Numerical Uncertainty in Linear Registration: An Experimental Study

*Niusha Mirhakimi[1[0009−0005−9558−1817]], Yohan Chatelain[2[0000−0001−7023−250X]], †Jean-Baptiste Poline[1[0000−0002−9794−749X]], and †Tristan Glatard[2[0000−0003−2620−5883]]

[1] McConnell Brain Imaging Centre, The Neuro, Faculty of Medicine and Health Sciences, McGill University, Montreal, Quebec, Canada
[2] Krembil Centre for Neuroinformatics, Centre for Addiction and Mental Health, Toronto, Ontario, Canada
*Corresponding Author: niusha.mirhakimi@mail.mcgill.ca
†Co–Lead Senior Authors

Abstract. While linear registration is a critical step in MRI preprocessing pipelines, its numerical uncertainty is understudied. Using Monte-Carlo Arithmetic (MCA) simulations, we assessed the most commonly used linear registration tools within major software packages—SPM, FSL, and ANTs—across multiple image similarity measures, two brain templates, and both healthy control (HC, n=50) and Parkinson's Disease (PD, n=50) cohorts. Our findings highlight the influence of linear registration tools and similarity measures on numerical stability. Among the evaluated tools and with default similarity measures, SPM exhibited the highest stability. FSL and ANTs showed greater and similar ranges of variability, with ANTs demonstrating particular sensitivity to numerical perturbations that occasionally led to registration failure. Furthermore, no significant differences were observed between healthy and PD cohorts, suggesting that numerical stability analyses obtained with healthy subjects may generalise to clinical populations. Finally, we also demonstrated how numerical uncertainty measures may support automated quality control (QC) of linear registration results. Overall, our experimental results characterize the numerical stability of linear registration experimentally and can serve as a basis for future uncertainty analyses.

Keywords: Linear Registration · MCA · Numerical Uncertainty.

1 Introduction

Neuroimaging preprocessing steps, including linear and non-linear registration, and segmentation, are sensitive to subtle numerical perturbations in data, pipeline configuration, or hardware environment [16, 20, 26]. In some cases, these instabilities propagate to downstream analyses, such as parcellation-based connectivity mapping [16], impacting derived findings.

Linear registration is a critical step in the vast majority of neuroimaging preprocessing pipelines, commonly formulated as an optimization problem that aims to align a subject's brain image to a common template. Small numerical errors introduced during optimization may steer the solution towards local minima or prevent convergence. Therefore, understanding how numerical errors impact linear registration is crucial. Previous studies [24, 26] demonstrated that MCA can effectively simulate the effects of software updates and hardware-induced variability, including for FSL's FLIRT registration tool.

In this paper, we studied the numerical stability of widely-used linear registration tools. We also investigated the feasibility of using MCA-derived measures for automated QC. Our findings provide guidance for tool selection and support the development of more reproducible preprocessing pipelines.

2 Material and Methods

2.1 Monte-Carlo Arithmetic

MCA is a commonly-used technique to investigate numerical instability in real-world software code bases [19]. It uses randomness to simulate the effect of finite precision in floating-point (FP) operations to mimic rounding errors and catastrophic cancellations [19]. In this study, we used the random rounding perturbation mode, which injects controlled amounts of noise into the output of FP functions using the following perturbation:

$$\text{random_rounding}(x \circ y) = \text{round}(\text{inexact}(x \circ y)) \tag{1}$$

where x and y are FP numbers that represent the function's inputs, $\circ$ is an arithmetic operation, and *inexact* is a random perturbation at a given virtual precision:

$$\text{inexact}(z) = z + 2^{e_z - t}\epsilon, \tag{2}$$

where z is the original FP value, e_z is the exponent of z's FP representation, t is the virtual precision, and ϵ is a random variable uniformly distributed in (-0.5, 0.5).

