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

**# 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 (Haegens et al., 2022; Hoy et al., 2021; Mathis & Mathis, 2020; Saez et al., 2018). 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 (Groppe et al., 2011; Maris, 2012; Maris & Oostenveld, 2007). 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 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 (CITATIONS). (Allen et al., 2024; Collins & Shenhav, 2022; 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). (CITATIONS). Conversely, analysis methods that preserve the complexity of neurophysiological data (i.e. two-dimensional timeseries) constrain behavioral predictors to discrete conditions (CITATIONS). 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; Candia-Rivera & Valenza, 2022; Maris & Oostenveld, 2007; Nichols & Holmes, 2002).(Başar et al., 2000; Burke et al., 2015; Domenech et al., 2020; Rey et al., 2015). 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 pixel-wise regression outputs are used to identify putative 2D clusters of activation related to the continuous predictor of interest. Then, 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 temporal windows and/or frequencies of interest, or to identify regional patterns of significant clusters within and between subjects. Future directions for NeuroCluster may implement mixed effects regressions, multiple cluster detection, and/or group-level analysis tools (CITATIONS). We demonstrate our approach with human intracranial local field potential data, but NeuroCluster provides functionality for all types of spatiotemporal or spectrotemporal neurophysiological measures (EEG, MEG) (CITATION) and may be applicable to phase-amplitude or phase-phase cross-frequency coupling analyses (CITATION). NeuroCluster is designed to supplement existing Python-based electrophysiological analysis toolboxes (CITATION FOOOF, MNE, eBOSC), such as MNE, which currently offers a cluster-based permutation testing approach for discrete group comparisons (mne.stats.permutation\_cluster\_test) (CITATION MNE).

Computational cognitive models are used to operationalize unobservable cognitive processes and provide estimates of latent cognitive variables, based on participants’ behaviors (Daw, 2009; Drummond & Niv, 2020; Pan et al., 2024). For instance, some cognitive models generate continuous, trial-wise value estimates, like reward prediction errors (RPEs) (Hoy et al., 2021; O’Doherty et al., 2007a).

However, current statistical methods are ill-equipped to interpret complex, cognitive behaviors (Collins & Shenhav, 2022; O’Doherty et al., 2007a; Wiecki et al., 2015), nor can they manage the high dimensionality of multi-region intracranial electrophysiological recordings (Buzsáki & Draguhn, 2004; Donoghue et al., 2022; Holdgraf et al., 2017; Siegel et al., 2012; Stringer et al., 2019; Vidaurre et al., 2018).

Additionally, NeuroCluster performs analyses with multivariate behavioral data by incorporating multiple predictors to model neural activity (CITATION?). NeuroCluster is amenable to analyses using the same statistical approach for model-based latent cognitive predictors (Drummond & Niv, 2020; O’Doherty et al., 2007b; Pan et al., 2024; Wiecki et al., 2015), model-free cognitive variables (CITATION), as well as continuous experimental (i.e., perceptual noise; Bang & Fleming (2018)) or behavioral (INTEROCEPTIVE?) (i.e., mood ratings; Blain & Rutledge (2020)) predictors.

The complexity of physiological timeseries data poses a challenge to neural data scientists trying to interpret or decode neurophysiological signals related to dynamic time-varying experimental events.

* + Emphasize importance of channel-specific null distribution & highlight that baseline oscillatory dynamics vary across regions/individuals so you must create a null distribution for each electrode
    - For each channel, generate a unique null distribution of time-frequency clusters related to permuted predictor data
    - A unique null distribution is generated for each channel to control for channel-specific spectrotemporal profiles
  + Highlight that this approach keeps time-frequency structure in-tact and allows us to test whether the cluster is specifically related to predictor of interest or whether it’s a false positive due to underlying structure in channel data

Emphasize importance of channel-specific null distribution & highlight that baseline oscillatory dynamics vary across regions/individuals so you must create a null distribution for each electrode

**# Documentation**

A diagram of cluster

Description automatically generatedNeuroCluster is accompanied by a detailed tutorial [link to Jupyter notebook] which outlines the workflow (Fig 1) for implementing this approach with time-frequency power estimates from multi-region LFP recording.

**Fig 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 Python-MNE 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).

1. ***Determine cluster statistic in true data*** 
   1. **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 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 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 coefficient to capture how significantly different it is from zero (**Fig 2B**).

A significance threshold is applied to the t-statistics of the coefficient for the regressor of interest (**Fig 3**). 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**).

* 1. **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**).

