

BioPsyKit: A Python package for the analysis of biopsychological data

Robert Richer^{*1}, Arne Küderle¹, Martin Ullrich¹, Nicolas Rohleder², and Bjoern M. Eskofier¹

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)

DOI: [10.21105/joss.03702](https://doi.org/10.21105/joss.03702)

Software

- [Review](#) ↗
- [Repository](#) ↗
- [Archive](#) ↗

Editor: [Øystein Sørensen](#) ↗

Reviewers:

- [@zen-juen](#)
- [@espenhgn](#)

Submitted: 01 September 2021

Published: 12 October 2021

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)).

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 modalities are used in the field of biopsychology, such as *electrophysiology*, assessed, for instance, via electrocardiography (ECG), electrodermal activity (EDA), or electroencephalography (EEG), *sleep, activity and movement*, assessed via inertial measurement units (IMUs), *neuroendocrine and inflammatory biomarkers*, assessed by saliva and blood samples, as well as *self-reports*, assessed via psychological questionnaires.

These different modalities are collected either “in the lab,” during standardized laboratory protocols, or “in the wild,” during unsupervised protocols in home environments. The collected data are typically analyzed using statistical methods, or, more recently, using machine learning methods.

While some software packages exist that allow for the analysis of single data modalities, such as electrophysiological data, or sleep, activity and movement data, no packages are available for the analysis of other modalities, such as neuroendocrine and inflammatory biomarker, and self-reports. In order to fill this gap, and, simultaneously, to combine all required tools analyzing biopsychological data from beginning to end into one single Python package, we developed BioPsyKit.

Statement of Need

Researchers in biopsychology often face the challenge of analyzing data from different assessment modalities during experiments in order to capture the complex interaction between biological and mental processes.

One example might be collecting (electro)physiological (e.g., ECG) or salivary biomarkers (e.g., cortisol) during an acute stress protocol, and investigating the correlation between biomarkers and psychometric data assessed via self-reports, such as perceived stress, state anxiety, or positive/negative affect. Another example is the assessment of relationships between sleep and neuroendocrine responses in the morning. To assess the beginning and end of sleep periods,

*corresponding author

as well as other sleep-related parameters, researchers typically use inertial measurement units (IMUs) or activity trackers. These data are then combined with psychometric data from self-reports (e.g., sleep quality, stress coping, etc.) and data from saliva samples to assess the cortisol awakening response (CAR) in the morning.

While some packages already address a subset of these different applications, such as `NeuroKit2` (Makowski et al., 2021) for the analysis of (electro)physiological data, `SleepPy` (Python) (Christakis et al., 2019) or `GGIR` (R) (Migueles et al., 2019) for sleep analysis from accelerometer data, no software package exists that unites all these different, heterogeneous data modalities under one umbrella. Furthermore, and to the best of our knowledge, no software packages exist that allow a standardized analysis of neuroendocrine biomarkers without the requirement to write analysis code from scratch. Likewise, no software packages that implement established psychological questionnaires, allowing to compute questionnaire (sub)scales from raw questionnaire items, have been published to date.

For that reason `BioPsyKit` addresses these limitations and offers all necessary building blocks for the analysis of biopsychological data. Our software package allows to systematically combine, process, and analyze data from different modalities using one common API. This enables researchers to write cleaner and more reproducible analysis code, to export results in a standardized format, and to create high quality figures 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](#).

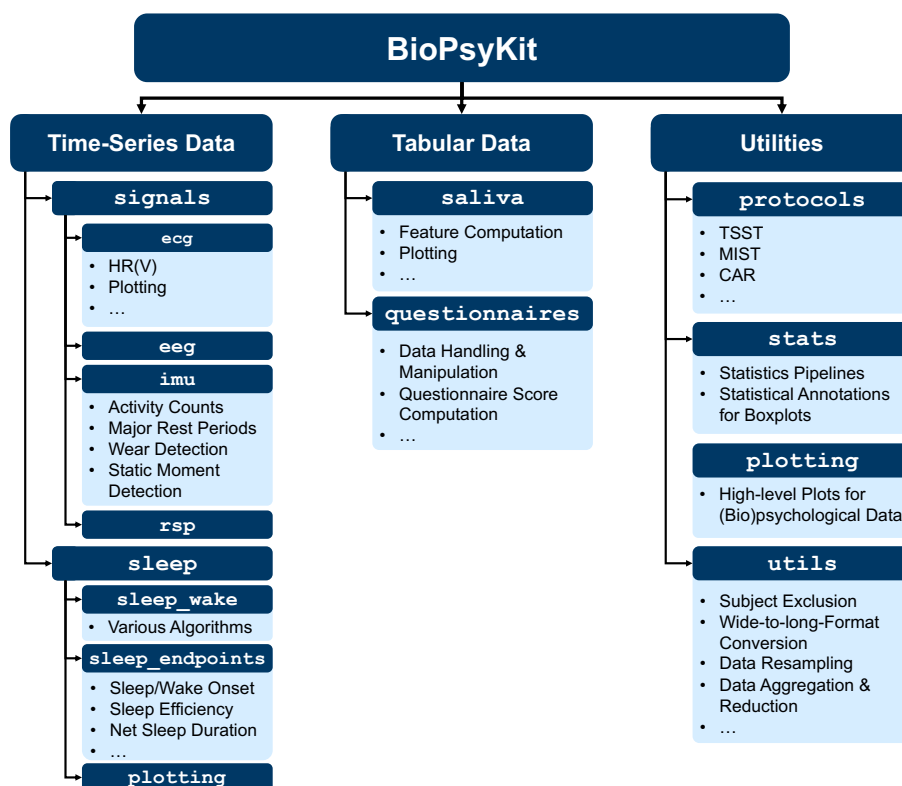


Figure 1: Overview of `BioPsyKit`.

