behavior relationships, by maintaining the complexity of dynamic behavior, without sacrificing the resolution of event-related neurophysiological activity.

`NeuroCluster` uses cluster-based permutation testing to identify significant two-dimensional clusters with respect to continuous task variables. Cluster-based nonparametric statistical testing is a standard approach to analyze two-dimensional event-related time series data, while controlling for multiple comparisons and reducing family-wise error rates (Cohen, 2014; Groppe et al., 2011; Maris, 2012; Maris & Oostenveld, 2007; Nichols & Holmes, 2002). Neurophysiological activity is typically aggregated by condition to perform a two-sample cluster-based permutation test, which tests whether the neuronal encoding patterns differ between two discrete task conditions, rather than continuous, trial-varying features (Bullmore et al., 1999; Maris & Oostenveld, 2007). While two-sample cluster-based permutation tests provide a nonparametric statistical inference tool for identifying the presence of significant clusters of activity between two conditions, they are insufficient for identifying the presence of clusters as a function of continuously varying predictors. `NeuroCluster` provides a solution to this analytical gap by performing linear regressions at individual points across the 2D neural matrix. This approach enables users to quantify the degree to which a continuous predictor is related to neurophysiological activity at the pixel-level and to perform analyses with multivariate behavioral data, by incorporating multiple continuous or categorical covariates in the regression models. The t-statistics corresponding to the predictor of interest from the pixel-wise regressions are thresholded by a critical t-statistic to control for the FDR, creating a binary 2D matrix (Genovese et al., 2002). The binary 2D matrix is then used to identify putative 2D clusters of activation related to the continuous predictor of interest. This process is repeated many times with the predictor of interest randomly permuted to produce a surrogate distribution of 2D clusters. Clusters that survive cluster-based permutation testing are classified as significant regions of activation with respect to the specified continuous predictor.

`NeuroCluster` is applicable for numerous analysis goals; the major use cases are performing an initial exploratory analysis to generate specific hypotheses, determine data-driven windows interest, or to identify regional patterns of significant clusters within and between subjects. Future adaptations of `NeuroCluster` may implement mixed effects regressions, nonlinear mapping models, or group-level analysis frameworks (Bianchi et al., 2019; Ivanova et al., 2022; König et al., 2024; Yu et al., 2022). We demonstrate our approach with human intracranial local field potential data, but NeuroCluster is applicable for all types of two-dimensional neurophysiological measures (e.g., spatiotemporal clusters from EEG/MEG, cross-frequency interactions). To our knowledge, `NeuroCluster` presents a novel Python-based statistical software package. `NeuroCluster` is designed to supplement existing Python-based electrophysiological analysis toolboxes (Donoghue et al., 2020; Gramfort, 2013; Kosciessa et al., 2020; Whitten et al., 2011), particularly MNE-Python.

NeuroCluster Documentation

`NeuroCluster` is accompanied by a detailed tutorial which outlines the workflow (Fig. 1) for implementing this approach with time-frequency power estimates from multi-region LFP recording.

