# Impact of bias-correction and skull-stripping pipelines on spatial normalisation using SPM5 in a phantom model

**J. Acosta-Cabronero[1], G. B. Williams[1], and P. J. Nestor[1]**
[1]Department of Clinical Neurosciences, University of Cambridge, Cambridge, United Kingdom

**Introduction**: In recent years, statistical parametric mapping (SPM) *[1]* has been widely used for the technique of optimised voxel-based morphometry (VBM). The VBM technique localises differences in structural magnetic resonance (MR) images between patient populations on a voxel-by-voxel basis *[2]*. Several studies have shown that VBM can be improved by skull-stripping *[3-9]* and correcting for intensity non-uniformity (radio frequency bias) *[10, 11]* in images prior to analysis. The latest SPM release, SPM5, enables spatial normalisation to standard space, tissue segmentation and bias correction to be combined within the same model *[12]*. The purpose of the present study was to investigate the impact that the pre-processing methods enumerated below have on normalised gray-matter (GM) segments in comparison to each other and to that derived from default SPM5 settings:
*Skull-stripping*. (i) Fully-automated hybrid watershed algorithm (HWA) using atlas information *[6]*; (ii) manually optimised brain surface extractor (BSE) *[3]* and (iii) fully-automated brain extraction tool v.2 (BET2) *[4]* with fractional intensity threshold, f, set to 0.4 and 0.5 (vertical gradient, g, set to 0).
*Bias-correction*. (i) Non-parametric non-uniform intensity normalisation (N3) *[10]* and (ii) bias field corrector (BFC) *[3]*, both fully-automated.

**Methods and Results**: The pre-processing pipelines were evaluated on a $T_1$-weighted MRI BrainWeb phantom (3% noise, 40% bias) *[13]* as shown in Fig. 1. A gold-standard dataset was obtained by manually delineating the cortical surface on the ground-truth image (0% bias). All volumes were skull-stripped and then intensity-corrected prior to being processed by SPM5. The full volume was also segmented using SPM5 in order to assess the need for pre-processing steps.

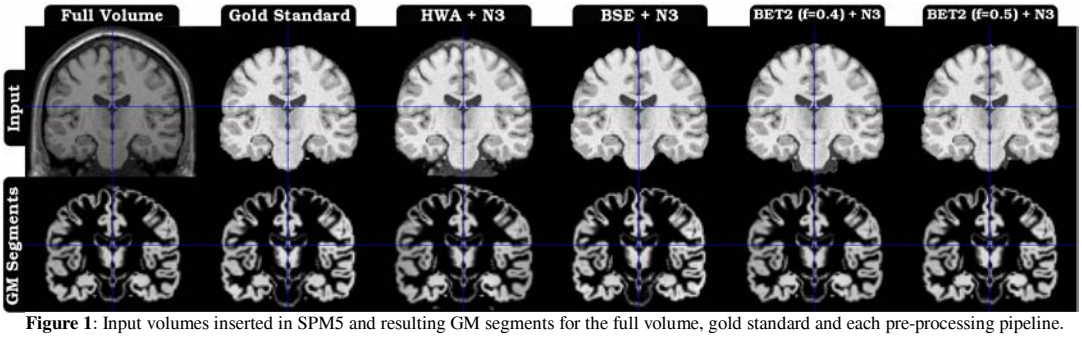


**Figure 1**: Input volumes inserted in SPM5 and resulting GM segments for the full volume, gold standard and each pre-processing pipeline.

*Bias-correction.* Ashburner et.al. *[12]* suggested that SPM5 is more accurate if it does not attempt to estimate bias fields when non-uniformities are not present. Hence, the performance of the bias-correction algorithms was evaluated by computing the root-mean-squared (RMS) difference between the intensity-uncorrected (BC off – bias regularisation: 10; bias FWHM: 150-mm cut-off) gold standard and the non-uniformity corrected (BC on - defaults) and uncorrected pre-processed (or full-volume) normalised outputs as a percentage of the maximum white matter (WM) signal intensity from the ground-truth image. N3 outperformed BFC and SPM5 for all skull-stripping methods as shown in Table 1. N3 performed consistently well for all brain-extraction methods, but the RMS error was lower when the bias correction within SPM5 was disabled (BC off). Table 1 also shows that any skull-stripping method and N3 correction used as pre-processing pipelines outperformed the bias correction implemented in SPM5 for the full volume.

