Extending provenance information in CBRAIN to address reproducibility issues across computing platforms

Tristan Glatard[1,2], Lindsay B. Lewis[1], Rafael Ferreira da Silva[3], Marc-Etienne Rousseau[1], Claude Lepage[1], Pierre Rioux[1], Najmeh Mahani[1], Ewa Deelman[3], and Alan C. Evans[1]

[1] McConnell Brain Imaging Centre, Montreal Neurological Institute, McGill University, Canada.
[2] University of Lyon, CNRS, INSERM, CREATIS, Villeurbanne, France.
[3] University of Southern California, Information Sciences Institute, Marina Del Rey, CA, USA.

Context: Neuroimaging tools are prone to reproducibility issues across computing platforms due to the propagation of numerical errors along pipelines (Gronenschild, et al., 2012). When different computing systems are used in the same study, these issues may alter the results and even generate false positives. We are designing a system to identify and mitigate reproducibility issues in experiments executed on distributed computing platforms. This system will extend the provenance information available in the CBRAIN web platform (Sherif, et al., 2014) with system-level monitoring information captured by the Kickstart tool (Deelman, et al., 2006). Method: We processed the 150-subject ICBM dataset (Mazziotta et al., 2001) with 3 pipelines: (i) brain tissue segmentation using FSL FAST (Zhang, et al., 2001) (ii) subcortical structure segmentation with FSL FIRST (Patenaude, et al., 2011) (iii) cortical thickness estimation with Freesurfer (Fischl and Dale, 2000). We used FSL 5.0 (build 506) to compare results obtained on two clusters running Linux CentOS 5 and Fedora 20 respectively. We used Freesurfer 5.3.0 and compared the results obtained with CentOS 4 and CentOS 6 x86_64 builds, executed on the Linux Fedora 20 cluster. Results: Brain tissue segmentations computed in FSL on CentOS5 vs. Fedora 20 have a Dice coefficient higher than 0.999 for grey matter, white matter, and CSF. Numerical differences result in discrete noise-like segmentation errors mostly located at the tissue interfaces (see Figure 1). Using ltrace (http://ltrace.org), we identified that these differences are due to different implementations of the exponential function (expf) between CentOS 5 (glibc 2.5) and Fedora 20 (glibc 2.18). Subcortical structure segmentations computed on CentOS5 vs. Fedora 20 have a Dice coefficient ranging from 0.59 to 1 (see Figure 2). Cortical thickness difference maps thresholded with random field theory (RFT) show significant differences between CentOS4 and CentOS6 Freesurfer builds for $p<0.05$ and $p<0.01$ (see Figure 3). Discussion: Different computing platforms may produce substantially different results in neuroimaging pipelines. Therefore it is legitimate to avoid using multiple computing platforms in a study. However, this drastically reduces the amount of available computing resources, which slows down experiments. Our provenance-based system will help identify the maximal set of resources that can be used in a study without altering its results.

Figure 1: Sum of binarized differences between brain tissue segmentations of the 150 ICBM subjects with FSL FAST on Linux CentOS 5 vs. Linux Fedora 20. From top to bottom and left to right: z=33,53,73,93,113.

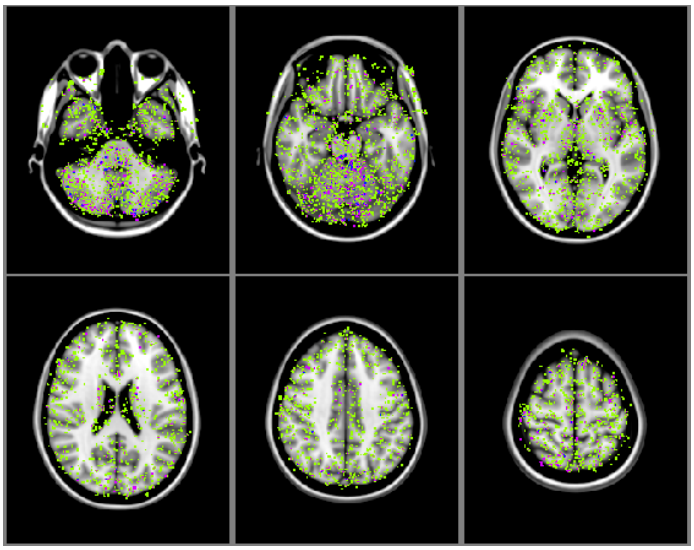
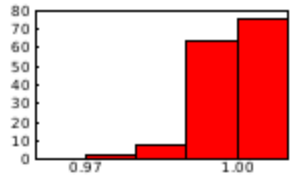
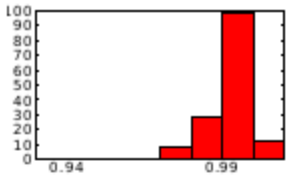