The Verificarlo [7] and Verrou [9] frameworks implement random rounding to assess numerical stability. Verificarlo is an MCA tool built on the LLVM compiler infrastructure, supporting a wide range of languages such as C, C++, and Fortran. Verrou, in contrast, uses dynamic binary instrumentation via Valgrind, allowing perturbation of FP operations at runtime. Due to its lower computational overhead, we used Verificarlo as the primary framework and employed Verrou to validate a key result. Specifically, we used *fuzzy-libm* [24], a lightweight Verificarlo backend that perturbs only the outputs of standard mathematical functions from the `libm` library (e.g., `exp`, `sin`, `log`). Although *fuzzy-libm* is suitable for assessing the stability of programs that rely heavily on `libm`, perturbing all FP operations (e.g. `+`, `-`, $\times$, $\div$) during execution provides a more general solution.

2.2 Numerical Uncertainty Metric

The Framewise Displacement (FD) metric was introduced as a single measure to characterize head movement through a subject's time series [21, 22]. FD summarizes motion parameters into a displacement measured at a distance of 50 mm from the origin, approximating the mean radius of an adult brain:

$$FD_i = \|t_i\| + 50.\left(\frac{\pi}{180}\right)\|r_i\|, \tag{3}$$

where FD_i represents the framewise displacement for a transformation labeled i, t_i is the translation vector in mm, r_i is the rotation vector expressed in Euler angles (degrees), and $\|.\|$ is the Euclidean norm. We assume that rotation and translation parameters correlate with shear and scaling parameters, which enables us to summarise 12-parameter affine registration with the FD measure.

To study how the FD varies under numerical perturbation for a given subject, we used the standard deviation (SD) of FD across the MCA runs as a measure of numerical uncertainty. Since the distributions of SD values were not normally distributed (see p-values from the Shapiro–Wilk test in Table **??**), we used non-parametric hypothesis testing to compare across tools and similarity measures.

2.3 Similarity Measures and Optimization Methods

Three linear registration tools were evaluated in this study. FMRIB's Linear Image Registration Tool (FLIRT), part of the FMRIB Software Library (FSL) [25], antsRegistrationSyN.sh script in Advanced Normalization Tools (ANTs) [4], and Statistical Parametric Mapping's (SPM) spm_affreg function. Both FSL and ANTs use multiresolution optimization strategies: registration begins with a coarse alignment at 8 mm resolution and is progressively refined through stages at 4 mm, 2 mm, and finally 1 mm [14, 3]. FSL uses the correlation ratio (CR) as its default similarity measure, whereas ANTs uses mutual information (MI). SPM adopts a different optimization approach, using a Bayesian framework to estimate the affine transformation by iteratively incorporating prior knowledge and minimizing alignment errors [2, 1]. SPM employs the Sum of Squared Differences (SSD) as its default similarity measure. All tools used trilinear interpolation by default, and default settings were applied during this study unless otherwise noted.

2.4 Dataset and Preprocessing

Fifty subjects with PD (age: 61.63 ± 7.29 years; 22 female; UPDRS3_OFF: 21.55 ± 11.42) and fifty HC subjects (age: 62.17 ± 10.48 years; 24 female; UPDRS3_OFF: 1.19 ± 2.21) were randomly selected from the baseline session of the Parkinson's Progression Markers Initiative (PPMI) dataset [18], an ongoing, multicenter observational study to identify PD biomarkers. Hoehn and Yahr (HY) scores were available for 49 of the 50 PD subjects: 25 were classified as

stage 1, 23 as stage 2, and 1 as stage 3. In the HC group, HY scores were available for 48 subjects: 47 were scored as stage 0, and 1 as stage 1.

T1-weighted images for both cohorts were processed using FSL's RobustFOV to remove the neck and BET to strip the skull. The resulting brain-extracted images were then used throughout the study as inputs to different linear registration pipelines. Each image was registered twice to a template for a given pipeline: first, using the standard (unperturbed) registration tools (referred to as the "IEEE registration"), and second, ten times using the MCA-perturbed registration tools (referred to as the "MCA registrations"). Manual quality control (QC) was performed on the registered images to ensure proper alignment: key brain structures, including the ventricles, corpus callosum, cerebellum, basal ganglia, brainstem, and Sylvian fissure, were assessed for alignment, along with the overall edge alignment of the brains with the template in sagittal, coronal, and axial planes. All registered images, either IEEE or MCA, were visually inspected and labeled as "failed" or "passed." However, a subject's overall QC classification was determined separately for each registration pipeline, based solely on the quality of their corresponding IEEE registration. This approach enables exploring the potential of using MCA-derived features for automatic QC, specifically to distinguish between subjects who passed and failed QC based on the IEEE standard for each template and tool combination.

