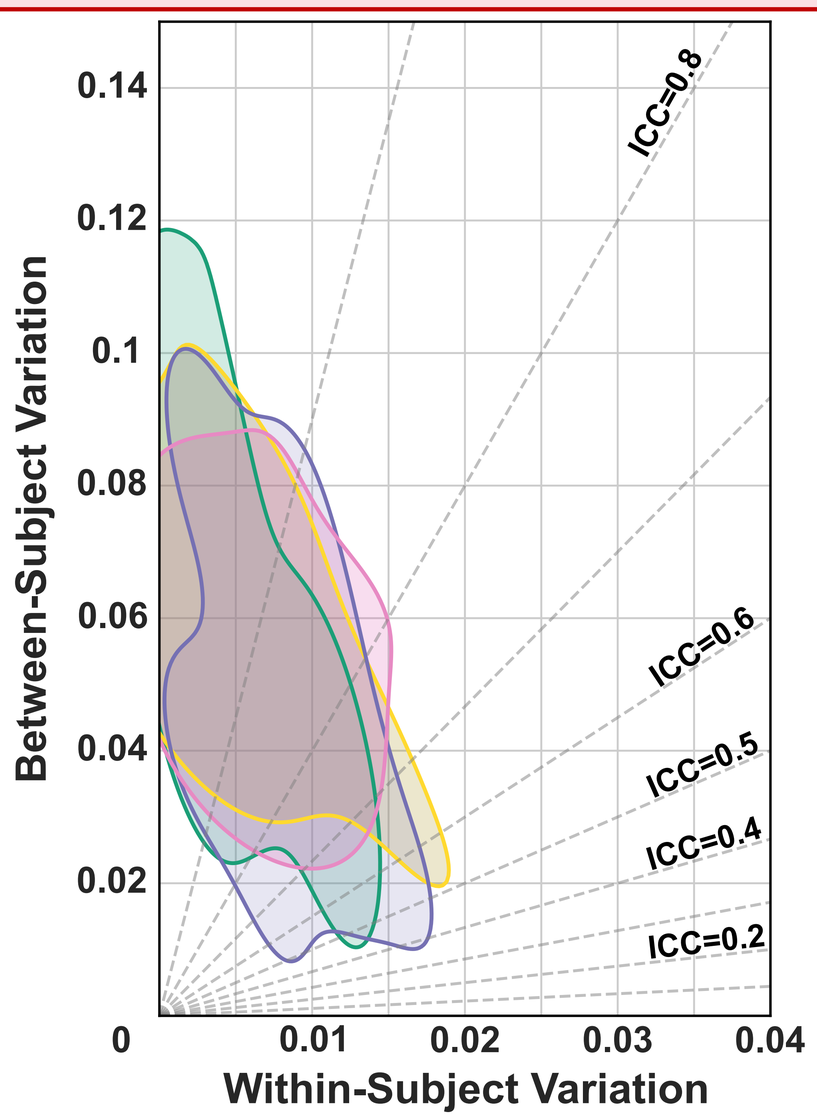
Reliability of cerebellar volumetry : a test / re-test study

Katia Chardon[1,*], Davide Boido[1], Charles Laidi[2,*]

[1]NeuroSpin, BAOBAB, CEA, CNRS, Université Paris-Saclay, Gif-sur-Yvette, France
[2]Univ. Paris Est- Créteil, Hôpitaux Universitaire Henri Mondor, INSERM & NeuroSpin CEA Paris, France
*katia.chardon@cea.fr *charles.laidi@cea.fr

Introduction

Accurate and reliable quantification of the cerebellum's structure present significant challenges for automated parcellation because of its foliated structure and fine subregional organization. Several tools, including traditional atlas-based approaches and recent deep learning methods, have been developed to segment and parcellate the cerebellum. While these tools are typically evaluated based on segmentation accuracy, their test-retest reliability, essential for longitudinal and individual difference studies, has seen limited comparison.



