Multiverse Analysis for Transparent and Replicable Neurometric Evaluations: Theoretical Plausibility of P300 Estimates Across Single Trial EEG Algorithms and Quantification Decisions

Rosina Zollner[1], Coşku Inceler[1], Cassie Ann Short[1]

[1]Psychological Methods and Statistics, Department of Psychology, School of Medicine and Health

Carl von Ossietzky Universität Oldenburg

Introduction

- **Replication crisis:** scientific findings difficult to replicate, raising concerns about credibility of science [1]
- Researchers highly flexible when it comes to decisions in the analysis process (*researcher's degrees of freedom*)
- (Exploratory) **Multiverse**: method to investigate several processing pipelines by creating a multiverse of datasets [2]
- We created a 4x4x4 multiverse (reference x electrode x single-trial algorithm)

- Single-trial latency estimation algorithm choice based on review from Ouyang et al. (2016) [3]
- **Theoretical plausibility** = logical expectation that reaction time and P300 latency are positively correlated

Aim of the study:
Compare robustness of theoretical plausibility of P300 latency estimates across 64 pipelines

Methods

① Dataset

- Previously collected and preprocessed by Nowparast Rostami and colleagues (2017) [4]
- 167 healthy adults
- Exclusion criteria: less than 20 correct trials in any condition
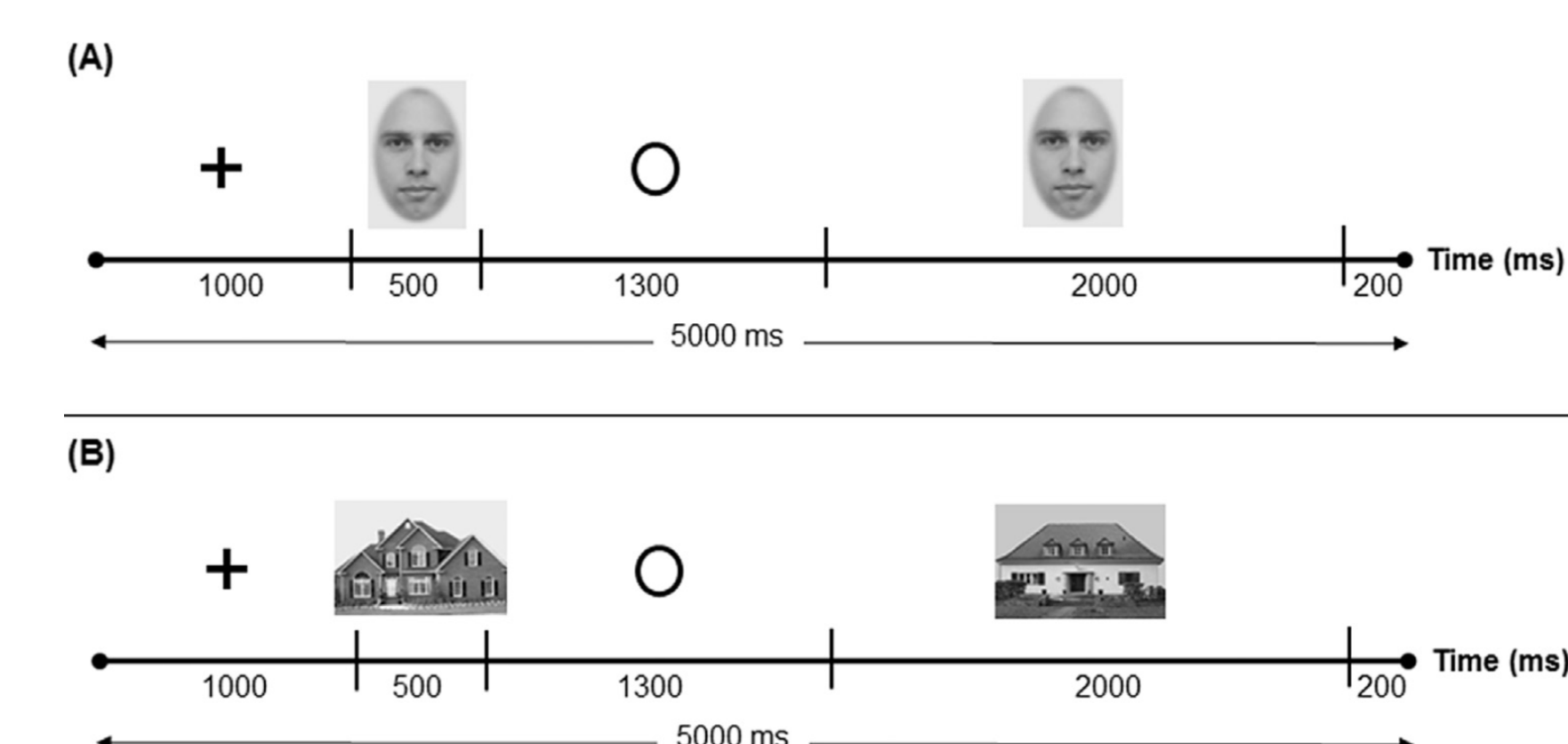- Primed learning and recognition task

