Factor number selection in the tensor decomposition of EEG data:

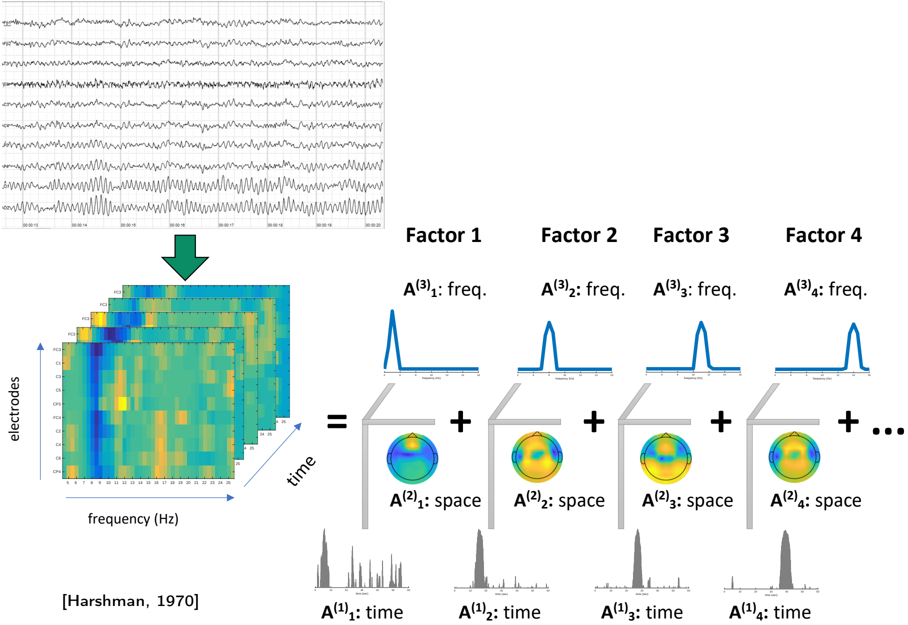
Mission (im)possible?

Zuzana Rošťáková, Roman Rosipal

Institute of Measurement Science
Slovak Academy of Sciences

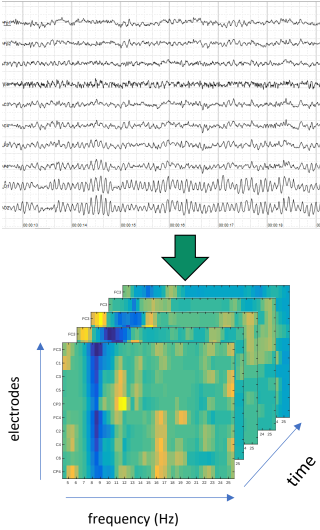
MEASUREMENT 2021

Parallel factor analysis (PARAFAC) and EEG

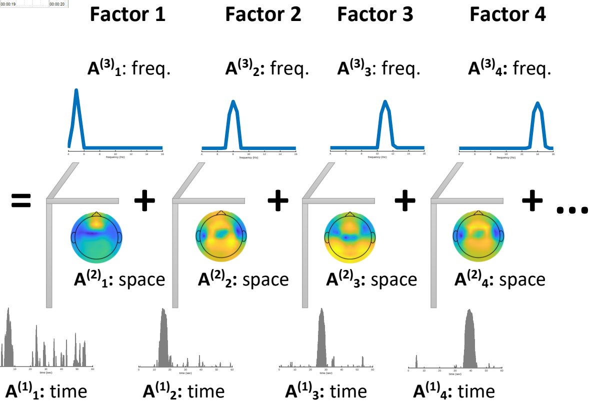


[Harshman, 1970]

Parallel factor analysis (PARAFAC) and EEG



number of factors F = ?