2.5 Experimental Setup

In this study, the symmetric (Sym) and asymmetric (Asym) versions of the MNI152NLin2009c template, each with a resolution of 1 mm were exclusively used. The templates are accessible through the TemplateFlow website at templateflow.org. The Dockerized formats of SPM12, ANTs v2.5.0, and FSL v6.0.4 were utilized throughout the study. All registration pipelines were executed using standardized Boutiques descriptors and invocations to ensure reproducibility of tool configuration and parameterization [11]. Docker recipes for both unperturbed and perturbed software versions, as well as Boutique descriptors, are available in this GitHub repository: github.com/mirhnius/mca_linear_registration.

3 Results

3.1 Numerical Uncertainty Can Reach Levels Comparable to the Template's Resolution or Motion Artifacts During Scanning

An examination of the median SD of FD across passed QC subjects for the three registration tools indicates generally low values: SPM (1.7×10^{-9}mm), FSL (0.05 mm), and ANTs (0.07 mm). However, in ANTs and FSL, some subjects exhibited SD values greater than 0.2 mm—with a substantial portion of these observed for the MI and NMI similarity measures in FSL—and several cases exceeded 1 mm, surpassing the resolution of the registration template. In neuroimaging studies, mean displacements exceeding 0.2 mm within a subject are typically considered substantial motion artifacts [12], as they represent a significant portion of

the voxel dimension. These findings reveal that numerical instability in linear registration can, in some instances, introduce spatial variability comparable in magnitude to subject motion during a recording session. Table **??** presents the mean SD per pipeline, highlighting that failed QC subjects consistently exhibit higher average variability in all pipelines.

3.2 SPM Is the Most Stable Tool, FSL and ANTs Are Comparable

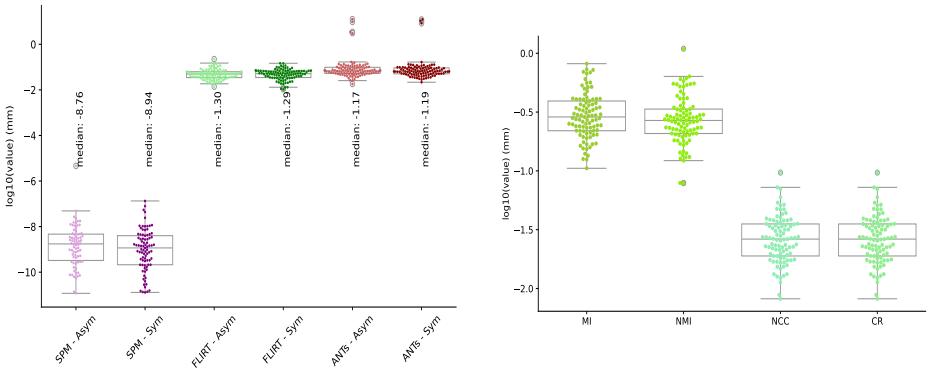
We computed the SD of FD in MCA runs per subject for each registration tool with its default settings. Comparing SD distributions revealed notable differences in numerical uncertainty associated with each tool, confirmed by Friedman tests ($p \approx 3 \times 10^{-27}$ for registering to the asymmetric template and $p \approx 3 \times 10^{-29}$ for registering to the symmetric template). While FSL and ANTs exhibit comparable median variability, SPM demonstrates significantly greater numerical stability. A closer examination of Figure 1(a) illustrates that despite the similar range of uncertainty for ANTs and FSL, ANTs produces outliers—subjects with high variability across MCA runs. We performed a detailed visual QC on all images registered using instrumented tools and discovered a unique sensitivity in ANTs: 4 subjects who previously passed QC under unperturbed conditions failed under MCA-perturbed conditions. These failures directly contribute to the observed outliers.