Fig. 1: Visualization of two trials of the recognition task conducted during EEG recording. A: primed face. B: unprimed house. Figure taken from Nowparast Rostami et al. (2017) [4].

② Multiverse

Template Matching (TM)
→ Peak of cross-correlation curve between template and single-trials [3]

Peak-Picking (PP)
→ low-pass filtering at 3 Hz and global peaks picked from single-trial epochs [3]

Maximal Likelihood (MLE)
→ Latency set to 0, constantly updated by small amount δ (Tuan et al. 1987) [3]

Residue Iteration Decomposition (RIDE)
→decompose ERP into stimulus-locked component S, central component C and RT-locked component R [3]
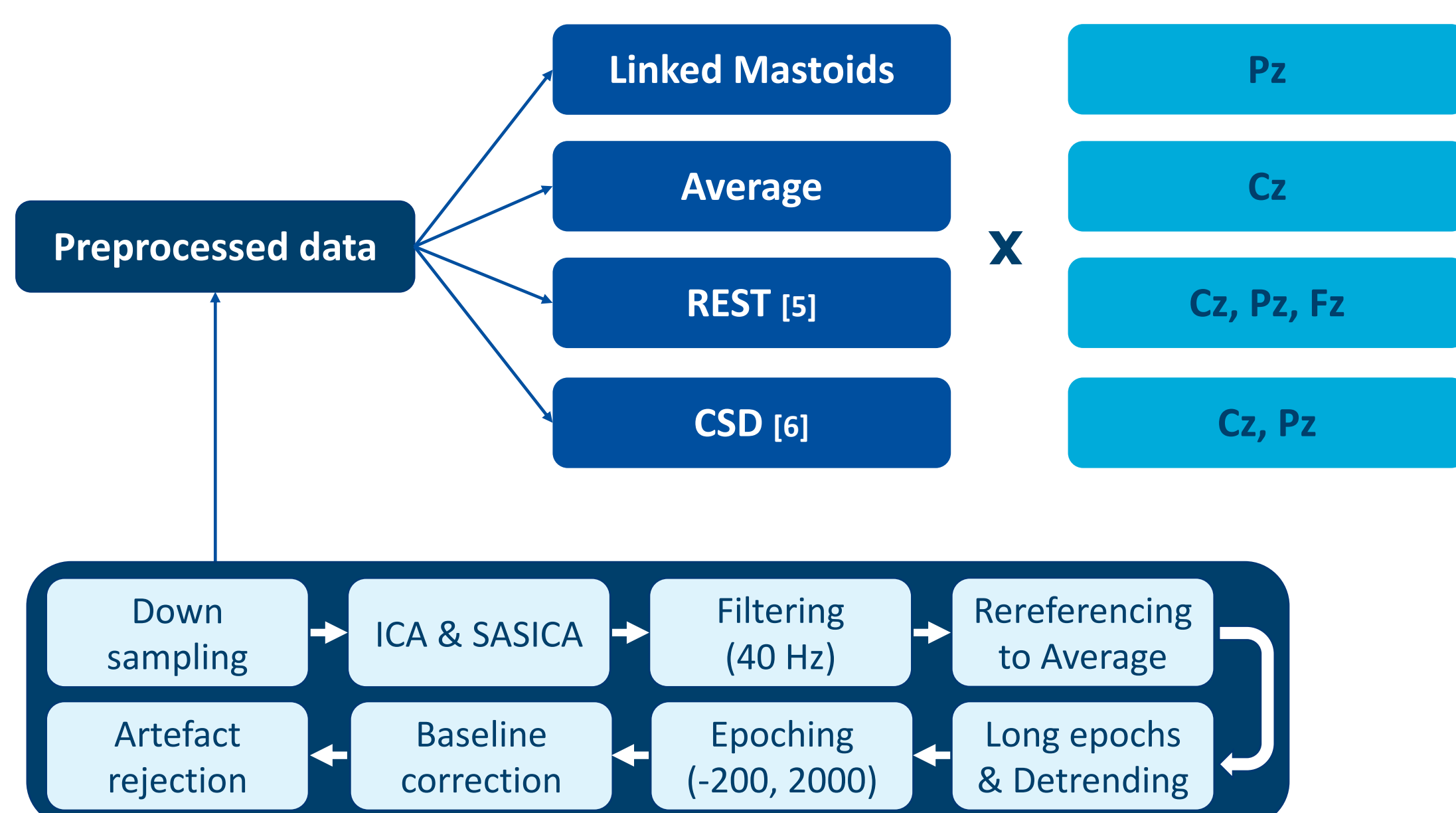
64 pipelines