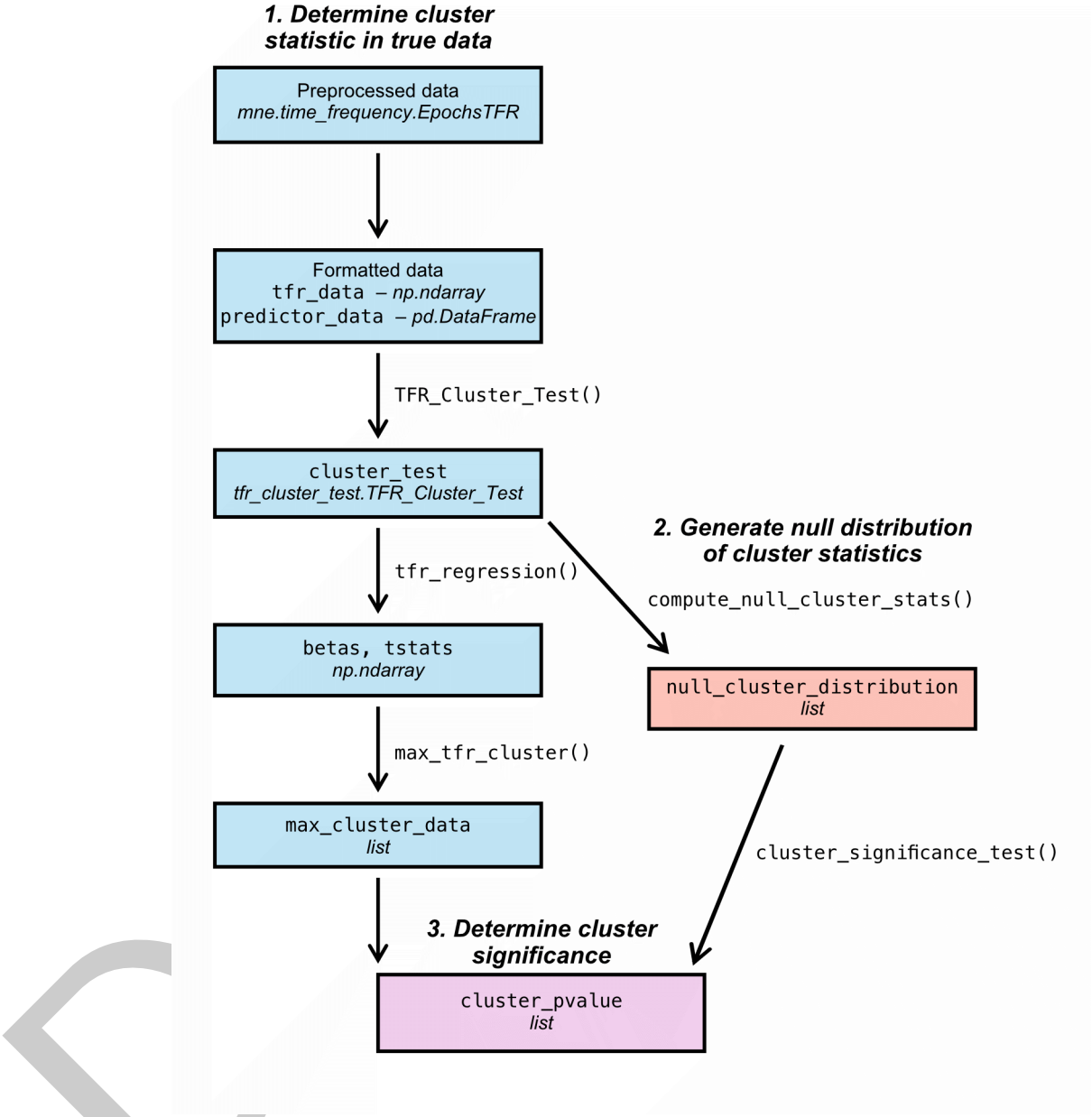

Figure 1: NeuroCluster workflow. This approach involves three key steps: (1) determine cluster statistic in true data, (2) generate a null distribution of cluster statistics by permuting dataset, (3) determine significance of true cluster statistic against null distribution.

Below we outline the statistical approach implemented by `NeuroCluster` for performing nonparametric permutation-based cluster testing using time-frequency resolved power estimates from neural data estimated using (**?**) and continuous predictors (i.e., latent cognitive processes, behavior, or experimental conditions). In these example data, we are testing the hypothesis that RPEs are significantly encoded in the electrophysiological signal from a given iEEG channel time-frequency representation (TFR). The following methodological description is based on data collected from a neurosurgical epilepsy patient undergoing stereotactic EEG (sEEG) monitoring for treatment-resistant depression. During the monitoring period, the patient performed a value-based decision-making task while local field potentials (LFPs) were recorded

from both cortical and subcortical sites. By analyzing the patient's behavior during the task, we derived continuous variables representing hypothesized latent cognitive processes—such as the trial-by-trial computation of reward prediction errors (RPEs)—to examine their relationship with neural activity.

1. Determine cluster statistic in true data

A. Define clusters: At each time-frequency index, we perform a linear univariate (or multivariate) regression using behaviorally-derived independent variables (e.g., latent cognitive variables, behavioral measures, task conditions) to predict neuronal activity (i.e., power). The β coefficient represents the strength and direction of the relationship between each independent variable and the dependent variable. It is estimated from the regression model and reflects how changes in the independent variable are associated with changes in power at the specific time-frequency pair. Pixel-wise regressions are parallelized for speed. For each time-frequency pair, the β coefficient for the regressor of interest (the independent variable of primary interest) is extracted from the regression results (Fig 2A). A t-statistic is computed for the β coefficient to capture how significantly different it is from zero (Fig 2B). A significance threshold is applied to the t-statistics of the β coefficient for the regressor of interest. If the t-statistic for a time-frequency pair exceeds the significance threshold, the pair is deemed significant. Clusters are then defined as adjacent time-frequency pairs where all pairs within the cluster have t-statistics exceeding the threshold, according to the test's desired tails (Fig 2C).