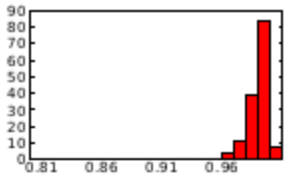
Figure 2: Histograms of DICE coefficients between segmentations obtained on CentOS5 vs. Fedora 20 with FSL FIRST. μ: mean; σ: standard deviation.



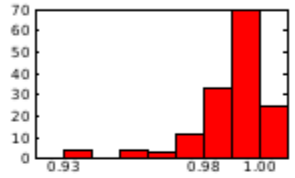
(a) Left-Thalamus-Proper μ=0.99, σ=0.01



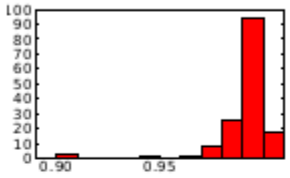
(b) Left-Caudate μ=0.99, σ=0.01



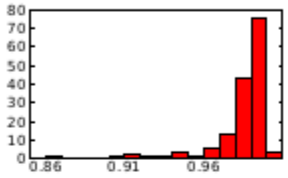
(c) Left-Putamen μ=0.98, σ=0.02



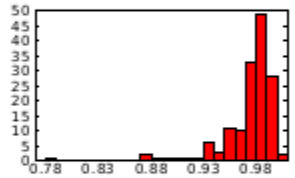
(d) Left-Pallidum μ=0.98, σ=0.01



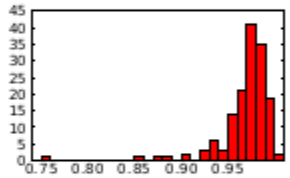
(e) Brain-Stem / 4th Ventricle μ=0.99, σ=0.01



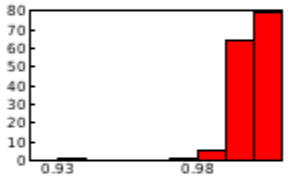
(f) Left-Hippocampus μ=0.98, σ=0.02



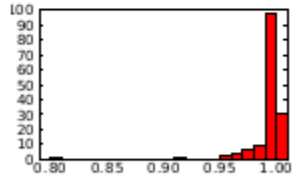
(g) Left-Amygdala μ=0.97, σ=0.03



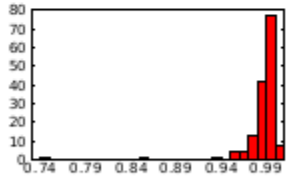
(h) Left-Accumbens-area μ=0.96, σ=0.02



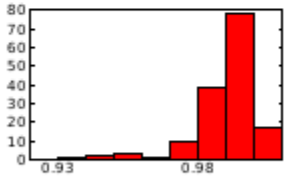
(i) Right-Thalamus-Proper μ=0.99, σ=0.01