3.3 Similarity Measure Significantly Affects Numerical Stability

To isolate the impact of similarity measure selection, we conducted further experiments using FLIRT. Subjects were registered to the asymmetric template using various similarity measures: Sum of Squared Differences (SSD), Normalized Cross Correlation (NCC), Correlation Ratio (CR), Mutual Information (MI), and Normalized Mutual Information (NMI)—as described in previous works [23, 8, 13]. SSD results were excluded since more than half of the unperturbed registrations failed. These failures were characterized by implausible transformation matrices, such as a large scaling factor, that still produced low loss values. This is a known limitation of multi-resolution optimization schemes [15]. Analyzing the SD of FD distributions across MCA runs revealed that the choice of similarity measure significantly affects numerical stability (Friedman test: $p \approx 8 \times 10^{-54}$). As shown in Figure 1(b), NCC and CR yielded more stable registrations than MI and NMI among subjects who passed QC. However, these results may not generalize to other registration tools, as discussed in the Supplementary **??**.

3.4 Numerical Uncertainty Metrics Hold Promise for Automated QC

We investigated the potential of numerical uncertainty measures for automated QC in preprocessing pipelines. A consistent pattern emerges in the standard deviation of framewise displacement across pipelines (Figure 2), with failed cases

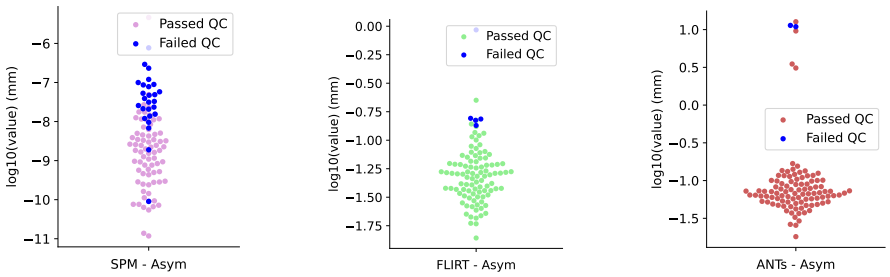


(a) Comparison across linear registration tools (SPM, FSL, and ANTs) and templates (symmetric and asymmetric), using their default similarity measures

(b) Comparison of linear registration across similarity measures using FSL registered to the asymmetric template.

Fig. 1. Standard deviation of framewise displacement across MCA runs for each passed QC subject and registration pipeline.

exhibiting greater variability than those that passed QC. The distinction was most pronounced in SPM, where the number of failed and passed subjects allowed for observation of two separate distributions. Given the class imbalance, we employed one-class classification models trained on passed QC subjects and treated failed cases as anomalies. Since all test cases were known failures, we report recall as the evaluation metric available in Table 1. While this setup is not intended as a definitive classification approach, it serves as a proof of concept, indicating that these measures reflect meaningful aspects of registration quality and may serve as promising features in future automated QC systems.



(a) (b) (c)

Fig. 2. Comparison of the standard deviation of framewise displacement between passed and failed QC subjects. (a) SPM, (b) FLIRT, and (c) ANTs.

Table 1. Evaluation of novelty detection methods for identifying failed subjects. Recall scores of three novelty detection methods in distinguishing passed vs. failed QC subjects based on the SD of FD. No threshold optimization or training was applied. Each linear registration pipeline processed the same 100 images; the number of passed cases is $100 - N$, where N is the number of QC failures.

