Individualized Brain Morphometry Through User-Controlled Normative Database

Campos, B.M.[1,2], Casseb, R.F.[1], Pimentel-Silva, L.R.[1], Alvim, M.K.M.[1], Yasuda, C.L.[1], Paschoal, A.M.[2], Cendes, F.[1,]

[1] Neuroimaging Laboratory, Faculty of Medical Science, University of Campinas, Brazil
[2] Lab. of Advanced MRI, Gleb-Wataghin Physics Institute, University of Campinas, Brazil

Supplementary Material

1. SSM Graphical User-Interface (GUI)

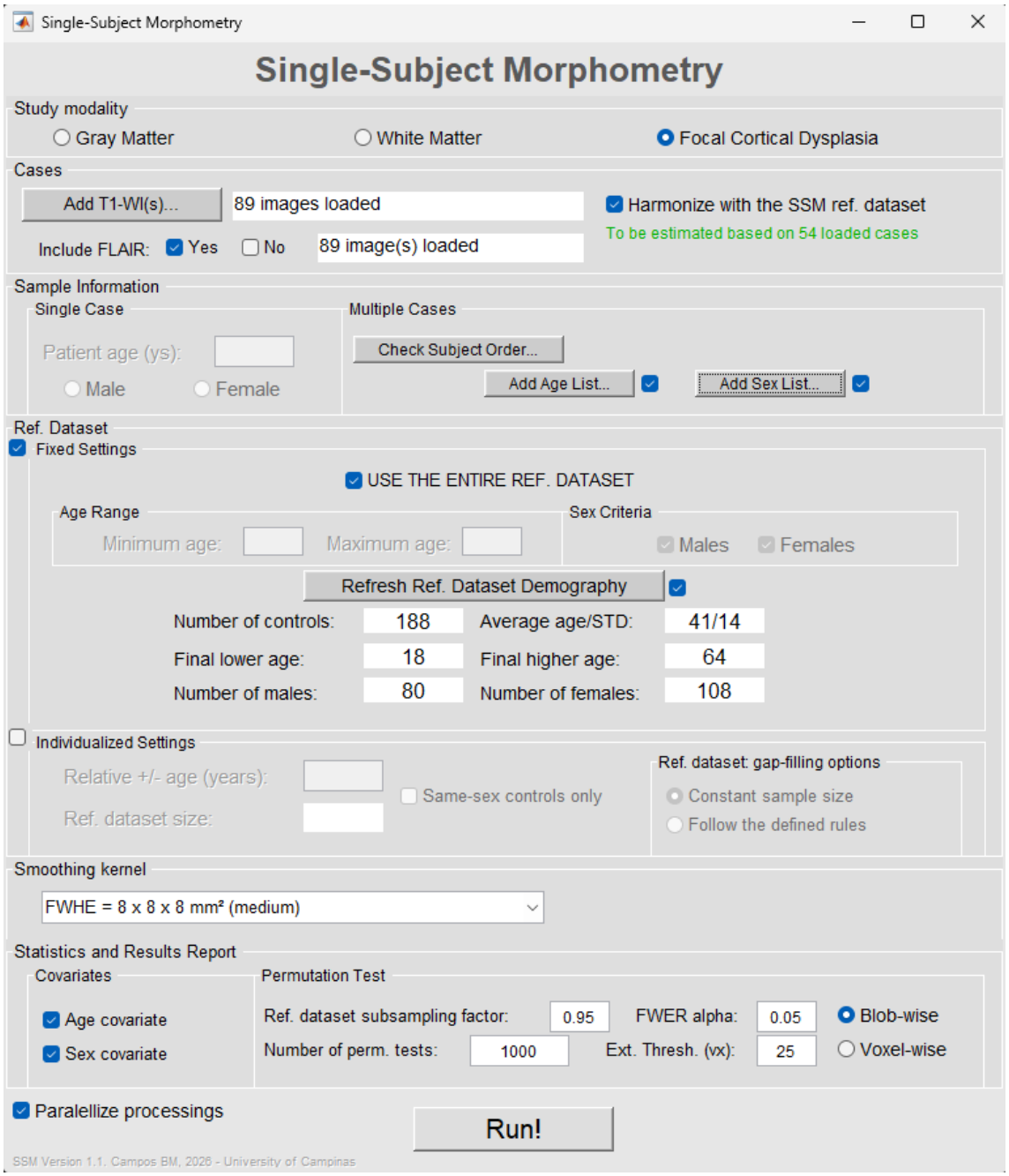

Figure 1: SSM graphical user-interface.