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 R peak outlier removal techniques for 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).

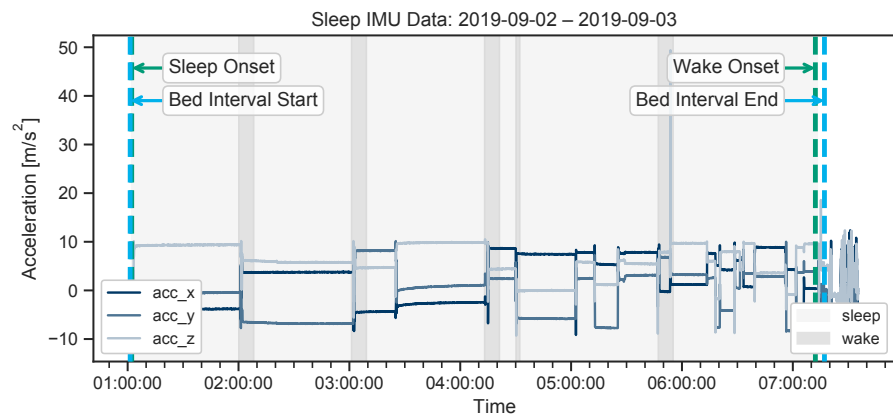


Figure 2: Example plot for visualizing computed sleep endpoints on IMU data.

Biomarker Analysis

The module `biopsykit.saliva` can be used for the analysis of saliva-based biomarkers, such as cortisol and alpha-amylase. This also includes the extraction of relevant parameters characterizing salivary biomarkers (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` provides an object-oriented interface for psychological protocols. On the one hand, it serves as data structure to store and access data collected during this psychological protocol from different modalities. On the other hand, the object-oriented interface allows to conveniently compute analysis results from the data added to the particular protocol instance, to export results, and to create plots for data visualization. This includes:

- Protocols for the assessment of biological rhythms, especially acute stress, in the laboratory, e.g., Trier Social Stress Test (TSST) (Kirschbaum et al., 1993) or 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).

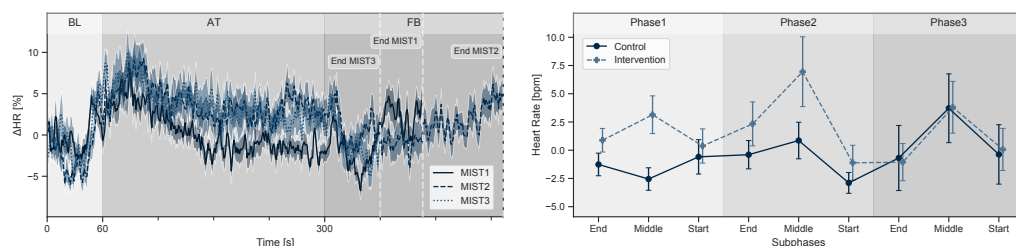


Figure 3: Example plots for visualizing heart rate data collected during the MIST.

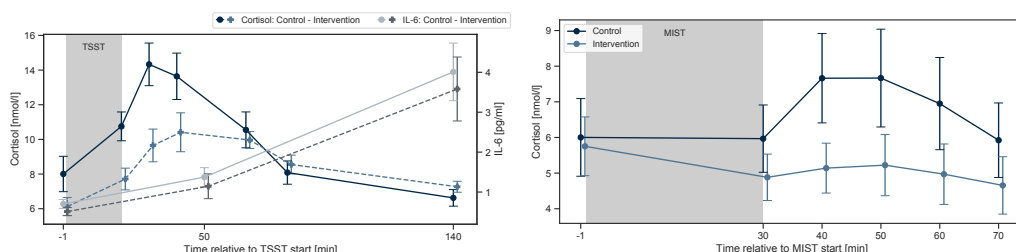


Figure 4: Example plots for visualizing biomarker data collected during the TSST (left) and the MIST (right).

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, e.g., Figure 5).

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

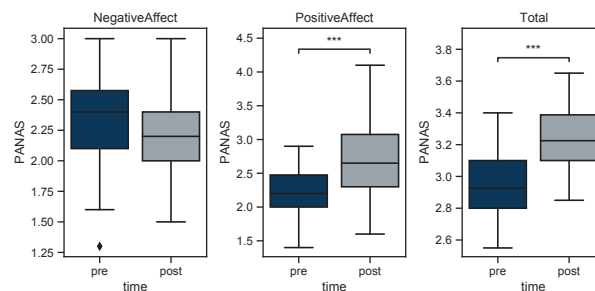


Figure 5: Example plot for adding statistical analysis results to boxplots.

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

We acknowledge contributions from Rebecca Lennartz, Daniel Krauß, Victoria Müller, and Janis Zenkner. Bjoern M. Eskofier gratefully acknowledges the support of the German Research Foundation (DFG) within the framework of the Heisenberg professorship programme (grant number ES 434/8-1). Furthermore, this work was partly supported by the DFG collaborative research center EmpkinS (CRC 1483).

References

- Christakis, Y., Mahadevan, N., & Patel, S. (2019). SleepPy: A python package for sleep analysis from accelerometer data. *Journal of Open Source Software*, 4(44), 4–7. <https://doi.org/10.21105/joss.01663>
- 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>
- Migueles, J. H., Rowlands, A. V., Huber, F., Sabia, S., & van Hees, V. T. (2019). GGIR: A Research CommunityDriven Open Source R Package for Generating Physical Activity and Sleep Outcomes From Multi-Day Raw Accelerometer Data. *Journal for the Measurement of Physical Behaviour*, 2(3), 188–196. <https://doi.org/10.1123/jmpb.2018-0063>
- 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](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>