Fig. 2: Simplified visualization of the multiverse with 3 decision knots: reference electrode (common average reference, linked mastoids, reference electrode standardization technique (REST) and current source density (CSD)), electrodes of interest (Cz, Pz; Fz, Cz, Pz; Cz; Pz) and single-trial analysis methods (Peak-Picking, Template Matching, Maximal Likelihood Estimation (MLE) and Residue Iteration Decomposition (RIDE)) (4x4x4). The preprocessing was conducted by the authors of the dataset [4].
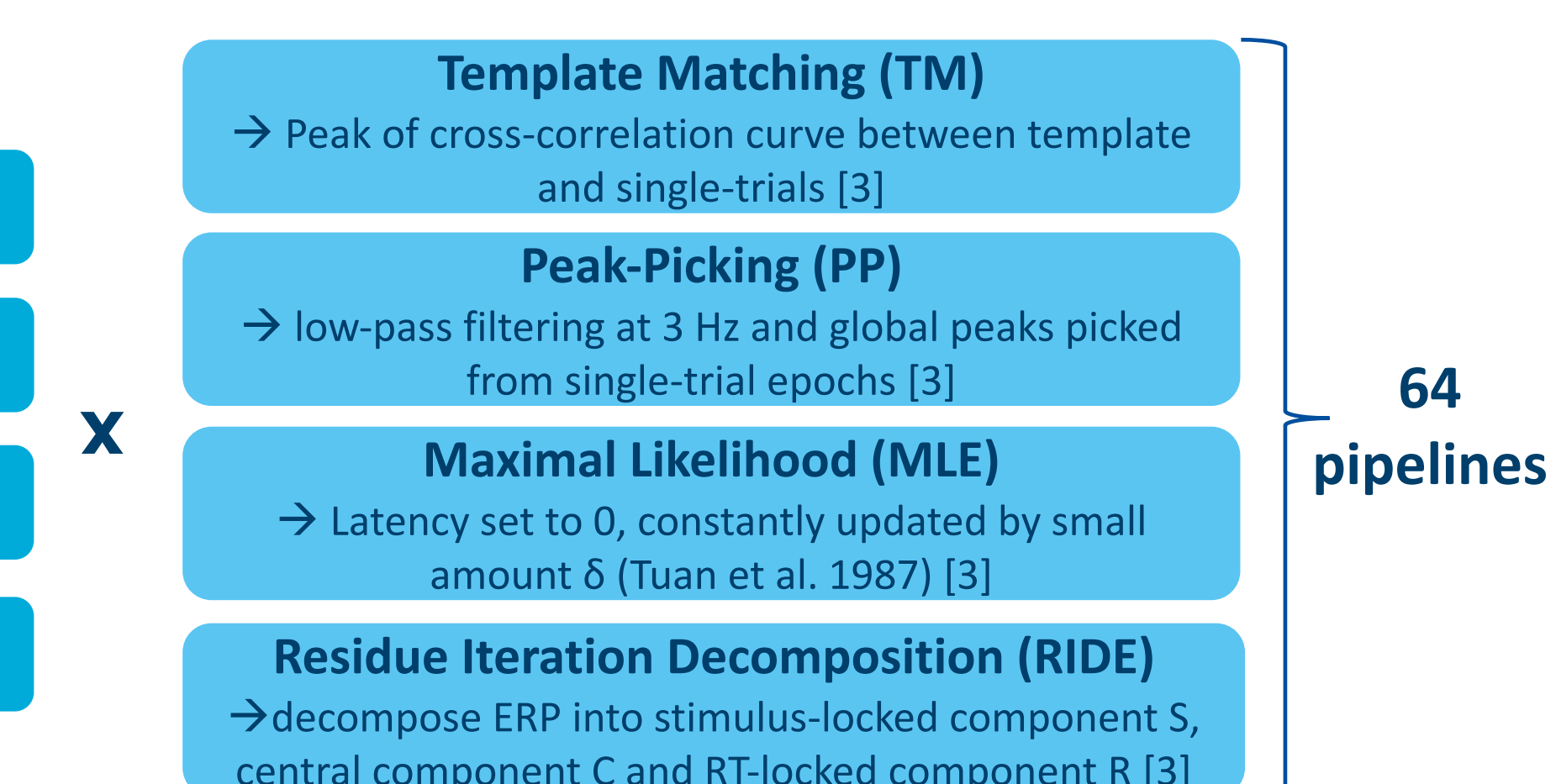
③ Theoretical plausibility

