BioPsyKit: A Python package for the analysis of biopsychological data

Robert Richer[*1], **Arne Küderle**[1], **Martin Ullrich**[1], **Nicolas Rohleder**[2], **and Bjoern M. Eskofier**[1]

1 Machine Learning and Data Analytics Lab (MaD Lab), Department Artificial Intelligence in Biomedical Engineering (AIBE), Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) **2** Chair of Health Psychology, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU)

Summary

Biopsychology is a field of psychology that analyzes how biological processes interact with behaviour, emotion, cognition, and other mental processes. Biopsychology covers, among others, the topics of sensation and perception, emotion regulation, movement (and control of such), sleep and biological rhythms, as well as acute and chronic stress.

To assess the interaction between biological and mental processes a variety of different methods are used in the field of biopsychology, such as *electrophysiology*, assessed via biosignals like electrocardiogram (ECG), electrodermal activity (EDA), electromyogram (EMG), or electroencephalogram (EEG), *neuroendocrine and inflammatory biomarker*, assessed via saliva- and blood-based samples, *self-reports*, assessed via psychological questionnaires, as well as *sleep, activity and movement*, assessed via inertial measurement units (IMUs).

These methods are used to collect data either during standardized procedures in the laboratory or in the wild. The collected data are typically analyzed using statistical methods, or, more recently, using machine learning methods.

In order to combine all these tools necessary for a researcher in the field of biopsychology into one single Python package we developed `BioPsyKit`.

Statement of need

Researchers in biopsychology often combine different assessment modalities during experiments in order to capture the interaction between biological and mental processes. One example might be collecting salivary biomarker (e.g., cortisol) during an acute stress protocol and investigating the correlation between biomarker and psychometric data assessed via self-reports, such as perceived stress, state anxiety, or positive/negative affect. However, currently, there exist no Python package that allows to systematically combine, process, and analyze these different data modalities out of one hand by using one common API. For that reason `BioPsyKit` enables researchers to write cleaner and reproducible analysis code, export analysis results in a standardized format, and create high quality figures with for scientific publications.

BioPsyKit Structure

The following section describes the structure and the core modules of `BioPsyKit`. An overview is also provided in Figure 1.
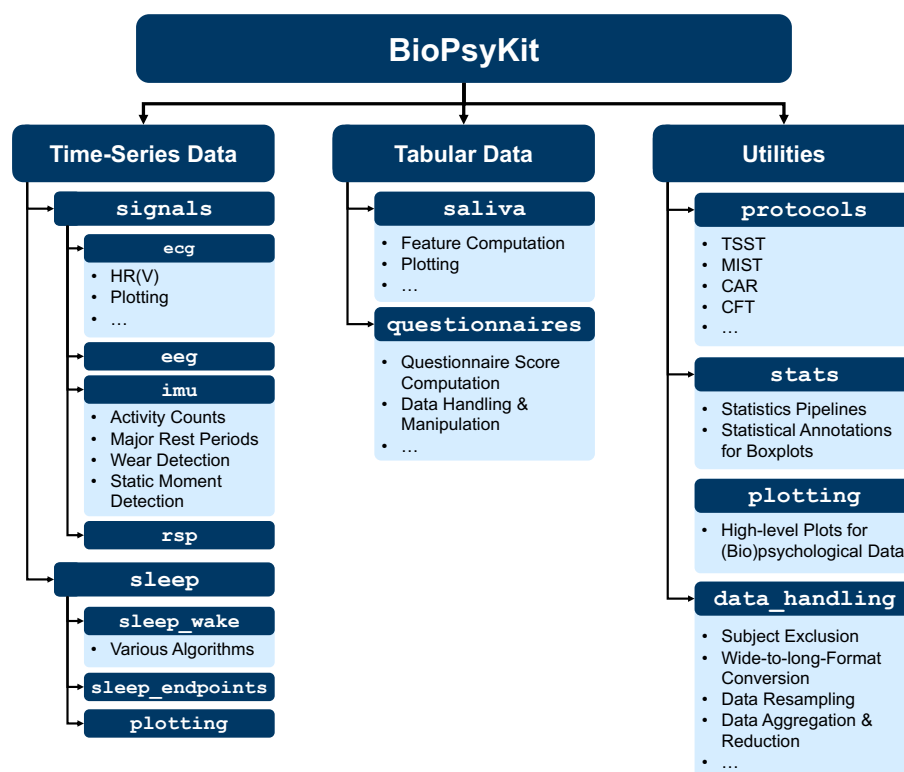
---

*corresponding author

Figure 1: Overview of BioPsyKit.