B. Compute cluster statistics: For each identified cluster, sum the t-statistics of all time-frequency pairs within the cluster. In a two-tailed test (the default), compute both the maximum and minimum cluster sums (Fig 2D).

2. Generate null distribution of cluster statistics

A. Permutation procedure: Labels for the behavioral predictor of interest are shuffled for the desired number of permutations.

B. Recalculate cluster statistic: Steps 1A/1B are repeated to define clusters and compute cluster statistics for each permuted dataset.

C. Construct null distribution: The cluster statistics from all permutations are compiled to create a null distribution, representing the distribution of cluster statistics under the null hypothesis (Fig 2E). The permuted TFR regressions are also parallelized at the pixel-level, while each permutation is performed sequentially. We tested many iterations of these functions with different parallelization approaches and sequential permutation-level computations with pixel-level parallelization within each TFR regression was the fastest method.

3. Determine cluster significance

A. Compare true cluster statistic to null distribution to compute p-values: The proportion of cluster statistics in the null distribution falling above (or below) the true cluster statistic(s) determines the p-value associated with the cluster(s) identified in the true data (Fig 2E).

The Journal of Open Source Software

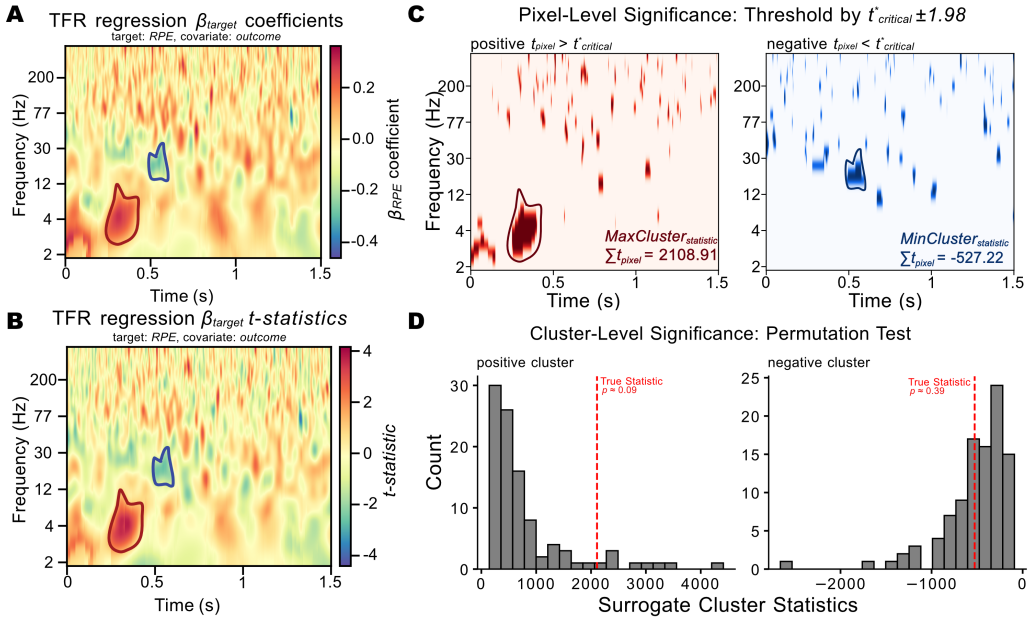

Figure 2: NeuroCluster methods. A. β coefficients for continuous predictor of interest (RPE) predicting power in given time-frequency pair (red outline = maximum positive cluster; blue outline = maximum negative cluster). B. T-statistics corresponding with βRPE coefficients. C. Clusters as determined using t-critical threshold. D. Maximum positive and negative clusters determined by summing t-statistics in identified clusters. E. Null distribution of cluster statistics generated by permuting dataset for predictor of interest (100 permutations; red dashed line = true cluster statistic.