- Theoretical plausibility measured in terms of correlation between reaction time and latency estimates [3]
- *Cor.test()* function in R with Pearson's correlation coefficient method
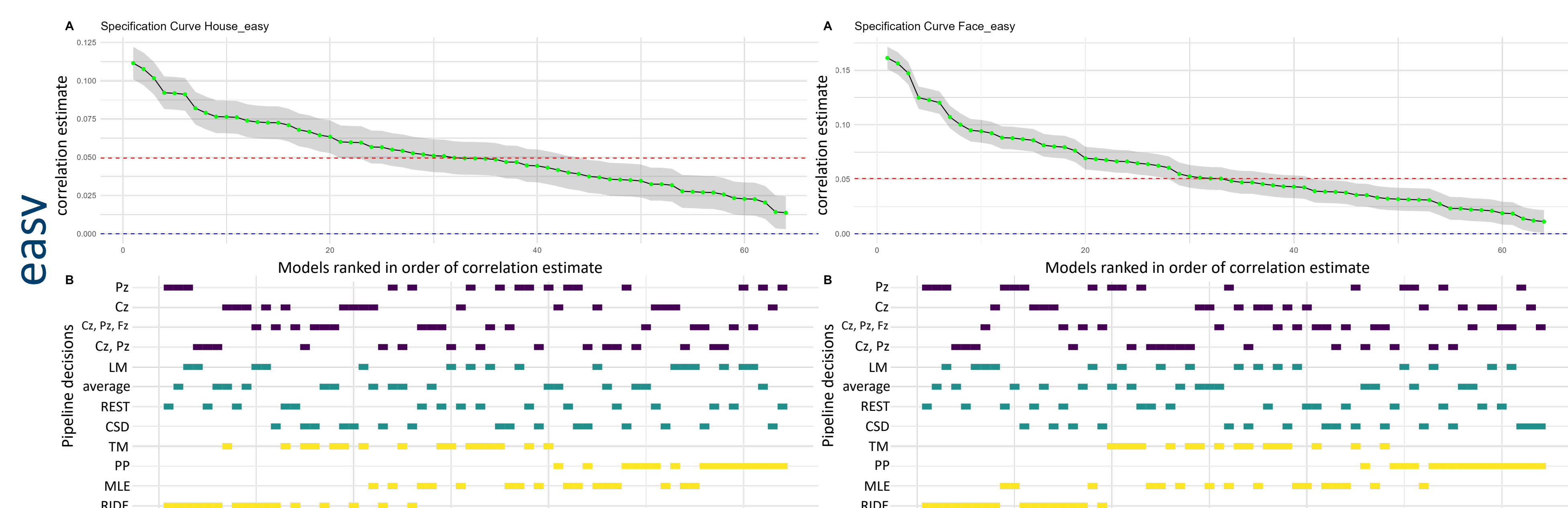
Results

