NeuroCluster: A Python package to detect electrophysiological signals related to complex behaviors using time-frequency resolved multiple regression and non-parametric cluster-based permutation testing.

**# Summary**

Interpreting neurophysiological signals related to precise events and/or behaviors is a fundamental goal of systems neuroscience. With the advent of new biotechnologies and neurosurgical practices, human (and animal) intracranial electrophysiological recordings are becoming widely accessible, both in academic research and for the development of Brain-Computer Interface technologies. However, the classic statistical methods for analyzing event-related time series data are ill-equipped to manage multi-region human electrophysiology, often with simultaneous recordings from 100s of channels, in conjunction with complex data from human behavioral experiments. Additionally, field potential data is inherently multi-dimensional due to the biophysical properties of neuronal oscillations, which are comprised of frequency, power, and phase components for every sample in the time series. The complexity of neurophysiological recordings poses a substantial challenge to neural data scientists trying to decode oscillatory signals related to specific trial-by-trial events.

* Ephys intro – richness of time-frequency resolution, multi-site simultaneous recordings
  + Richness of time frequency resolution and why it matters
  + frequency-specific modulation is widely accepted and understood that oscillations of different frequencies correspond to distinct underlying mechanisms
  + temporal profile of signal – encoding latencies, pseudo network construction for further hypothesis-driven connectivity tests
* Particularly with human experiments where behavioral task readouts are often continuous variables, but classic approaches only compare two conditions, not continuous trial-by-trial behavioral measures
* Non parametric statistics is the gold standard for statistical testing of event related signals to control for FWER, type I errors – can be applied to univariate + multivariate data
* GOAL: Identify significant clusters of oscillatory power modulations with time-frequency resolution that encode trial-by-trial complex behavioral variables
* 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

Unique approach allows 1) multiple regression with time frequency resolution 2) non-parametric cluster-based permutation testing specific to behavioral variable of interest, controlling for null hypothesis that cluster is due to unrelated structure in data

**# Statement of Need**

Interpreting neurophysiological signals related to precise events and/or behaviors is a fundamental goal of systems neuroscience. With the advent of new biotechnologies and neurosurgical practices, human (and animal) intracranial electrophysiological recordings are becoming widely accessible, both in academic research and for the development of Brain-Computer Interface technologies. However, the classic statistical methods for analyzing event-related time series data are ill-equipped to manage multi-region human electrophysiology, often with simultaneous recordings from 100s of channels, in conjunction with complex data from human behavioral experiments. Additionally, field potential data is inherently multi-dimensional due to the biophysical properties of neuronal oscillations, which are comprised of frequency, power, and phase components for every sample in the time series. The complexity of neurophysiological recordings poses a substantial challenge to neural data scientists trying to decode oscillatory signals related to specific trial-by-trial events.

Ephys intro – richness of time-frequency resolution, multi-site simultaneous recordings \*\* define TFR clearly here so the entire function syntax isn’t confusing lol

* + Richness of time frequency resolution and why it matters
  + frequency-specific modulation is widely accepted and understood that oscillations of different frequencies correspond to distinct underlying mechanisms
  + temporal profile of signal – encoding latencies, pseudo network construction for further hypothesis-driven connectivity tests
* Particularly with human experiments where behavioral task readouts are often continuous variables, but classic approaches only compare two conditions, not continuous trial-by-trial behavioral measures
* GOAL: Identify significant clusters of oscillatory power modulations with time-frequency resolution that encode trial-by-trial complex behavioral variables
* 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
* Lack of time-frequency resolved regression approaches for complex behaviors
  + Frequency resolution is important – frequency-specific modulation is widely accepted and understood that oscillations of different frequencies correspond to distinct underlying mechanisms
  + Time resolution is important – can understand the temporal encoding properties of a given region – ie how long does it take information about signal X to get to Region Y. AND timing resolution allows you to construct a pseudo network, where the timing of Regions A,B,C can be related to the involvement of these regions on a mechanistic level – if region A has a cluster that’s 100 ms after region B, that’s a hint that the signal from B may be transmitted to A and gives potential hypothesis-driven directions for connectivity metrics
* Because of the nature of electrophysiological data, one cannot assume that you can determine task-relevant power modulations based on the assumption that any power modulations are going to be task-related
  + Because of the biophysical properties of electrophysiological data, there is always going to be some time-frequency structure in the data – how do you parse what is not just event related, but event related signals specific to your predictor of interest
  + Requires a priori hypothesis for relevant frequencies/times
  + Current approach 1: Collapsing relevant multi-dimensional data into trial-averaged signals from pre-defined frequency bands loses the richness and uniqueness of spectrotemporal data
  + Current approach 2: two-sample cluster based permutation testing – no behavioral complexity, can’t account for possible covariates

Applications of neurocluster:

* Start looking at data – find regions with significant clusters to look at which regions + freqs + times are potentially encoding task variables
* Look at how timing/duration/intensity of clusters/oscillatory bursts varies as a function of trial outcome
* Differences in timing of clusters in specific regions reflecting flow of processing
* “parallel vs sequential sampling”
* Data-driven epoch of interest encoding
* Data-driven frequency of interest encoding

**# Methods/Documentation**

## Multiple Comparisons Correction Using Parametric Test Statistic Thresholding

For example, a hypothetical iEEG study has a dataset of N=15 subjects that engage in a cognitive task with n=150 trials. For each task trial, you extract 3.0 seconds from every neural recording. If there are approximately n=750 channels across subjects, each time series will contain 2.25×10^5 data points in the time domain (n=150 trials, epochs = 3.0 s, sampling rate = 500Hz). After spectral decomposition (n=30 frequencies, Morlet wavelets), the spectral domain expands from N=1 dimensions to N=30 of power estimates for every frequency, for every time point, giving . Now each time series has N=6.75x10^7 data points, or 1.0125x10^8 data points across subjects. In addition to the computational load limitations for a dataset of this magnitude, hypothesis testing using standard statistical inference to relate time-frequency power modulations to trial-wise continuous variables is uninterpretable without multiple comparisons correction (MCC) to prevent inflated Type I & II error rates. Standard MCC approaches cannot sufficiently address a problem of such a magnitude. For example, using Bonferroni correction to reduce the significance threshold (alpha=0.05) by the number of independent tests (6.75x10^7), increasing the False Discovery Rate (FDR) beyond acceptable limits [Maris, 2011].

Here we propose a novel statistical framework to perform time-frequency resolved multiple regression to predict trial-wise power modulations from multivariate predictors. For every pixel in time-frequency space, the following regression is performed:

Yf,t [matrix notation for n epochs] = A black and white text

Description automatically generated

\*\*\*make formula in latex + make sure to cite statsmodels OLS [summarize: β i = partial slope coefficient (also called partial regression coefficient, metric coefficient). It represents the change in E(Y) associated with a one-unit increase in Xi when all other IVs are held constant. α = the intercept. Geometrically, it represents the value of E(Y) where the regression surface (or plane) crosses the Y axis. Substantively, it is the expected value of Y when all the IVs equal 0. ε = the deviation of the value Yj from the mean value of the distribution given X. This error term may be conceived as representing (1) the effects on Y of variables not explicitly included in the equation, and (2) a residual random element in the dependent variable. From <https://www3.nd.edu/~rwilliam/stats2/l02.pdf>]

to estimate the partial regression coefficient of the predictor of interest for every TFR pixel (t,f).

A chart of error and time

Description automatically generated with medium confidence\*insert betas plot into markdown file\*

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.

\*\*insert t stat equation + t values heatmap\*\*

\*use this fig to explain possible null hypotheses, defaults, and rationale\*

A diagram of a normal distribution

Description automatically generated

* Pixel-level correction of FWER –
  + T-statistic from regression is to test the hypothesis that the regression coefficients are not zero – NOT the power values

### Multiple Comparisons Correction using Cluster-Based Permutation Tests

* Non parametric statistics is the gold standard for statistical testing of event related signals to control for FWER, type I errors – can be applied to univariate + multivariate 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

Unique approach allows 1) multiple regression with time frequency resolution 2) non-parametric cluster-based permutation testing specific to behavioral variable of interest, controlling for null hypothesis that cluster is due to unrelated structure in data

Methods description/functionality:

* Cluster-based permutation testing approach: must test clusters against non-parametric null hypothesis, rather than frequentist null hypothesis.
  + Cluster detection identifies latent structures in multi-dimensional data that are not just spurious signals (one pixel signal can’t be a significant ‘encoding’ profile)
  + Cannot assume time-freq power is normally distributed around zero – must build a null distribution from permuted regressions to capture spectrotemporal clusters specifically related to parameter of interest, not all event related signals
  + Non parametric statistics: doesn’t rely on assumptions about data distributions + makes multiple comparisons correction easy \* give example of Bonferroni correction p value\* (Cohen, 2014) [chapter 13] errors – can be applied to univariate + multivariate 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
    - 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

Documentation:

* Comment on things in code for speed?
* Make sure to clarify that these should be standardized regression coefficients (maybe cite salman’s 2std standardization)

Future directions:

* Discuss implementation for mixed effects regression
* Allow for multiple cluster detection

Applications of neurocluster:

* Start looking at data – find regions with significant clusters to look at which regions + freqs + times are potentially encoding task variables
* Look at how timing/duration/intensity of clusters/oscillatory bursts varies as a function of trial outcome
* Differences in timing of clusters in specific regions reflecting flow of processing
* “parallel vs sequential sampling”
* Data-driven epoch of interest encoding
* Data-driven frequency of interest encoding

Acknowledgements

* Shawn, other PIs