4. Comparison of results to existing methods.

To evaluate the advantages of NeuroCluster, we compared its results to those obtained using MNE-Python's two-sample cluster-based permutation test. This approach requires discretizing the continuous variable of interest (RPE) into distinct categories, which reduces the resolution of the behavioral predictor. Additionally, MNE-Python's implementation does not support multivariate analyses, limiting the ability to model multiple behavioral covariates simultaneously. When applying the two-sample cluster test to our data, we did not identify any significant clusters of increased or decreased activity related to RPE. In contrast, NeuroCluster successfully detected significant clusters (Fig. 2), demonstrating its ability to preserve the richness of continuous behavioral variables and reduce the likelihood of false negatives. This comparison highlights NeuroCluster as a powerful and flexible alternative to existing statistical methods for analyzing continuous brain-behavior relationships.

5. Metric validation in synthetic data with known ground truth.

Thus far, we have demonstrated NeuroCluster using biological data. However, because these data are experimental, there is no definitive ground truth for the observed neural fluctuations associated with behavioral predictors. To validate the NeuroCluster method, we generated synthetic TFR data (2-200 Hz, sampling rate = 250, -1 to +1 seconds around "choice", 1 channel, 100 trials) with a known linear association between power in a specific time-frequency region and a continuous behavioral variable—in this case, the expected value of choice. Code for simulating these data is provided in the NeuroCluster repository. We then applied NeuroCluster to the synthetic dataset and, as expected, successfully identified a significant positive cluster corresponding to the known association embedded in the data (Fig. 3). This validation confirms the accuracy of NeuroCluster and provides evidence against its susceptibility to false positives.

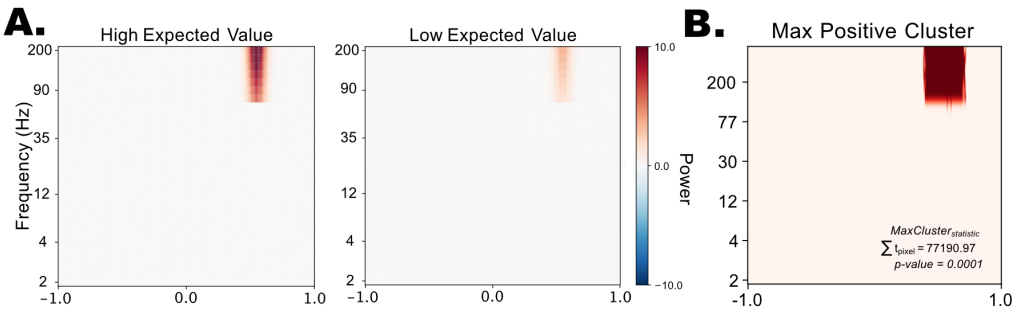

Figure 3: NeuroCluster validation in synthetic data. A. Time-frequency representation (TFR) showing power differences between high (>0.50) and low (<0.50) expected value trials in synthetic data (1 channel, 100 trials, time-locked to "choice"). B. A significant positive cluster identified in the expected time-frequency region, consistent with the predefined association embedded in the synthetic dataset.

Acknowledgements

We acknowledge feedback from Shawn Rhoads and support from Xiaosi Gu and Angela Radulescu during the genesis of this project.

References

Bianchi, B., Shalom, D. E., & Kamienkowski, J. E. (2019). Predicting known sentences: Neural basis of proverb reading using non-parametric statistical testing and mixed-effects models. *Frontiers in Human Neuroscience*, *13*, 82. https://doi.org/10.3389/fnhum.2019.00082

Bullmore, E. T., Suckling, J., Overmeyer, S., Rabe-Hesketh, S., Taylor, E., & Brammer, M. J. (1999). Global, voxel, and cluster tests, by theory and permutation, for a difference between two groups of structural MR images of the brain. *IEEE Transactions on Medical Imaging*, *18*(1), 32–42. https://doi.org/10.1109/42.750253

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

Collins, A. G., & Shenhav, A. (2022). Advances in modeling learning and decision-making in neuroscience. *Neuropsychopharmacology*, *47*(1), 104–118. https://doi.org/10.1038/s41386-021-01126-y

Crosse, M. J., Di Liberto, G. M., Bednar, A., & Lalor, E. C. (2016). The multivariate temporal response function (mTRF) toolbox: A MATLAB toolbox for relating neural signals to continuous stimuli. *Frontiers in Human Neuroscience*, *10*. https://doi.org/10.3389/fnhum.2016.00604