House

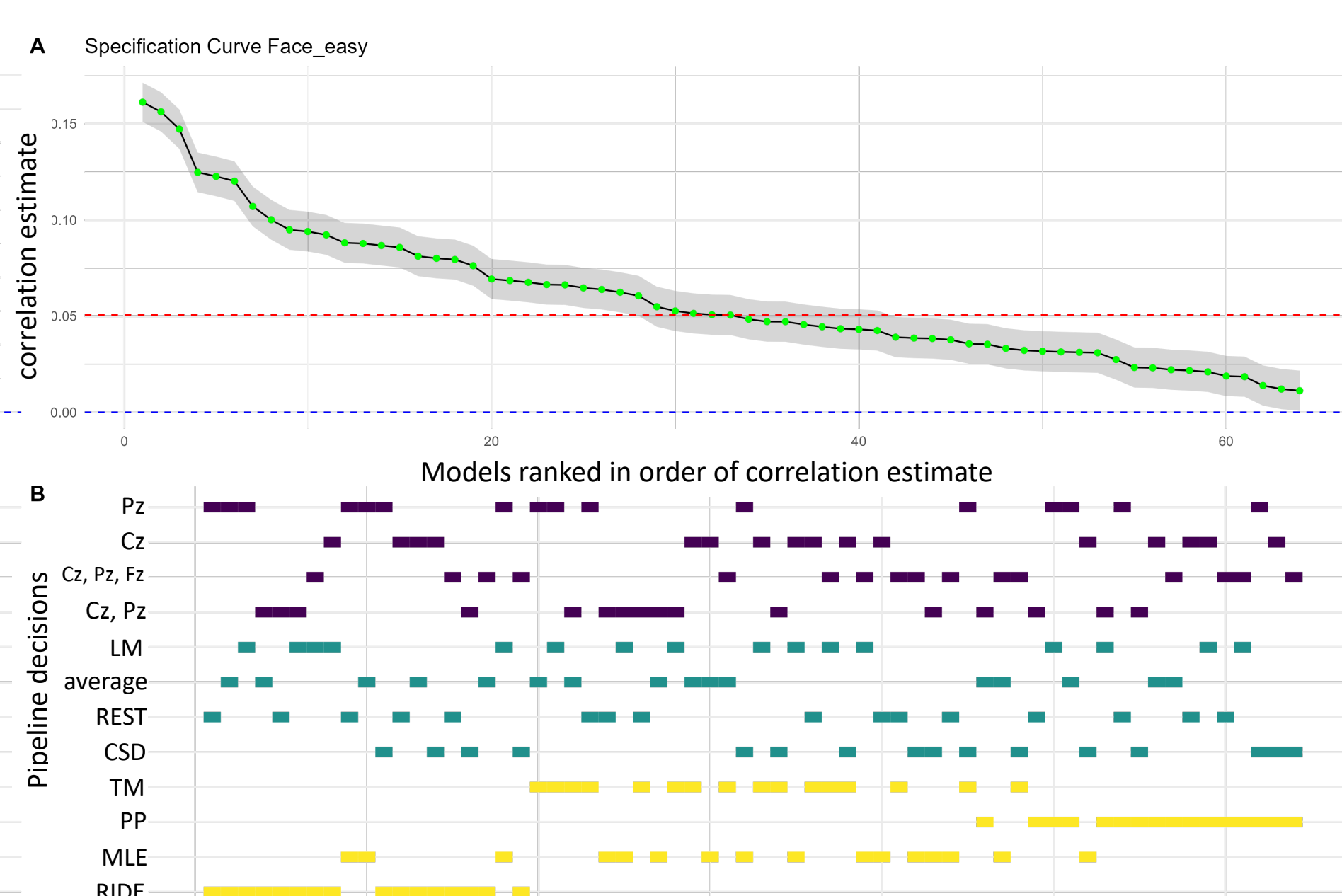**Fig. 3**: Specification curve in the condition *House easy*.