2. Procedures for FLAIR Image Inclusion

FLAIR images are preprocessed using dedicated procedures implemented exclusively with SPM functions, and subsequently combined with the T1-WI segmented maps to generate final gray and white matter maps influenced by FLAIR-derived signal patterns. Briefly, FLAIR images are first registered to the corresponding T1-WI images using rigid-body transformations. They are then segmented and non-linearly normalized into the T1-WI space using the T1-derived tissue segmentations as probabilistic priors. Finally, the resulting FLAIR-derived segmented maps are normalized (and modulated) to the postprocessed gray and white matter spaces (MNI-space) by applying the deformation fields estimated during the T1-WI CAT12 processing.

The final FLAIR-derived GM and WM maps are subsequently merged with the T1-derived maps according to the following voxel-wise rules:

- For GM, the final voxel intensity is defined as the maximum value between the T1-derived and FLAIR-derived maps.
- For WM, the final voxel intensity is defined as the minimum value between the T1-derived and FLAIR-derived maps.

3. Methodological considerations

- On the SSM permutation approach: although subsampling introduces partial overlap between iterations, this overlap is substantially lower than that observed in a standard leave-one-out scheme. The use of unique subsamples increases the diversity of the sampled configurations, improving the approximation of the null distribution while maintaining a consistent reference pool. The choice of a 95% subsampling fraction reflects a trade-off between increasing permutation diversity and preserving the statistical properties of the full reference dataset. Lower subsampling fractions would increase variability across permutations but at the cost of reducing the effective sample size and altering the underlying statistical model. Therefore, a high subsampling ratio was adopted to maintain consistency with the full reference distribution while still expanding the permutation space.

4. SSM harmonization procedures

This section illustrates, using *Figure 1*, the pattern of findings and the limitations of the different SSM processing approaches for the FCD modality.