[Harshman, 1970]

Factor number selection methods

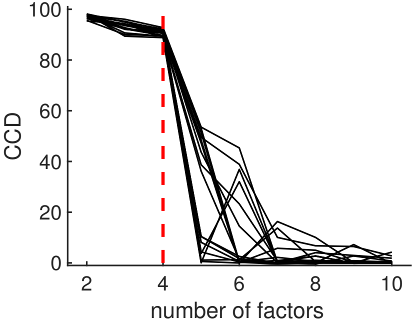
- **scree-plot**
  - Variance Explained (VarExpl)
  - **Core Consistency Diagnostics (CCD)**
  - Core consistency diagnostics aided by reconstruction error (CCDaRE)
  - ...
- **redundant factors**
  - **Non-redundant model order selection (NORMO)**
  - ...
- **model fit**
  - Difference in fit (DIFFIT)
  - Fast difference in fit (fastDIFFIT)
  - **Numerical convex hull (NCH)**
  - ...
- **eigenvalues**
  - **Minimal description length (MDL)**
  - Modified eigenvalue estimation for Tucker rank determination (MEET)
  - Sparse core method (SCORE)
  - ...
- **Bayesian statistics**
  - **Automatic relevance determination (ARD)**
  - ...
- **...**

Core consistency diagnostics (CCD)

- [Bro and Kiers, 2003]
- Parallel factor analysis (PARAFAC) models with $f = 1, \ldots, F_{max}$ factors

$$CCD(f) \in (-\infty, 100]$$

- an "elbow"/rapid change in the CCD decreasing profile

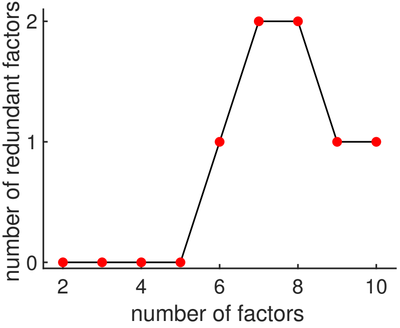


Non-redundant model order selection (NORMO)

- [Fernandes et al., 2020]
- PARAFAC models with $f = 1, \ldots, F_{max}$ factors

$$NORMO(f) = \text{ number of redundant factors } \rightarrow \text{ correlation } > 0.7$$

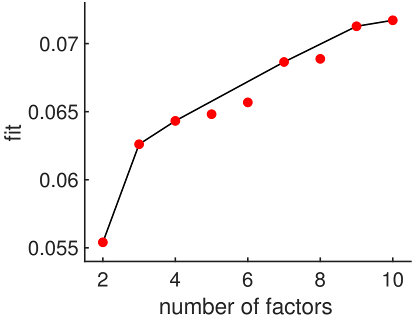
- **F** $\rightarrow$ the **largest** f with $NORMO(f) = 0$ and $NORMO(f+1) > 0$
- NORMO$_E$/NORMO$_B$ - all/selected PARAFAC models are considered



Numerical convex hull (NCH)

- [Ceulemans and Kiers, 2006]
- PARAFAC models with $f = 1, \ldots, F_{max}$ factors
- maximal change in fit between models with a consecutive number of factors

→ only models, which fit belongs to the fit convex hull boundary, are considered

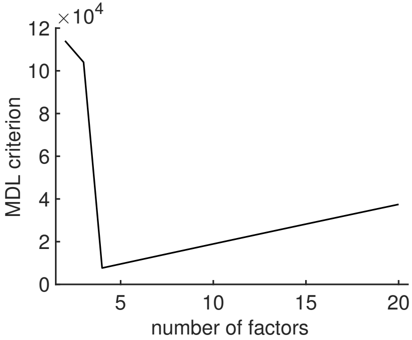


Minimal description length (MDL)

- [Liu et al., 2016]
- eigenvalues of matricized versions of the tensor

⇒ "indirect" assumption of factors orthogonality

- $F \in \text{argmin}_{f \in \{2,\dots,F_{max}\}}$ MDL(f)



Automatic relevance determination (ARD)