Fig. 5: Specification curve in the condition *House difficult*.
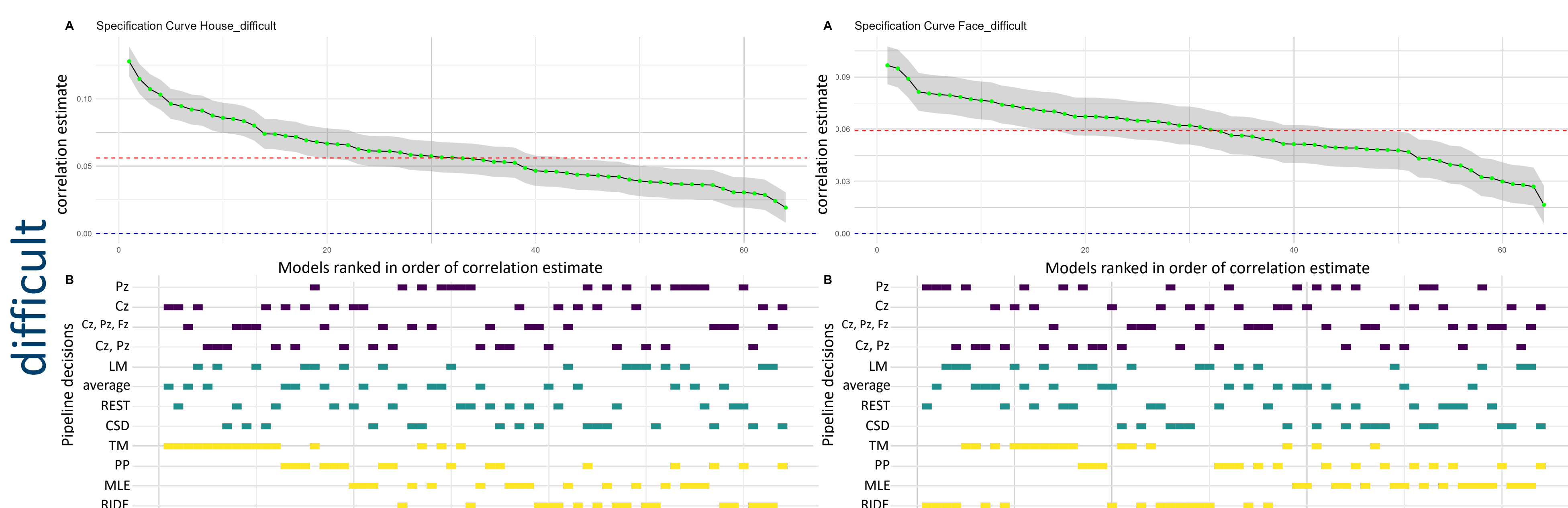
Face