Template	Software	Similarity Measure	N	95th Quantile	KDE (5%)	1-Class SVM
Asym	FSL	NCC	6	0.83	0.83	1.0
Asym	FSL	NMI	5	0.8	1.0	0.4
Asym	FSL	MI	7	1.0	1.0	1.0
Asym	FSL	CR	5	1.0	1.0	0.2
Asym	ANTs	MI	2	1.0	1.0	1.0
Asym	SPM	SSD	26	0.73	0.73	1.0
Sym	FSL	CR	6	0.5	0.5	0.5
Sym	ANTs	MI	2	1.0	1.0	1.0
Sym	SPM	SSD	15	0.6	0.6	1.0

3.5 Cohort and Template Choices Show No Statistically Significant Impact on Numerical Stability

We investigated whether cohort (PD vs. HC) or template choice (asymmetric vs. symmetric) significantly affects the numerical stability of linear registration, as measured by the SD of FD across MCA runs. For cohort effects, Mann–Whitney U tests revealed no statistically significant differences in numerical uncertainty between PD and HC subjects across all registration pipelines (Table **??**). Regarding template choice, Wilcoxon signed-rank tests comparing registrations to symmetric and asymmetric templates showed no significant differences for FSL ($p \approx 0.479$) and ANTs ($p \approx 0.329$). While SPM showed a marginally significant difference ($p \approx 0.0194$), this was likely driven by a single outlier, with minimal visual difference observed in the overall distributions. These findings, together, suggest that neither cohort differences nor template choice substantially influences numerical stability in the evaluated registration tools.

4 Discussion and Conclusion

This work investigates the numerical uncertainty of linear registration. While focusing on a single dataset cannot fully disentangle the effects of software choice, similarity measures, demographics, and template selection, it nonetheless serves as an initial case study, highlighting the importance of investigating numerical uncertainty as a previously overlooked source of variability. Ensuring numerical stability is essential for building robust pipelines, which in turn is a prerequisite for developing reliable imaging biomarkers. This study sets the stage for more comprehensive evaluations and future efforts to develop more reliable neuroimaging workflows.

Numerical variability in linear registration can be significant, in some cases comparable to template resolution and head movement in the scanner, suggesting it may lead to substantial discrepancies in preprocessing and downstream analyses for some subjects.

The numerical stability of linear registration is strongly software-dependent. Among the tools evaluated, SPM exhibited markedly greater stability, while FSL and ANTs showed heightened sensitivity to numerical perturbations. We considered two potential explanations for this difference: either SPM was not properly instrumented with Verificarlo, or its underlying optimization strategy contributes to its stability. To rule out instrumentation issues, we conducted supplementary validation experiments (Supplementary **??**). We confirmed that Octave was instrumented correctly, and additionally, we instrumented SPM with the Verrou framework. The comparable variability patterns observed across both tools suggest that the instrumentation was effective, supporting the hypothesis that SPM's increased numerical stability stems from its fundamentally different optimization approach.

Multi-resolution approaches are predicated on the assumption that the minima identified at lower resolutions are sufficiently close to the global minima at higher resolutions, which is necessary for convergence [13]. However, there is no guarantee that this assumption holds, as local minima can shift across different resolutions. Additionally, the processes of subsampling and interpolation introduce noise, in which multi-resolution approaches not only fail to simplify the optimization landscape but also introduce new local optima, adding complexity to the problem [15]. Multi-resolution methods involve complex subsampling and interpolation steps that can amplify small perturbations, potentially leading to suboptimal solutions.

SPM employs a Bayesian optimization framework that incorporates prior knowledge about variability in head shape and size, using a Maximum A Posteriori approach to enhance robustness and convergence speed [2]. This method is particularly advantageous when dealing with low-quality data, as it reframes the optimization objective to not only maximize image similarity but also penalize deviations from expected parameter values based on prior distributions. This strong regularization may explain SPM's observed numerical stability and its resilience to small perturbations.

The choice of similarity measure influences numerical stability, and this effect varies across software. FSL was selected to evaluate similarity measures, as it supports both SSD (used by SPM) and MI (used by ANTs). ANTs does not support CR, and SPM's spm_affreg hardcodes SSD. A comparison of MI in FSL and ANTs (Supplementary **??**) showed that stability depends on both the cost function and the tool. This suggests that each tool should be evaluated with multiple similarity measures, as optimization strategies may interact with cost functions. Further investigation is needed to understand ANTs' specific sensitivity and assess the linear registration's configurations in a broader sense.