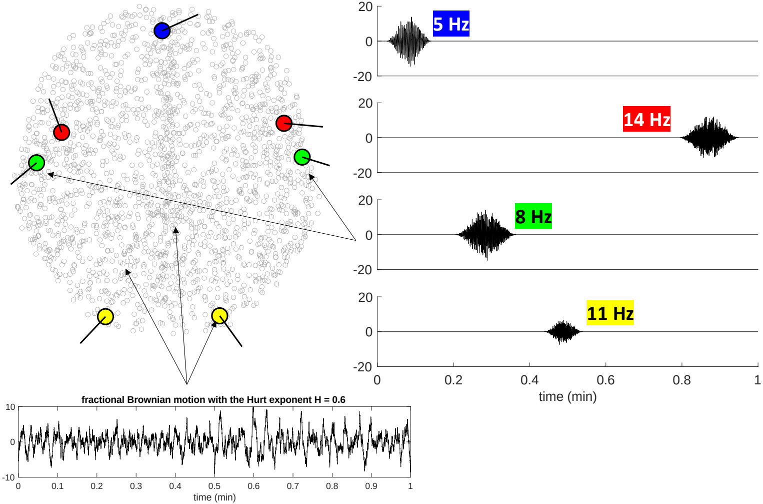
- [Mørup and Hansen, 2009]
- factor elements are assumed to follow
  - **normal** distribution $\rightarrow$ **ridge** version of ARD (ARD_R)
  - **Laplacian** distribution $\rightarrow$ **sparse** version of ARD (ARD_S)
- **iterative algorithm:**
  - k^{th} step: PARAFAC model with F_k factors is estimated

    $\rightarrow$ factors with lowest weights are pruned out

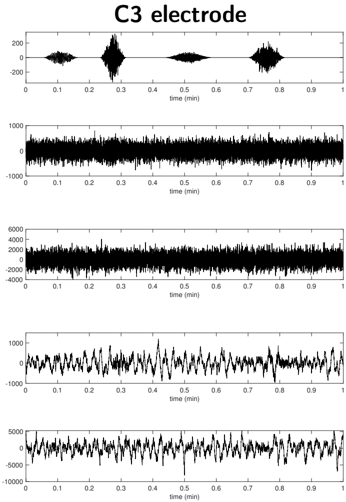
    $\Rightarrow$ number of factors decreases to $F_{k+1} \leq F_k$
- modified selection criterion

Simulated EEG data

S_{EEG}	$=$	C	$\times$	$S_{\text{source signal}}$
$64 \times T$		64×2004		$2004 \times T$ (number of time points)
EEG signal		forward model		oscillations + broadband brain activity



Simulated EEG data



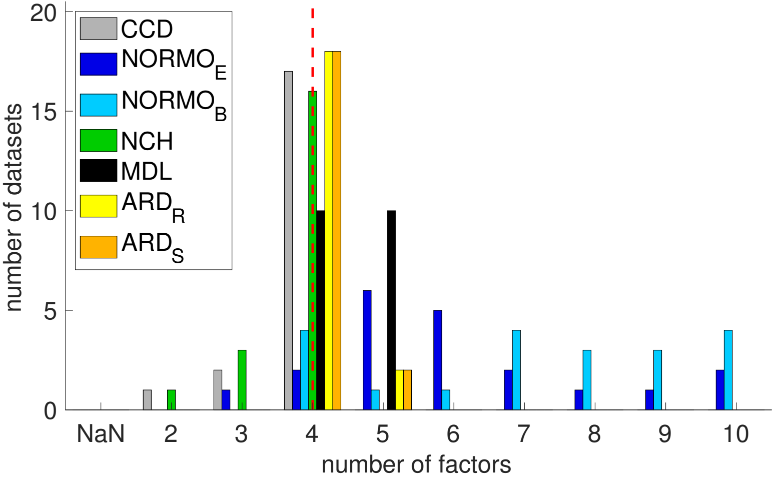
- N_0 - only 4 oscillations
  5 Hz, 8 Hz, 11 Hz, 14 Hz
- NG_{low} - N_0 + gaussian noise $\mathcal{N}(0, 200)$
- NG_{high} - N_0 + gaussian noise $\mathcal{N}(0, 1000)$
- $NBBA_{low}$ - 4 oscillations and "low" BBA
- $NBBA_{high}$ - 4 oscillations + "high" BBA