Fig. 4: Specification curve in the condition *Face easy*.
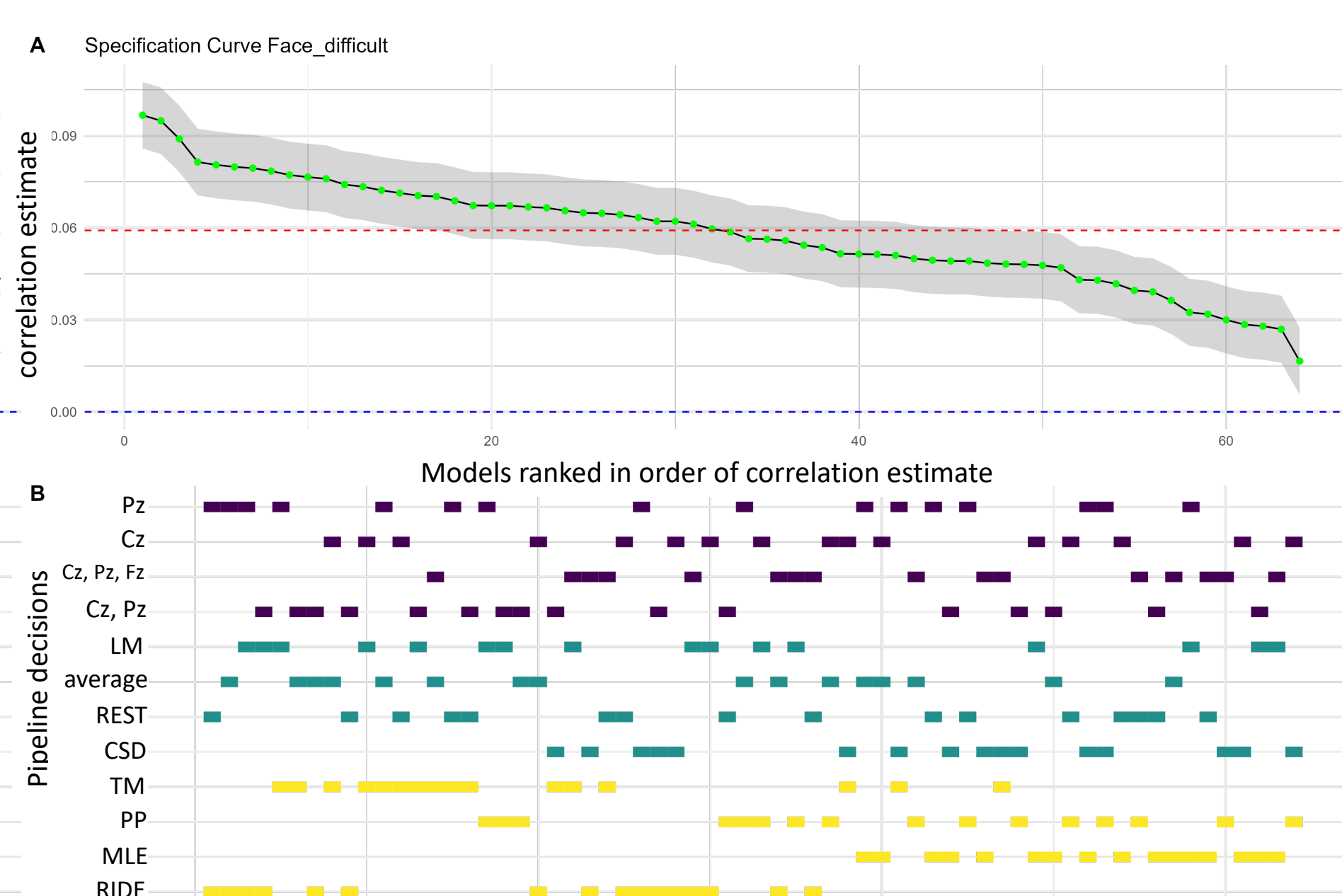
Fig. 6: Specification curve in the condition *Face difficult*.

Discussion

- No obvious pattern in theoretical plausibility across reference schemes and electrode groups
- RIDE leads to consistently highest correlation values in 3 of 4 conditions
- Template Matching leads to consistently high to medium plausibility values
- Peak-Picking and Maximal Likelihood Method are the methods with the lowest theoretical plausibility
- These findings are in line with the findings of Ouyang et al. (2017) [3]
- Single-trial P300 latency difficult to measure, possible that algorithms would perform differently in an easier component e.g. N170
- Difference in House difficult vs. Face difficult could be due to subjective difference in task difficultness, choosing a later time window for the cognitive component may lead to different results

Limitations & Outlook

- Small number of Quantification Decisions and single-trial algorithms due to timing
- Exploratory Multiverse: not all decisions are equally defendable
- Future studies could investigate whether these findings hold for other components and conditions

References:
[1] Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS medicine*, 2(8), e124.
[2] Del Giudice, M., & Gangestad, S. W. (2021). A traveler's guide to the multiverse: Promises, pitfalls, and a framework for the evaluation of analytic decisions. *Advances in Methods and Practices in Psychological Science*, 4(1), 2515245920954925.
[3] Ouyang, G., Hildebrandt, A., Sommer, W., & Zhou, C. (2017). Exploiting the intra-subject latency variability from single-trial event-related potentials in the P3 time range: a review and comparative evaluation of methods. *Neuroscience & Biobehavioral Reviews*, 75, 1-21.
[4] Rostami, H. N., Sommer, W., Zhou, C., Wilhelm, O., & Hildebrandt, A. (2017). Structural encoding processes contribute to individual differences in face and object cognition: Inferences from psychometric test performance and event-related brain potentials. *Cortex*, 95, 192- 210.
[5] Yao, J. (2001). A method to standardize a reference of scalp EEG recordings to a point at infinity. *Physiological measurement*, 22(4), 693.
[6] Cohen, M. X. (2014). *Analyzing neural time series data: theory and practice*. MIT press.

NEUROCOGNITIVE PSYCHOLOGY

Check out my GitHub repository

Contact: rosina.zollner@uol.de