Domenech, P., Rheims, S., & Koechlin, E. (2020). Neural mechanisms resolving exploitation-exploration dilemmas in the medial prefrontal cortex. *Science*, *369*(6507), eabb0184. https://doi.org/10.1126/science.abb0184

Donoghue, T., Haller, M., Peterson, E. J., Varma, P., Sebastian, P., Gao, R., Noto, T., Lara, A. H., Wallis, J. D., Knight, R. T., & others. (2020). Parameterizing neural power spectra into periodic and aperiodic components. *Nature Neuroscience*, *23*(12), 1655–1665. https://doi.org/10.1038/s41593-020-00744-x

Genovese, C. R., Lazar, N. A., & Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *NeuroImage*, *15*(4), 870–878. https://doi.org/10.1006/nimg.2001.1037

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

Groppe, D. M., Urbach, T. P., & Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields i: A critical tutorial review. *Psychophysiology*, *48*(12), 1711–1725. https://doi.org/10.1111/j.1469-8986.2011.01273.x

Hoy, C. W., Steiner, S. C., & Knight, R. T. (2021). Single-trial modeling separates multiple overlapping prediction errors during reward processing in human EEG. *Communications Biology*, *4*(1), 910. https://doi.org/10.1038/s42003-021-02426-1

Ivanova, A. A., Schrimpf, M., Anzellotti, S., Zaslavsky, N., Fedorenko, E., & Isik, L. (2022). Beyond linear regression: Mapping models in cognitive neuroscience should align with research goals. *Neurons, Behavior, Data Analysis, and Theory*, *1*. https://doi.org/10.51628/001c.37507

König, S. D., Safo, S., Miller, K., Herman, A. B., & Darrow, D. P. (2024). Flexible multi-step hypothesis testing of human ECoG data using cluster-based permutation tests with GLMEs. *NeuroImage*, *290*, 120557. https://doi.org/10.1016/j.neuroimage.2024.120557

Kosciessa, J. Q., Grandy, T. H., Garrett, D. D., & Werkle-Bergner, M. (2020). Single-trial characterization of neural rhythms: Potential and challenges. *NeuroImage*, *206*, 116331. https://doi.org/10.1016/j.neuroimage.2019.116331

Maris, E. (2012). Statistical testing in electrophysiological studies. *Psychophysiology*, *49*(4), 549–565. https://doi.org/10.1111/j.1469-8986.2011.01320.x

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Mathis, M. W., & Mathis, A. (2020). Deep learning tools for the measurement of animal behavior in neuroscience. *Current Opinion in Neurobiology*, *60*, 1–11. https://doi.org/10.1016/j.conb.2019.10.008

Nichols, T. E., & Holmes, A. P. (2002). Nonparametric permutation tests for functional neuroimaging: A primer with examples. *Human Brain Mapping*, *15*(1), 1–25. https://doi.org/10.1002/hbm.1058

Rey, H. G., Ahmadi, M., & Quian Quiroga, R. (2015). Single trial analysis of field potentials in perception, learning and memory. *Current Opinion in Neurobiology*, *31*, 148–155. https://doi.org/10.1016/j.conb.2014.10.009

Saez, I., Lin, J., Stolk, A., Chang, E., Parvizi, J., Schalk, G., Knight, R. T., & Hsu, M. (2018). Encoding of multiple reward-related computations in transient and sustained high-frequency activity in human OFC. *Current Biology*, *28*(18), 2889–2899.e3. https://doi.org/10.1016/j.cub.2018.07.045

Stokes, M., & Spaak, E. (2016). The importance of single-trial analyses in cognitive neuroscience. *Trends in Cognitive Sciences*, *20*(7), 483–486. https://doi.org/10.1016/j.tics.2016.05.008

Whitten, T. A., Hughes, A. M., Dickson, C. T., & Caplan, J. B. (2011). A better oscillation detection method robustly extracts EEG rhythms across brain state changes: The human alpha rhythm as a test case. *NeuroImage*, *54*(2), 860–874. https://doi.org/10.1016/j.neuroimage.2010.08.064

Yu, Z., Guindani, M., Grieco, S. F., Chen, L., Holmes, T.-C., & Xu, X. (2022). Beyond t test and ANOVA: Applications of mixed-effects models for more rigorous statistical analysis in neuroscience research. *Neuron*, *110*(1), 21–35. https://doi.org/10.1016/j.neuron.2021.10.030