$\rightarrow$ 20 sets for each type of data

Results

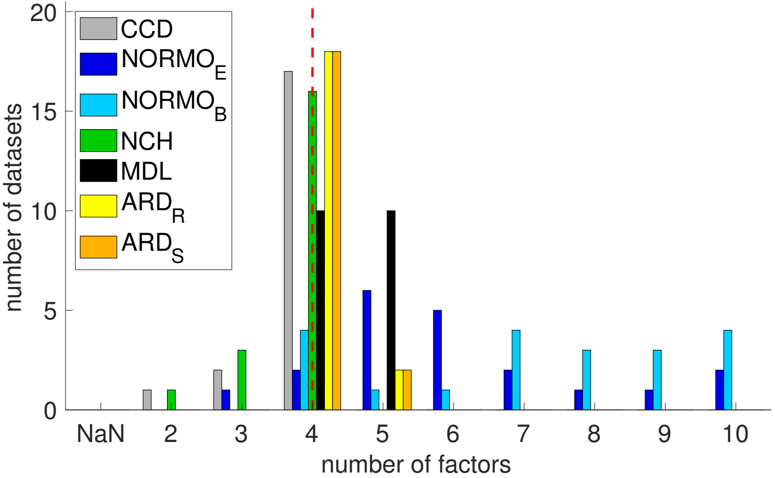
Mission 1 - N_0 data

CCD **NORMO_E, NORMO_B** **NCH** **MDL** **ARD_R, ARD_S**



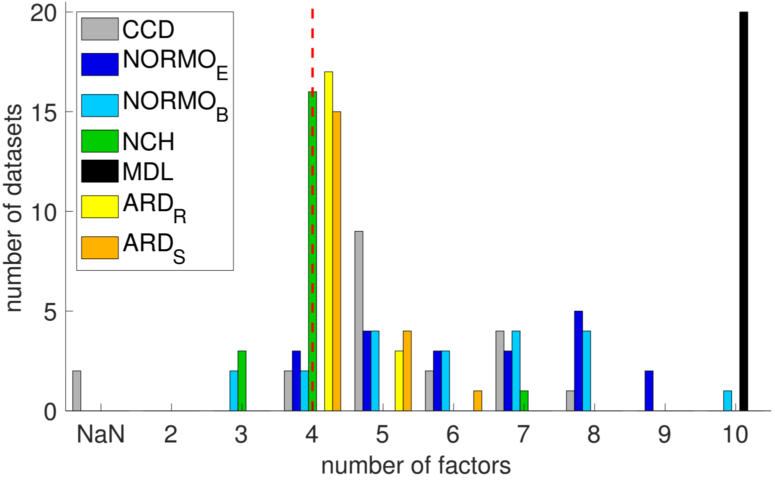
Mission 1 - N_0 data

CCD **$NORMO_E$, $NORMO_B$** **NCH** **MDL** **ARD_R, ARD_S**



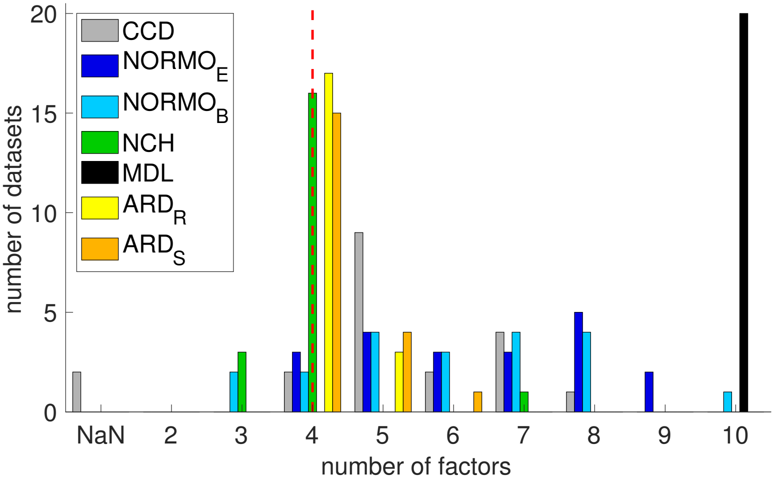