1. ***Generate null distribution of cluster statistics***
   1. **Permutation procedure:** Labels for the behavioral predictor of interest are shuffled for the desired number of permutations.
   2. **Recalculate cluster statistic:** Steps 1A/1B are repeated to define clusters and compute cluster statistics for each permuted dataset**.**
   3. **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.
2. ***Determine cluster significance*** 
   1. **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**).

***A screenshot of a graph

Description automatically generated***

**Fig 2. NeuroCluster methods. A.**  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 RPE 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).

A diagram of a normal distribution

Description automatically generated

**Fig 2. NeuroCluster methods. A.**  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 RPE 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).

The significance of partial slope coefficients is computed via parametric significance testing under given null hypothesis using an observed test statistic, specifically the *t*-statistic. For every pixel, we calculate the observed *t*-value of the partial slope for the regressor of interest from a Student *t*-distribution with df=N-K-1. Though each pixel(t,f) regression is unique, the degrees of freedom for every regression model are equal because the predictor data is identical across pixels, making the *t*-distributions for every regression identical (T~tdf). Since each pixel-wise *T*-distribution is identical, we are able to compute a global *t*-statistic, *t-critical*, that can be used to simultaneously threshold the observed *t*-values in time-frequency space.

**Acknowledgements**

Shawn, other PIs

**References**

CM References

[**https://www.pnas.org/doi/full/10.1073/pnas.1800795115**](https://www.pnas.org/doi/full/10.1073/pnas.1800795115) **- Bang & Fleming (2018)**

[**https://elifesciences.org/articles/57977**](https://elifesciences.org/articles/57977) **- Blain & Rutledge (2020)**

[**https://arxiv.org/abs/2406.14742**](https://arxiv.org/abs/2406.14742) **- Pan et al (2024)**

[**https://pubmed.ncbi.nlm.nih.gov/17416921/**](https://pubmed.ncbi.nlm.nih.gov/17416921/) **- O’Doherty et al. (2007)**

Donoghue T, Haller M, Peterson EJ, Varma P, Sebastian P, Gao R, Noto T, Lara AH, Wallis JD, Knight RT, Shestyuk A, Voytek B (2020). Parameterizing neural power spectra into periodic and aperiodic components. Nature Neuroscience, 23, 1655-1665. DOI: 10.1038/s41593-020-00744-x **– FOOOF reference**

Alexandre Gramfort, Martin Luessi, Eric Larson, Denis A. Engemann, Daniel Strohmeier, Christian Brodbeck, Roman Goj, Mainak Jas, Teon Brooks, Lauri Parkkonen, and Matti S. Hämäläinen. MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, 7(267):1–13, 2013. [doi:10.3389/fnins.2013.00267](https://doi.org/10.3389/fnins.2013.00267). **– MNE-Python**

**eBOSC references:**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. <http://doi.org/10.1016/j.neuroimage.2019.116331>

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. <http://doi.org/10.1016/j.neuroimage.2010.08.064>

**# 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 (Haegens et al., 2022; Hoy et al., 2021; Mathis & Mathis, 2020; Saez et al., 2018). 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 (Groppe et al., 2011; Maris, 2012; Maris & Oostenveld, 2007). 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 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 (CITATIONS). (Allen et al., 2024; Collins & Shenhav, 2022; 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). (CITATIONS). Conversely, analysis methods that preserve the complexity of neurophysiological data (i.e. two-dimensional timeseries) constrain behavioral predictors to discrete conditions (CITATIONS). 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; Candia-Rivera & Valenza, 2022; Maris & Oostenveld, 2007; Nichols & Holmes, 2002).(Başar et al., 2000; Burke et al., 2015; Domenech et al., 2020; Rey et al., 2015). 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 pixel-wise regression outputs are used to identify putative 2D clusters of activation related to the continuous predictor of interest. Then, 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 temporal windows and/or frequencies of interest, or to identify regional patterns of significant clusters within and between subjects. Future directions for NeuroCluster may implement mixed effects regressions, multiple cluster detection, and/or group-level analysis tools (CITATIONS). We demonstrate our approach with human intracranial local field potential data, but NeuroCluster provides functionality for all types of spatiotemporal or spectrotemporal neurophysiological measures (EEG, MEG) (CITATION) and may be applicable to phase-amplitude or phase-phase cross-frequency coupling analyses (CITATION). NeuroCluster is designed to supplement existing Python-based electrophysiological analysis toolboxes (CITATION FOOOF, MNE, eBOSC), such as MNE, which currently offers a cluster-based permutation testing approach for discrete group comparisons (mne.stats.permutation\_cluster\_test) (CITATION MNE).