Physiological Signal Analysis

The module `biopsykit.signals` can be used for the analysis of various (electro)physiological signals (ECG, EEG, Respiration, Motion, and more). This includes:

- Classes to create processing pipelines for various physiological signals and for extracting relevant parameters from these signals. For physiological signal processing, `BioPsyKit` internally relies on the `neurokit2` Python library (Makowski et al., 2021), but offers further functionalities (e.g., the possibility to apply different outlier removal techniques R peaks extracted from ECG data).

- Plotting functions specialized for visualizing different physiological signals.

Sleep Analysis

The module `biopsykit.sleep` can be used for the analysis of motion data collected during sleep. This includes:

- Different algorithms for sleep/wake detection from wrist-worn activity or IMU data, such as the Cole/Kripke (Cole et al., 1992) or the Sadeh algorithm (Sadeh et al., 1994).
- Computation of sleep endpoints from detected sleep and wake phases and functions for plotting sleep processing results (e.g., Figure 2).
- Functions to import and process data from commercially available sleep trackers (e.g., Withings Sleep Analyzer).
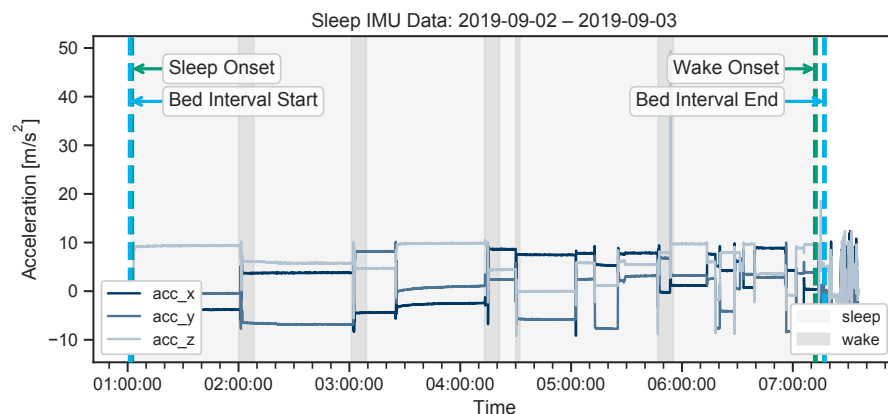
2

Figure 2: Example plot for visualizing computed sleep endpoints on IMU data.

Biomarker Analysis

The module `biopsykit.saliva` can be used for the analysis of saliva-based biomarker, such as cortisol and alpha-amylase. This also includes the extraction of relevant parameters characterizing salivary biomarker (e.g., area under the curve (Pruessner et al., 2003), slope, maximum increase, and more) and specialized plotting functions.

Self-report Analysis

The module `biopsykit.questionnaires` can be used for the analysis of psychometric self-reports, assessed via questionnaires. This includes:

- Functions to convert, clean, and impute tabular data from questionnaire studies.
- Implementation of various established psychological questionnaires, such as Perceived Stress Scale (PSS) (Cohen et al., 1983), Primary Appraisal Secondary Appraisal Scale (PASA) (Gaab et al., 2005) and functions to compute scores from questionnaire data.

Support for Psychological Protocols

The module `biopsykit.protocols` can be used for analyzing data collected during various psychological protocols. This includes:

- Protocols for the assessment of acute stress in the laboratory, e.g., Trier Social Stress Test (TSST) (Kirschbaum et al., 1993), Montreal Imaging Stress Task (MIST) (Dedovic et al., 2005).
- Protocols for the assessment of biological rhythms in the wild (e.g., Cortisol Awakening Response (CAR)).
- Specialized plotting functions for standardized visualization of data collected during these psychological protocols (such as, heart rate data: Figure 3, saliva data: Figure 4).

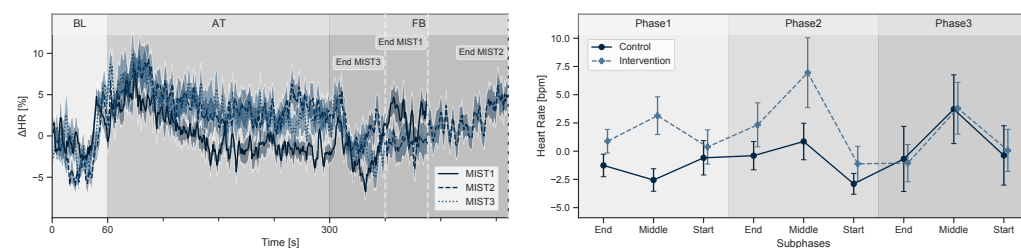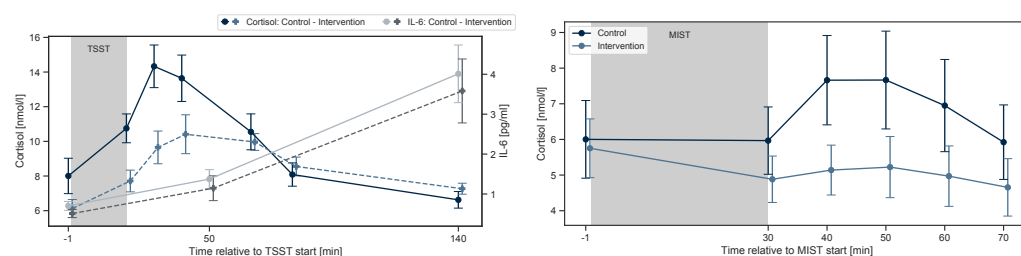**Figure 3:** Example plots for visualizing heart rate data collected during the MIST.