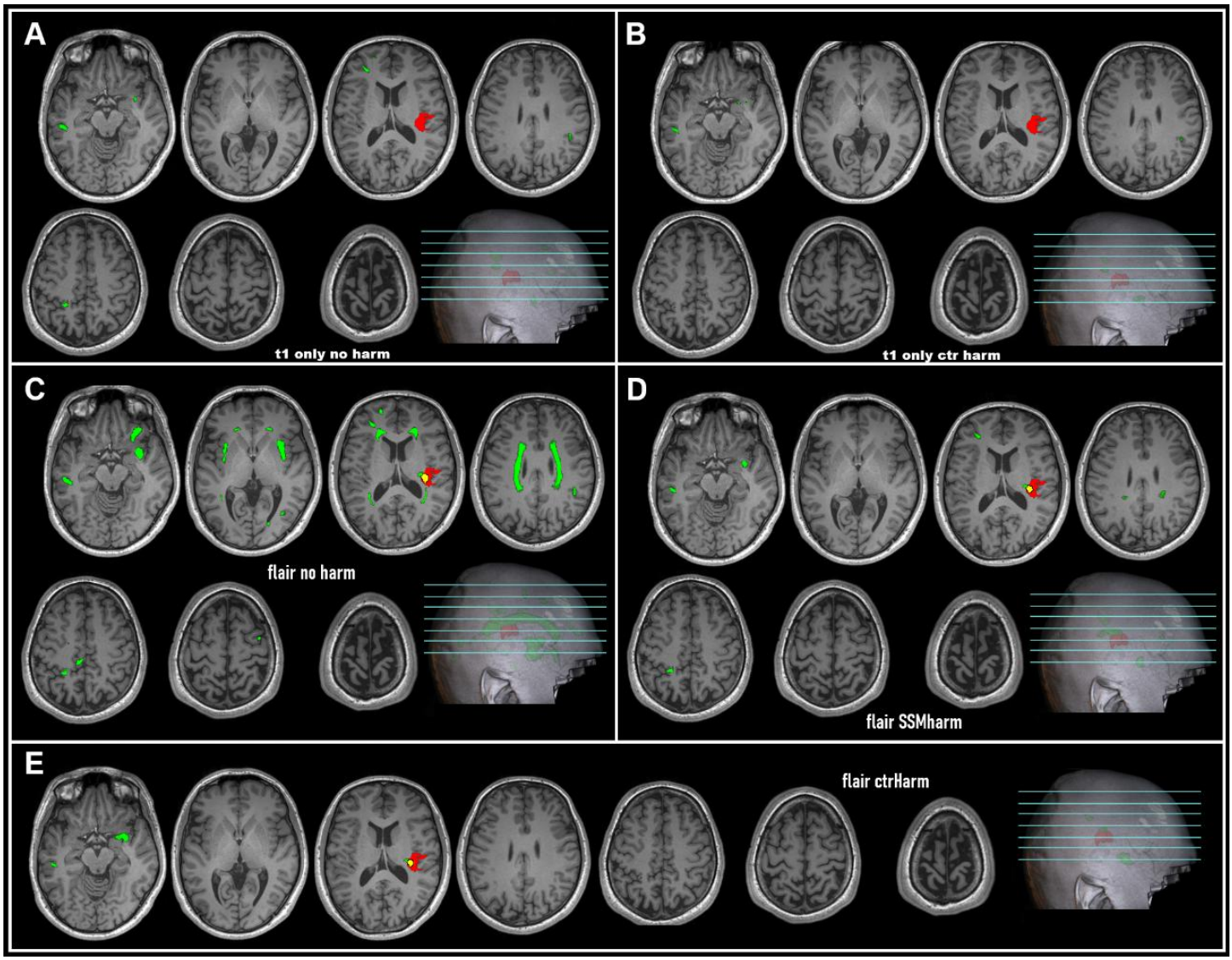

Figure 2. Comparison of FCD detection results across different SSM approaches and harmonization procedures. Red ROIs represent manually labeled FCD regions, green ROIs indicate the SSM results, and yellow represents the overlap (true positive) between them. Panel (A) shows the SSM result using only T1WI without any harmonization procedure (resulting in no lesion overlap). Panel (B) shows the result using only T1WI but applying a control-based harmonization—representing the optimal harmonization scenario for T1-only data—which still yields no lesion overlap. Panel (C) demonstrates the impact of adding FLAIR images without any harmonization; a cluster precisely overlaps the lesion (yellow), but a widespread pattern of FLAIR-induced false-positive clusters is introduced. Panel (D) shows the results after adding FLAIR and applying the SSM included harmonization parameters, where harmonization parameters were estimated by calibrating local site controls (with FLAIR) against the default SSM database (controls without FLAIR); this mitigates non-idiosyncratic FLAIR artifacts while preserving the lesion-overlapping cluster. Finally, panel (E) illustrates the ideal scenario: combining T1WI and FLAIR images with a proper site-specific harmonization calculated from local controls with FLAIR, demonstrating subtle improvements and a cleaner background.

5. Computational Hardware and Performance

- Machine 1: MELD-Graph was tested on a Linux workstation running Ubuntu 20.04.6 LTS, equipped with a 12th-generation Intel® Core™ i7-12700KF processor (3.60 GHz), 128 GB

RAM, and an NVIDIA GeForce RTX A600 GPU (48 GB VRAM). Data were processed using an M.2 NVMe solid-state drive (SSD).

- Machine 2: AID-HS was tested on a Linux workstation running Ubuntu 22.04.5 LTS, equipped with an 8th-generation Intel® Core™ i7-8700K processor (3.70 GHz), 32 GB RAM, and an NVIDIA GeForce GTX 1050 GPU (4 GB VRAM). Data were processed using a SATA SSD.

