Massive generalized additive models of neurophysiological time-series

Joshua Krause, Jelmer P Borst, Jacolien van Rij
Bernoulli Institute, University of Groningen

Using statistical models for the analysis of neurophysiological time-series, such as electroencephalography (EEG) or pupil dilation recordings, is complicated by the fact that these signals change non-linearly over time. Additionally, other experimental continuous variables might have a non-linear effect on the measured signal as well. The analysis is typically further complicated by substantial between-subject and between-trial heterogeneity of these non-linear relationships. Generalized additive models (GAMs) are theoretically well-equipped to address both challenges, allowing the estimation of non-linear functions of predictor variables as well as random effects to account for these sources of heterogeneity. However, in practice it is often computationally intractable to include sufficient random effects, as it is not uncommon for cognitive experiments to involve thousands of trials across participants. Here, we combined and extended recently proposed strategies to reduce memory requirements and matrix infill into a sparse GAM estimation algorithm capable of handling previously impossible (non-linear) random effect structures. This allowed us to compute proper GAM models of pupil dilation data with $\sim$1.8 million observations and EEG data with $\sim$23 million observations. Fitting these models introduces new challenges for established model comparison strategies, which we investigated with simulation studies. Based on the results we established guidelines on how to identify the optimal model. To further facilitate this model-based analysis approach, we provide an openly available Python package of the algorithm and the investigated model comparison strategies.