Mission 2 - NG_{low} data

CCD | NORMO$_E$, NORMO$_B$ | NCH | MDL | ARD$_R$, ARD$_S$

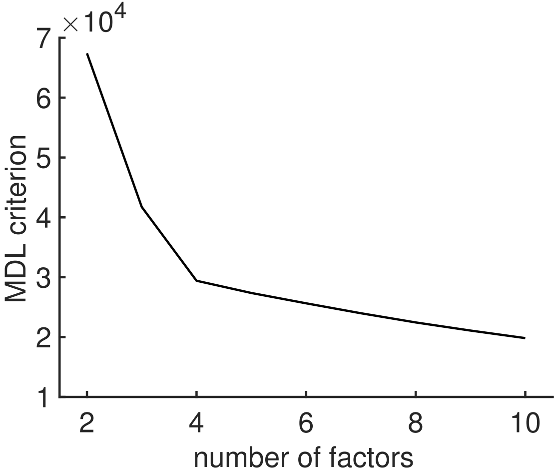


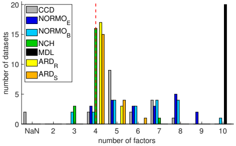
Mission 2 - NG_{low} data

CCD **NORMO_E, NORMO_B** **NCH** **MDL** **ARD_R, ARD_S**



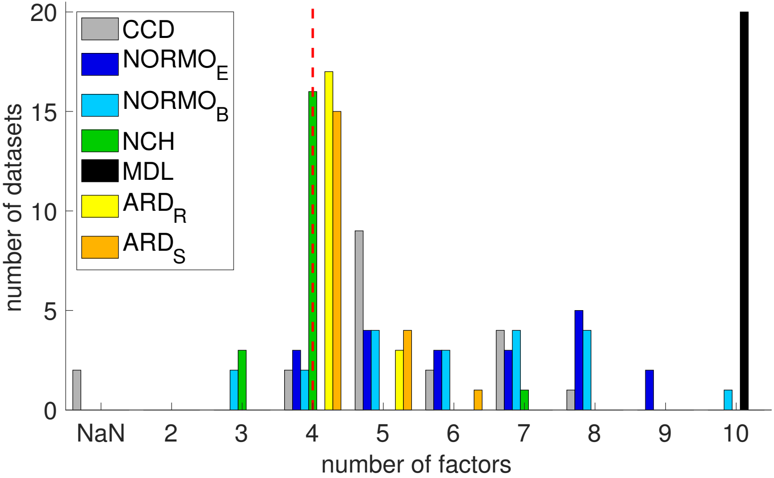
CCD **NORMO$_E$, NORMO$_B$** **NCH** **MDL** **ARD$_R$, ARD$_S$**