Given that linear registration tends to struggle in the presence of brain atrophy [6], that patient data often suffers from motion-related artifacts[10], and

that poor image quality can amplify numerical instability during processing [5, 24], the PD cohort was expected to exhibit reduced numerical stability. However, no significant differences were observed between the PD and control groups, suggesting that numerical uncertainty findings from healthy populations may be generalizable to pathological cohorts. Nevertheless, we hypothesize that even small perturbations introduced during registration may propagate through the full preprocessing pipeline, potentially influencing downstream metrics such as cortical thickness and altering observed effect sizes between groups.

This study demonstrated that numerical variability measures hold promise for integration into automated QC algorithms within preprocessing pipelines. The runtime overhead introduced by MCA is reported in Table **??**. MCA overhead is feasible for linear registration, but in more complex pipelines, accumulated numerical errors may hinder convergence and substantially increase runtime. We used variability in FD as a proxy for numerical uncertainty and showed that subjects who failed QC exhibited higher variability. As a future direction, an alternative metric could be developed based on the Anatomical Fiducials Registration Error, a method introduced by [17], which uses 32 anatomical fiducial points identified on brain scans to assess registration accuracy. Studying the variability of these fiducial points across perturbed runs may offer a more localized and anatomically meaningful estimation of numerical uncertainty.

Acknowledgments. This work was supported by the Michael J. Fox Foundation (LivingPark) and FRSQ – Réseaux de recherche thématiques (35450).

Disclosure of Interests. The authors have no competing interests to declare.

References

1. Ashburner, J., Barnes, G., Chen, C.C., Daunizeau, J., Flandin, G., Friston, K., Kiebel, S., Kilner, J., Litvak, V., Moran, R., et al.: Spm12 manual. Wellcome Trust Centre for Neuroimaging, London, UK **2464**(4), 53 (2014)
2. Ashburner, J., Neelin, P., Collins, D., Evans, A., Friston, K.: Incorporating prior knowledge into image registration. Neuroimage **6**(4), 344–352 (1997)
3. Avants, B.B., Tustison, N.J., Song, G., Cook, P.A., Klein, A., Gee, J.C.: A reproducible evaluation of ants similarity metric performance in brain image registration. Neuroimage **54**(3), 2033–2044 (2011)
4. Avants, B.B., Tustison, N.J., Stauffer, M., Song, G., Wu, B., Gee, J.C.: The insight toolkit image registration framework. Frontiers in neuroinformatics **8**, 44 (2014)
5. Chatelain, Y., Sao Young, N.Y., Kiar, G., Glatard, T.: Pytracer: Automatically profiling numerical instabilities in python. IEEE Transactions on Computers **72**(6), 1792–1803 (2022)
6. Dadar, M., Fonov, V.S., Collins, D.L., Initiative, A.D.N., et al.: A comparison of publicly available linear mri stereotaxic registration techniques. Neuroimage **174**, 191–200 (2018)
7. Denis, C., Castro, P.D.O., Petit, E.: Verificarlo: Checking floating point accuracy through monte carlo arithmetic. arXiv preprint arXiv:1509.01347 (2015)