SSM was evaluated on both workstations described above, as these represented either the required or the most suitable hardware configurations available for running the MELD tools. AID-HS was not compatible with the newer laboratory workstations and GPU architectures and could only be successfully installed and executed in the older Machine 2. To further assess cross-platform compatibility, SSM was additionally tested on a Windows 11 Pro workstation (Machine 3) equipped with a 12th-generation Intel® Core™ i7-12700KF processor (3.60 GHz), 32 GB RAM, and an NVIDIA GeForce RTX 3060 GPU (12 GB VRAM), using an M.2 NVMe SSD.

For the FCD modality, processing a single subject with SSM required 20 minutes on Machine 1 when all processing steps were executed from scratch, including CAT preprocessing and excluding any previously cached results. Under the same conditions and using the same image and hardware, MELD-Graph required 3 hours and 3 minutes. For the GM modality, processing a single subject with SSM required approximately 27 minutes on Machine 2, whereas AID-HS required approximately 25 minutes using the same image and hardware.

Performance on Machine 3 (Windows) was comparable to that observed on Linux systems. Processing a single subject required approximately 25 minutes for the GM modality and 35 minutes for the FCD modality, including FLAIR image processing and the SSM-integrated FLAIR harmonization procedure.

6. Blob-Wise FWER Option

As a robust alternative to standard single-voxel maximum statistics (T_{max}) in permutation testing, a local spatial summary metric ("blob-wise FWER") was implemented as an alternative option. In each permutation iteration, upon identifying the global maximum voxel, a spherical region of interest (ROI) with a radius equal to half the smoothing kernel size, rounded up to the nearest integer (Eq. 1) is centered on this peak

$$R = \left\lceil \frac{FWHM}{2} \right\rceil \quad \text{Eq. 1}$$

The summary statistic for the permutation distribution is then computed as the median value of all non-zero voxels within this sphere. This local neighborhood (Eq. 1) directly reflects the spatial autocorrelation induced by the smoothing filter, encompassing the volume where spatial dependency remains high ($\rho \geq 0.71$). Consequently, evaluating the local median incorporates spatial coherence into the permutation threshold, rather than relying on single-voxel peak intensity alone. This reduces sensitivity to isolated high-amplitude variance spikes or boundary artifacts while maintaining strict non-parametric FWER control for the spatial cluster metric.

7. Final sample distribution and main result

Table 1. Final sample sizes and main results for each tested modality and tools.

	SSM		MELD/AID-HS	
	GM (HS)	**FCD**	**GM (HS)**	**FCD**
Initial Totals	211	112	211	112
Data problems	0	7	0	7
Processing Fails	0	2	0	6
Final Totals	211	103	211	99

Wrong results	13	31	9	27
Correct results	198	72	202	72
Precision (%)	94%	70%	96%	73%

GM: grey matter; HS: hippocampal sclerosis; FCD: focal cortical dysplasia;

8. Alternative MELD-Graph Harmonization Procedure

To evaluate MELD-Graph under the most favorable conditions, we additionally tested an alternative harmonization strategy for the *Bonn-Batch-1* dataset. Instead of using only healthy controls to estimate the harmonization parameters, as described in the main manuscript, the harmonization procedure was performed using the entire cohort (healthy controls and patients). As expected, this strategy resulted in a slightly lower performance.

MELD_Scores			
Bonn		**UCLH + Bonn**	
Total:	75	ALL Total:	101
Errors:	22	Errors:	28
Hits:	53	Hits:	73
Hits %	70.67%	Hits %	72.28%

Table 2. Performance of MELD-Graph using the alternative harmonization procedure for the *Bonn_Batch_1* dataset. The overall detection accuracy was 71%, compared with 72% obtained using the harmonization strategy described in the main text. MELD-Graph final score of 72% (considering *Bonn-* and *IDEAS-dataset*, instead of 73%)

Interestingly, the alternative harmonization strategy substantially changed the pattern of correctly and incorrectly classified subjects. Although the overall performance decreased by only one correctly detected case (54 to 53), nearly 30% of the failed cases (22) differed between the two approaches. In other words, several cases that were correctly localized using the harmonization strategy adopted in the main manuscript became false negatives with the alternative approach, while other previously missed cases were correctly detected. This finding suggests that the harmonization procedure has a considerable impact on the stability of individual predictions, even when the overall performance remains nearly unchanged.