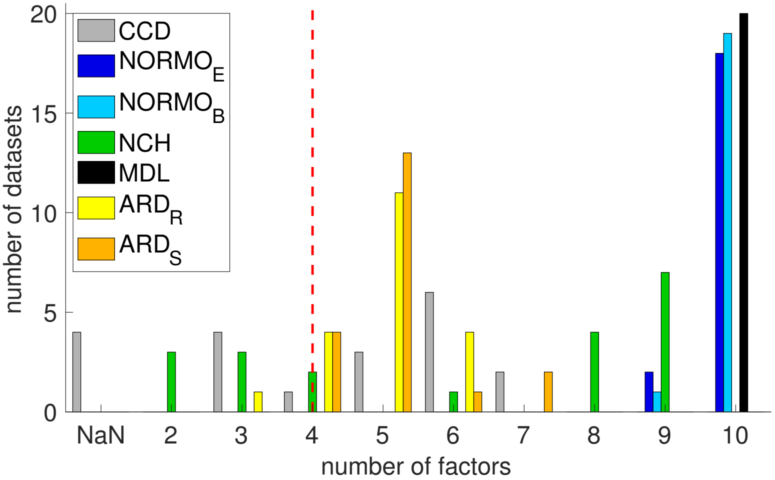
Mission 2 - NG_{low} data

CCD | NORMO$_E$, NORMO$_B$ | NCH | MDL | ARD$_R$, ARD$_S$

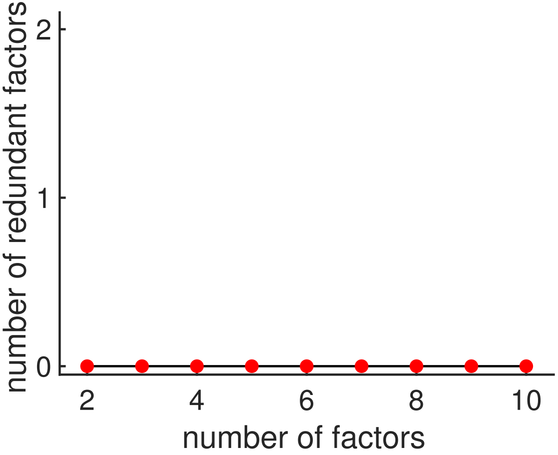


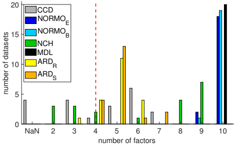
Mission 3 - NG_{high} data

CCD | NORMO$_E$, NORMO$_B$ | NCH | MDL | ARD$_R$, ARD$_S$



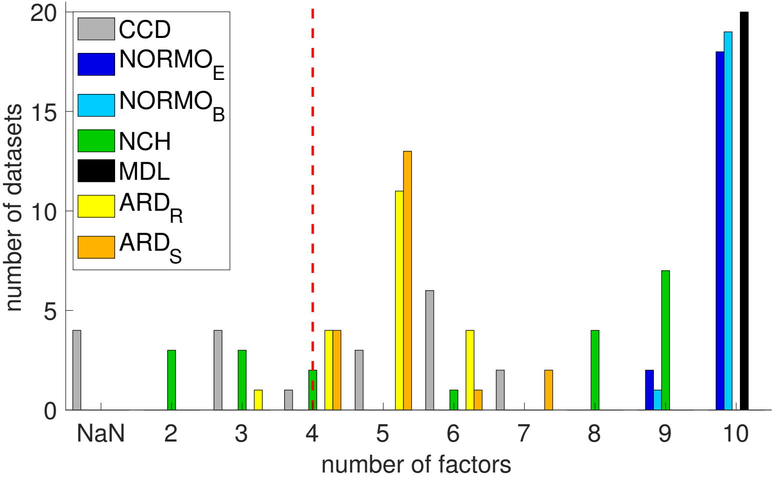
CCD NORMO$_E$, NORMO$_B$ NCH MDL ARD$_R$, ARD$_S$