8. Deserno, T.M.: Fundamentals of biomedical image processing. In: Biomedical image processing, pp. 1–51. Springer (2010)
9. Févotte, F., Lathuilière, B.: Verrou: Assessing floating-point accuracy without recompiling (2016), https://hal.science/hal-01383417
10. Gilmore, A.D., Buser, N.J., Hanson, J.L.: Variations in structural mri quality significantly impact commonly used measures of brain anatomy. Brain informatics **8**, 1–15 (2021)
11. Glatard, T., Kiar, G., Aumentado-Armstrong, T., Beck, N., Bellec, P., Bernard, R., Bonnet, A., Camarasu-Pop, S., Cervenansky, F., Das, S., et al.: Boutiques: a flexible framework for automated application integration in computing platforms. arXiv preprint arXiv:1711.09713 (2017)
12. Gu, S., Satterthwaite, T.D., Medaglia, J.D., Yang, M., Gur, R.E., Gur, R.C., Bassett, D.S.: Emergence of system roles in normative neurodevelopment. Proceedings of the National Academy of Sciences **112**(44), 13681–13686 (2015)
13. Hill, D.L., Batchelor, P.G., Holden, M., Hawkes, D.J.: Medical image registration. Physics in medicine & biology **46**(3), R1 (2001)
14. Jenkinson, M., Bannister, P., Brady, M., Smith, S.: Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage **17**(2), 825–841 (2002)
15. Jenkinson, M., Smith, S.: A global optimisation method for robust affine registration of brain images. Medical image analysis **5**(2), 143–156 (2001)
16. Kiar, G., de Oliveira Castro, P., Rioux, P., Petit, E., Brown, S.T., Evans, A.C., Glatard, T.: Comparing perturbation models for evaluating stability of neuroimaging pipelines. The International Journal of High Performance Computing Applications **34**(5), 491–501 (2020)
17. Lau, J.C., Parrent, A.G., Demarco, J., Gupta, G., Kai, J., Stanley, O.W., Kuehn, T., Park, P.J., Ferko, K., Khan, A.R., et al.: A framework for evaluating correspondence between brain images using anatomical fiducials. Human Brain Mapping **40**(14), 4163–4179 (2019)
18. Marek, K., Jennings, D., Lasch, S., Siderowf, A., Tanner, C., Simuni, T., Coffey, C., Kieburtz, K., Flagg, E., Chowdhury, S., et al.: The parkinson progression marker initiative (ppmi). Progress in neurobiology **95**(4), 629–635 (2011)
19. Parker, D.S.: Monte Carlo arithmetic: exploiting randomness in floating-point arithmetic. Citeseer (1997)
20. Pepe, I.G., Sivakolunthu, V., Park, H.L., Chatelain, Y., Glatard, T.: Numerical uncertainty of convolutional neural networks inference for structural brain mri analysis. In: International Workshop on Uncertainty for Safe Utilization of Machine Learning in Medical Imaging. pp. 64–73. Springer (2023)
21. Power, J.D., Barnes, K.A., Snyder, A.Z., Schlaggar, B.L., Petersen, S.E.: Spurious but systematic correlations in functional connectivity mri networks arise from subject motion. Neuroimage **59**(3), 2142–2154 (2012)
22. Power, J.D., Mitra, A., Laumann, T.O., Snyder, A.Z., Schlaggar, B.L., Petersen, S.E.: Methods to detect, characterize, and remove motion artifact in resting state fmri. Neuroimage **84**, 320–341 (2014)
23. Roche, A., Malandain, G., Pennec, X., Ayache, N.: The correlation ratio as a new similarity measure for multimodal image registration. In: Medical Image Computing and Computer-Assisted Intervention—MICCAI'98: First International Conference Cambridge, MA, USA, October 11–13, 1998 Proceedings 1. pp. 1115–1124. Springer (1998)

24. Salari, A., Chatelain, Y., Kiar, G., Glatard, T.: Accurate simulation of operating system updates in neuroimaging using monte-carlo arithmetic. In: Uncertainty for Safe Utilization of Machine Learning in Medical Imaging, and Perinatal Imaging, Placental and Preterm Image Analysis: 3rd International Workshop, UNSURE 2021, and 6th International Workshop, PIPPI 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, October 1, 2021, Proceedings 3. pp. 14–23. Springer (2021)
25. Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobnjak, I., Flitney, D.E., et al.: Advances in functional and structural mr image analysis and implementation as fsl. Neuroimage **23**, S208–S219 (2004)
26. Vila, G., Medernach, E., Gonzalez Pepe, I., Bonnet, A., Chatelain, Y., Sdika, M., Glatard, T., Camarasu Pop, S.: The impact of hardware variability on applications packaged with docker and guix: A case study in neuroimaging. In: Proceedings of the 2nd ACM Conference on Reproducibility and Replicability. pp. 75–84 (2024)