| *Method* | *BC off* | | *BC on* | |
|---|---|---|---|---|
| | *N3* | *BFC* | *N3* | |
| Full Volume | — | — | — | 4.37 |
| HWA | 4.26 | 7.34 | 4.43 | 4.40 |
| BSE | 4.22 | 4.60 | 4.49 | 4.44 |
| BET2, f=0.4 | 4.23 | 5.57 | 4.49 | 4.41 |
| BET2, f=0.5 | 4.22 | 5.51 | 4.46 | 4.42 |

**Table 1**: RMS error of the bias-corrected images for the full-volume and pre-processing methods.

*Warping to template*. The impact on spatial normalisation was evaluated by computing the Jaccard similarity (J) as the ratio of the size (*i.e.* number of voxels, N) of the intersection between the unmodulated normalised GM segment (thresholded to 0.5) for each pipeline and for the gold standard, divided by the size of their union. False negative (FN) and false positive rates (FP) as a percentage of the size of the thresholded gold-standard GM segment were also calculated. In addition, the difference between the size (ΔN) of the pre-processed and gold-standard final outcomes as a percentage of the latter was also computed. Table 2 shows that optimised BSE and BET2 (f=0.5) had the best similarity with regard to the gold-standard boundary. Note that a more conservative setting of BET2 (f=0.4) produced a higher similarity coefficient than the full-volume segment. HWA performed poorly as this method aims to conservatively bound the pial surface and the *unified segmentation* process in SPM5 classifies the sinuses as GM tissue. FN, FP and ΔN results clearly illustrate the general behaviour of each technique: (i) As expected, HWA produced the highest difference in number of voxels and FP rate. However, although the HWA segment had a very low FN rate in native space, the redundant GM tissue induced substantial mis-registration in standard space and therefore, high FN rate. (ii) BSE is the most specific of all methods, hence the lowest FP rate. However, ΔN was negative, indicating that some GM voxels were removed. This is the opposite effect to that for HWA, but has a similar impact on the FN rate after warping to the template. (iii) BET2 was the most consistent method; it kept the lowest FN rates, and FN, FP and ΔN could be lowered by increasing f. Table 2 also shows that the full-volume method performed better than HWA, but ostensibly worse than the BSE, BET2 + N3 pipelines.

| *Method* | *J* | *FN (%)* | *FP (%)* | *ΔN (%)* |
|---|---|---|---|---|
| Full Volume | 0.84 | 9.0 | 8.3 | -0.7 |
| HWA + N3 | 0.80 | 8.3 | 15.2 | 6.9 |
| BSE + N3 | 0.90 | 6.4 | 4.4 | -2.0 |
| BET2, f=0.4 + N3 | 0.86 | 5.5 | 9.5 | 3.9 |
| BET2, f=0.5 + N3 | 0.89 | 4.8 | 6.9 | 2.1 |

**Table 2**: Jaccard similarity coeffienct, false negative rates, false positive rates and differences in number of voxels for the normalised GM segments obtained from the full-volume and pre-processing methods.

**Discussion and Conclusions**: This study evaluated the performance of SPM5 using different methods to pre-process structural MRI data. Running BET2 to skull-strip and N3 to bias-correct the BrainWeb phantom performed especially well. BSE + N3 also performed better than the full-volume and HWA + N3 method, but its high specificity came at the price of GM tissue removal that, in turn, could adversely impact on statistical analyses. It was also demonstrated that removing brain tissue or including non-brain voxels have a negative effect on the warping process. In summary, this phantom study suggests that BET2 + N3 and BSE + N3 pre-processing may offer greater normalisation accuracy compared to SPM5 alone.

**References:** [1] Available from http://www.fil.ion.ucl.ac.uk/spm; [2] Ashburner J. et.al., Neuroimage 8:1105 (1997); [3] Shattuck, D.W. et.al., Neuroimage 13: 856 (2001); [4] Smith S.M. Hum. Brain Mapp. 17: 143 (2002); [5] Boesen K. et.al., Neuroimage 22: 1255 (2004); [6] Ségonne F. et.al., Neuroimage 22: 1060 (2004); [7] Rex D.E. et.al., Neuroimage 23: 625 (2004); [8] Zhuang A.H. et.al., Neuroimage 32: 79 (2006); [9] Fennema-Notestine C. et.al., Hum. Brain Mapp. 27: 99 (2006); [10] Sled J.G. et.al., IEEE Trans. Med. Imag. 17: 87 (1998); [11] Arnold J.B. et.al., Neuroimage 13: 931 (2001); [12] Ashburner J. et.al., Neuroimage 26: 839 (2005); [13] Available from http://www.bic.mni.mcgill.ca/brainweb/selection_normal.html.