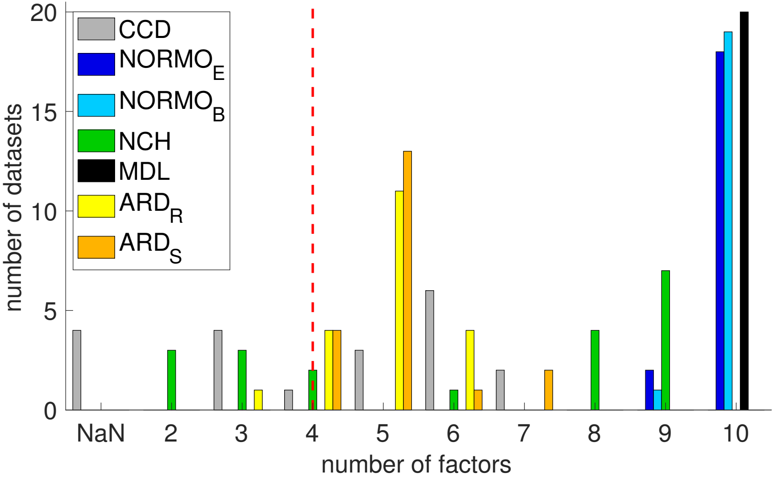
Mission 3 - NG_{high} data

CCD NORMO$_E$, NORMO$_B$ NCH MDL ARD$_R$, ARD$_S$



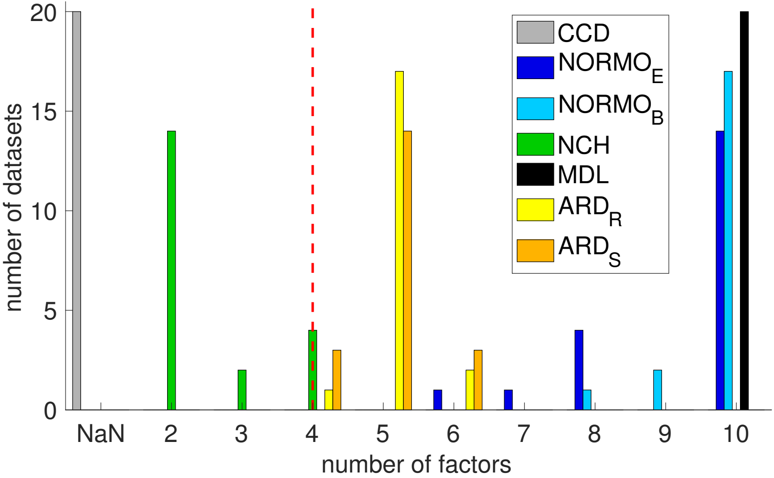
Mission 3 - NG_{high} data

CCD NORMO$_E$, NORMO$_B$ NCH MDL ARD$_R$, ARD$_S$



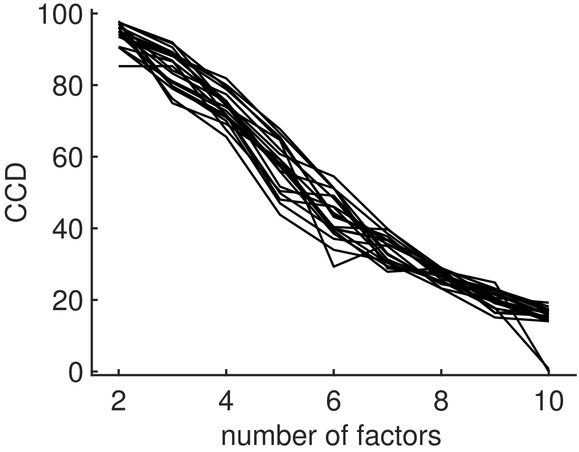
Mission 4 - $NBBA_{low}$ data

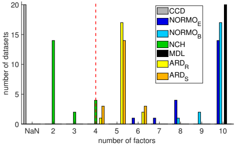
CCD NORMO$_E$, NORMO$_B$ **NCH** MDL **ARD$_R$, ARD$_S$**



Mission 4 - $NBBA_{low}$ data

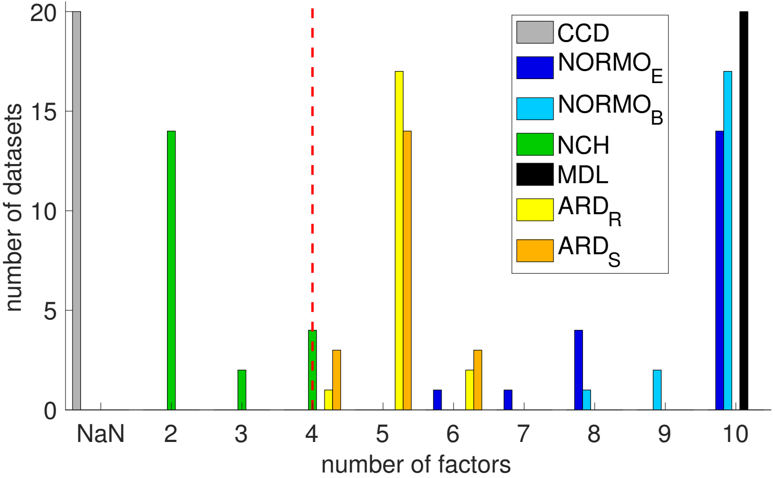
CCD NORMO$_E$, NORMO$_B$ NCH MDL ARD$_R$, ARD$_S$





