Title: Influence of imputation strategies on the identification of brain functional connectivity networks

Authors: Britta Pester, Thomas Lehmann, Lutz Leistritz, Herbert Witte, Carolin Ligges

PII: S0165-0270(18)30288-7
DOI: https://doi.org/10.1016/j.jneumeth.2018.09.021
Reference: NSM 8123

To appear in: *Journal of Neuroscience Methods*

Received date: 7-6-2018
Revised date: 17-9-2018
Accepted date: 17-9-2018

Please cite this article as: Pester B, Lehmann T, Leistritz L, Witte H, Ligges C, Influence of imputation strategies on the identification of brain functional connectivity networks, *Journal of Neuroscience Methods* (2018), https://doi.org/10.1016/j.jneumeth.2018.09.021

Influence of imputation strategies on the identification of brain functional connectivity networks

Britta Pester[1], Thomas Lehmann[2], Lutz Leistritz[2], Herbert Witte[2], Carolin Ligges[3]

[1]Chair of Clinical Psychology and Behavioural Neuroscience, Department of Psychology, Technische Universität Dresden
[2]Institute of Medical Statistics, Computer and Data Sciences, Jena University Hospital, Friedrich Schiller University, Jena
[3]Department of Child and Adolescent Psychiatry, Psychosomatic Medicine and Psychotherapy, Jena University Hospital, Friedrich Schiller University Jena

Highlights

- EEG missing values
- Handling of imputation methods
- Imputation for neuronal networks

Abstract

Whenever neurophysiological data, such as EEG data are recorded, occurring artifacts pose an essential problem. This study addresses this issue by using imputation methods whereby whole data sets of a trial, or distinct electrodes, are not removed from the analysis of the EEG data but are replaced. We present different imputation strategies but use only two which are optimal for this particular study; predictive mean matching and data augmentation. The study addresses the as of yet unresolved question if the quality of derived brain functional networks is improved by imputation methods compared to traditional exclusion techniques which drop data, and will finally assesses the differences between the two imputation methods themselves used here.
In this study, EEG data from a study evaluating dyslexia-specific therapy on a neurophysiological level were used to investigate imputation strategies in research of cortical interaction. Several recorded values were artificially declared as 'missing'. This enables the comparison of networks based on the complete data set without any missing values (pseudo ground truth) and those derived from imputation approaches in a realistic situation of disturbed data. Functional connectivity was quantified by time-variant partial directed coherence, providing a directed, temporally varying and frequency-selective connectivity measure.
Based on the comparison between pseudo ground truth and networks of data with excluded missing values and data with imputed values, we found that any imputation strategy is preferable to the entire exclusion of data. The study also showed that the choice of the applied imputation algorithm impacts the resulting networks only marginally.