Figure 4: Example plots for visualizing biomarker data collected during the TSST (left) and the MIST (right).

Simplified Evaluation

The module `biopsykit.stats` and `biopsykit.classification` can be used for simplified evaluation of statistical analyses and machine learning pipelines that are frequently used in biopsychological research. `biopsykit.stats` provides functions to easily set up statistical analysis pipelines (using `pingouin` (Vallat, 2018)) and to visualize and export statistical analysis results in a standardized way (see, for example, Figure 5).

`biopsykit.classification` provides functions to set up, optimize and evaluate different machine learning pipelines for biopsychological problems.
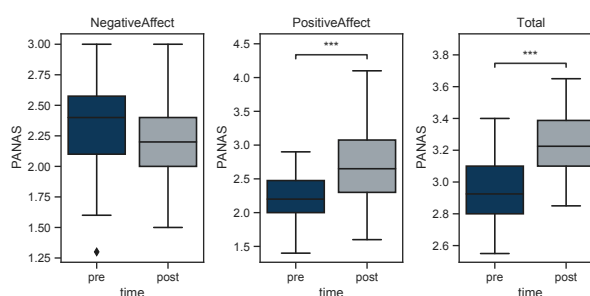

Figure 5: Example plot for adding statistical analysis results to boxplots.

Availability

The software is available as a pip installable package (`pip install biopsykit`), as well as on GitHub at: https://github.com/mad-lab-fau/BioPsyKit.

Acknowledgements

References

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A Global Measure of Perceived Stress. *Journal of Health and Social Behavior*, *24*(4), 385. https://doi.org/10.2307/2136404

Cole, R. J., Kripke, D. F., Gruen, W., Mullaney, D. J., & Gillin, J. C. (1992). Automatic Sleep/Wake Identification From Wrist Activity. *Sleep*, *15*(5), 461–469. https://doi.org/10.1093/sleep/15.5.461

Dedovic, K., Renwick, R., Mahani, N. K., Engert, V., Lupien, S. J., & Pruessner, J. C. (2005). The Montreal Imaging Stress Task: Using functional imaging to investigate the effects of perceiving and processing psychosocial stress in the human brain. *Journal of Psychiatry and Neuroscience*, *30*(5), 319–325.

Gaab, J., Rohleder, N., Nater, U. M., & Ehlert, U. (2005). Psychological determinants of the cortisol stress response: The role of anticipatory cognitive appraisal. *Psychoneuroendocrinology*, *30*(6), 599–610. https://doi.org/10.1016/J.PSYNEUEN.2005.02.001

Kirschbaum, C., Pirke, K.-M., & Hellhammer, D. H. (1993). The 'Trier Social Stress Test' A Tool for Investigating Psychobiological Stress Responses in a Laboratory Setting. *Neuropsychobiology*, *28*, 76–81. https://doi.org/10.1159/000119004

Makowski, D., Pham, T., Lau, Z. J., Brammer, J. C., Lespinasse, F., Pham, H., Schölzel, C., & Chen, S. H. A. (2021). NeuroKit2: A Python toolbox for neurophysiological signal processing. *Behavior Research Methods*, *53*(4), 1689–1696. https://doi.org/10.3758/s13428-020-01516-y

Pruessner, J. C., Kirschbaum, C., Meinlschmid, G., & Hellhammer, D. H. (2003). Two formulas for computation of the area under the curve represent measures of total hormone concentration versus time-dependent change. *Psychoneuroendocrinology*, *28*(7), 916–931. https://doi.org/10.1016/S0306-4530(02)00108-7

Sadeh, A., Sharkey, K. M., & Carskadon, M. A. (1994). Activity-Based Sleep-Wake Identification: An Empirical Test of Methodological Issues. *Sleep*, *17*(3), 201–207. https://doi.org/10.1093/sleep/17.3.201

Vallat, R. (2018). Pingouin: Statistics in Python. *Journal of Open Source Software*, *3*(31), 1026. https://doi.org/10.21105/joss.01026