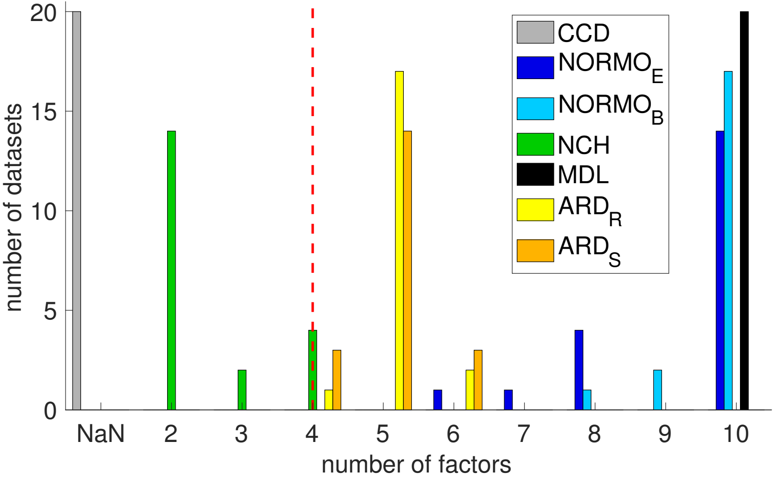
Mission 4 - $NBBA_{low}$ data

CCD $NORMO_E$, $NORMO_B$ NCH   MDL ARD_R, ARD_S



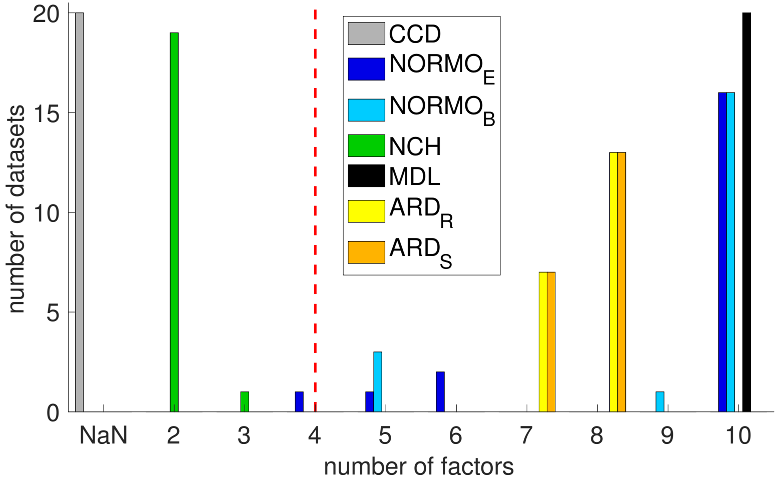
Mission 4 - $NBBA_{low}$ data

CCD NORMO$_E$, NORMO$_B$ NCH MDL ARD$_R$, ARD$_S$



Mission 5 - $NBBA_{high}$ data

CCD NORMO$_E$, NORMO$_B$ NCH MDL **ARD$_R$, ARD$_S$**



Conclusion: Mission impossible ?

$\Rightarrow$ none method produced satisfactory results for PARAFAC decomposition of simulated EEG data

- **NORMO, MDL** - problems already for noiseless data
- **CCD** - no visible change-point in CCD due to the gaussian noise/BBA
- **NCH** - data with BBA $\rightarrow F \approx$ minimal allowed value 2
- **ARD** - the most promising method $\rightarrow$ modification in the future

$\rightarrow$ new approaches are needed

Literature

Bro, R. and Kiers, H. A. L. (2003).
A new efficient method for determining the number of components in PARAFAC models.
Journal of Chemometrics, 17(5):274–286.

Ceulemans, E. and Kiers, H. A. L. (2006).
Selecting among three-mode principal component models of different types and complexities: A numerical convex hull based method.
British Journal of Mathematical and Statistical Psychology, 59(1):133–150.

Fernandes, S., Fanaee-T, H., and Gama, J. (2020).
NORMO: A new method for estimating the number of components in CP tensor decomposition.
Engineering Applications of Artificial Intelligence, 96.

Harshman, R. A. (1970).
Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multimodal factor analysis.
UCLA Working Papers in Phonetics, 16:1–84.

Liu, K., da Costa, J., So, H. C., Huang, L., and Ye, J. (2016).
Detection of number of components in CANDECOMP/PARAFAC models via minimum description length.
Digital Signal Processing, 51:110 – 123.

Mørup, M. and Hansen, L. K. (2009).
Automatic relevance determination for multi-way models.
Journal of Chemometrics, 23(7-8):352–363.

Rošťáková, Z., Rosipal, R., Seifpour, S., and Trejo, L. J. (2020).
A comparison of non-negative Tucker decomposition and Parallel Factor Analysis for identification and measurement of human EEG rhythms.
Measurement Science Review, 20(3):126 – 138.