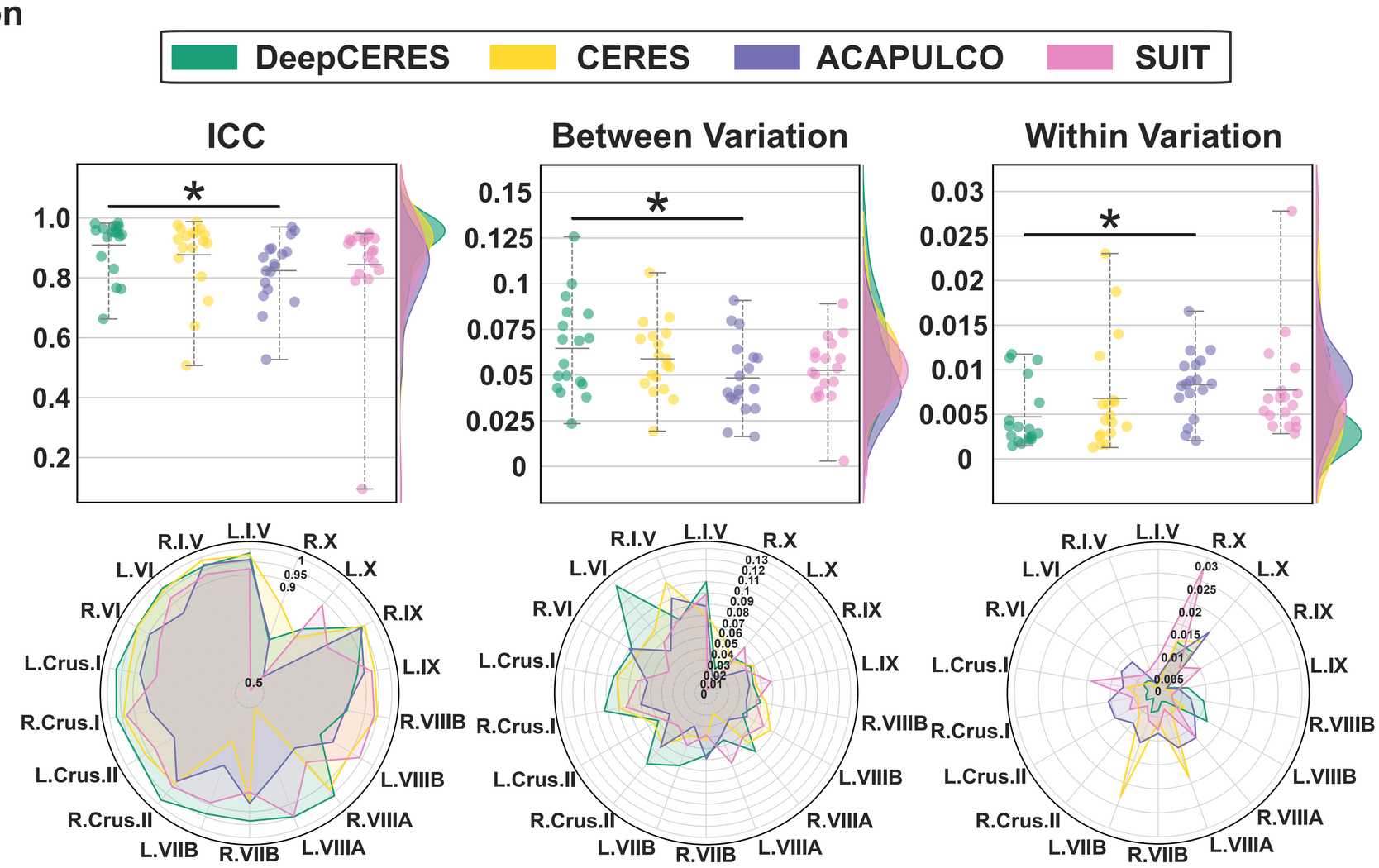
Methods

We assessed the test-retest reliability of four cerebellar segmentation pipelines : **SUIT-VBM** [1], **ACAPULCO** [2], **CERES** [3], and **DeepCERES** [4]. We used high-resolution T1-weighted images from the **Hangzhou Normal University open dataset (HNU)** [5], which includes 10 MRI sessions for healthy participants scanned every three days over one month. We selected **11 participants**, including 5 females and 6 males of mean age 24.1 ± 2.7 years.

Volumes of cerebellar subregions were extracted with each method, normalized by intracranial volume to account for head size, and scaled across lobules to correct for size differences. We then computed **intra-individual variability, inter-individual variability, and Intraclass Correlation Coefficients (ICC)** using the **ReX toolbox** [6].

Results

- **All pipelines exhibited strong test-retest reliability**, with average ICC values above 0.8.
- **DeepCERES was the most robust method**, significantly more reliable than ACAPULCO ($p < 0.05$), showing high ICC values, high inter-individual variability, and low intra-individual variability.
- **Lobules X and VIIIB displayed lower reliability** across all methods.



Conclusion

Our results inform the state-of-the-art in cerebellum parcellation. We provide guidance for selecting the most reliable parcellation method and point out the most challenging lobules for this analysis.

References

1. Diedrichsen, J. (2006). A spatially unbiased atlas template of the human cerebellum. Neuroimage. 33(1), 127-138.
2. Han, S., et al. (2020). Automatic cerebellum anatomical parcellation using U-Net with locally constrained optimization. NeuroImage, 218, 116819.
3. Romero, J. E., et al. (2017). CERES: A new cerebellum lobule segmentation method. NeuroImage, 147, 916–924.
4. Morell-Ortega, S., et al. (2025). DeepCERES: A deep learning method for cerebellar lobule segmentation using ultra-high resolution multimodal MRI. NeuroImage, 308, 121063.
5. Zuo, X. N., et al. (2014). An open science resource for establishing reliability and reproducibility in functional connectomics. Scientific data, 1, 140049.
6. Xu, T., et al. (2023). ReX: an integrative tool for quantifying and optimizing measurement reliability for the study of individual differences. Nature methods, 20(7), 1025–1028.