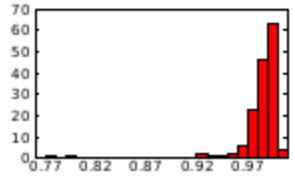
(j) Right-Caudate μ=0.99, σ=0.02



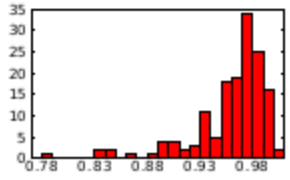
(k) Right-Putamen μ=0.98, σ=0.03



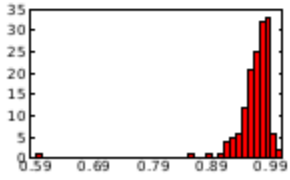
(l) Right-Pallidum μ=0.99, σ=0.01



(m) Right-Hippocampus μ=0.98, σ=0.03



(n) Right-Amygdala μ=0.95, σ=0.04



(o) Right-Accumbens-area μ=0.96, σ=0.04

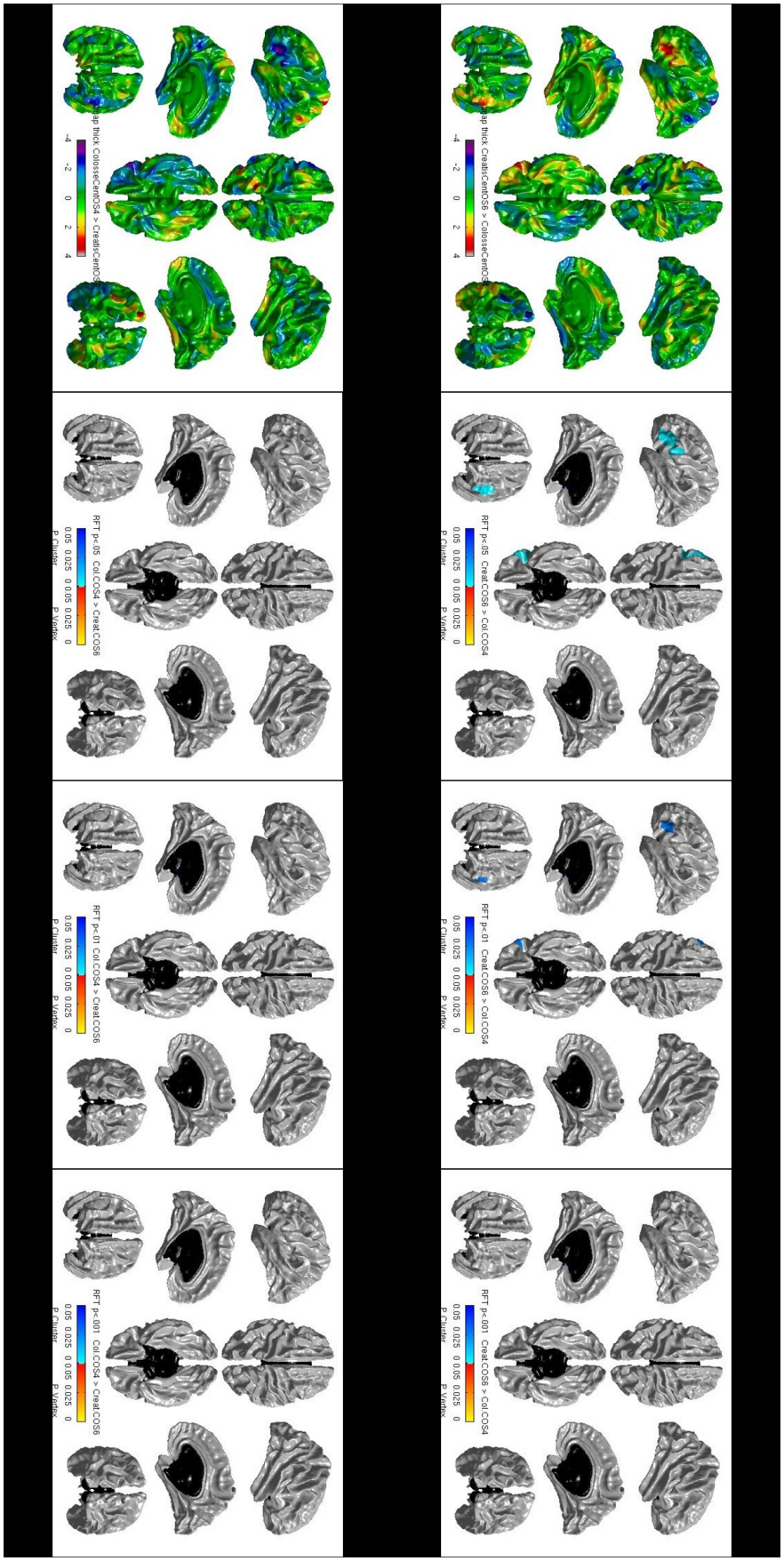



Figure 3: Comparison of cortical thickness maps between CentOS4 and CentOS6 Freesurfer builds. Top row: CentOS6 vs CentOS4; bottom row: CentOS4 vs. CentOS6. From left to right, column (1): t statistics; columns (2)-(4): random field theory (RFT) maps thresholded at $p<0.05$, $p<0.01$ and $p<0.001$, respectively.