Template-free estimation of intracranial volume: A preterm birth animal model study

Juan Eugenio Iglesias[1], Sebastiano Ferraris[1], Marc Modat[1], Willy Gsell[2], Jan Deprest[2,3], Johannes L. van der Merwe[2], and Tom Vercauteren[2,3]

[1] Translational Imaging Group, University College London (UCL), UK
[2] Biomedical Sciences Group, KU Leuven, Belgium
[3] Wellcome / EPSRC Centre for Interventional and Surgical Sciences, UCL, UK

Abstract. Accurate estimation of intracranial volume (ICV) is key in neuro-imaging-based volumetric studies, since estimation errors directly propagate to the ICV-corrected volumes used in subsequent analyses. ICV estimation through registration to a reference atlas has the advantage of not requiring manually delineated data, and can thus be applied to populations for which labeled data might be inexistent or scarce, e.g., preterm born animal models. However, such method is not robust, since the estimation depends on a single registration. Here we present a group-wise, template-free ICV estimation method that overcomes this limitation. The method quickly aligns pairs of images using linear registration at low resolution, and then computes the most likely ICV values using a Bayesian framework. The algorithm is robust against single registration errors, which are corrected by registrations to other subjects. The algorithm was evaluated on a pilot dataset of rabbit brain MRI ($N = 7$), in which the estimated ICV was highly correlated ($\rho = 0.99$) with ground truth values derived from manual delineations. Additional regression and discrimination experiments with human hippocampal volume on a subset of ADNI ($N = 150$) yielded reduced sample sizes and increased classification accuracy, compared with using a reference atlas.

1   Introduction

Background. Intracranial volume (ICV) is a crucial covariate in MRI-based neuroimaging studies. Correcting for ICV, by division or regression [1], is necessary for comparing volume estimates of brain structures from cases with different head sizes. While the automated segmentation of brain structures has received a considerable amount of attention in the literature, ICV estimation is often overlooked, despite the fact that a poor ICV estimate can have a very detrimental impact on the corrected volume of an otherwise very well segmented structure. Compared with skull stripping, ICV estimation needs to account for all the tissue and fluid inside the skull, not only the brain. Otherwise, atrophy or growth of brain structures would be partially explained by changes of the whole brain.

The literature of ICV estimation is dominated by methods designed for adult human brain MRI. Earlier approaches relied on simple thresholding and morphological operations [2]. In the 2000s, methods based on linear registration [3,
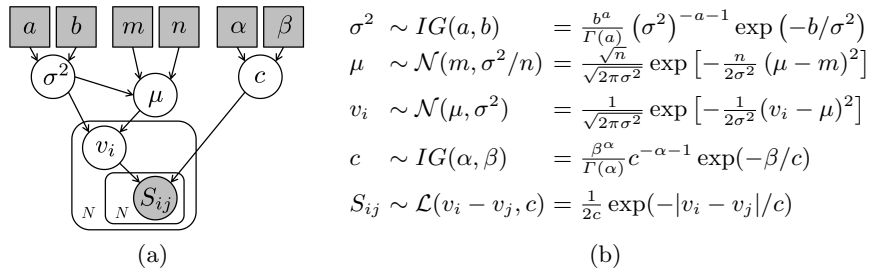
4] gained popularity: scaling factors are derived from an affine registration to a reference atlas, and multiplied by the ground truth ICV of the atlas to yield the estimates. These methods are implemented in widespread neuroimaging packages (e.g., FreeSurfer [5], FSL [6]). Other approaches rely on explicit segmentation of the intracranial cavity, typically with supervised methods based on parametric or non-parametric models. Representative examples of the former are the Bayesian segmentation [7] implemented in the SPM package [8], or variants thereof [9]. Examples of the latter include multi-atlas methods [10] and patch matching [11].

Although supervised methods can potentially yield better results, registration-based algorithms for ICV estimation are still widely used in adult human brain MRI, e.g., in FreeSurfer and FSL. The reason is threefold: they are fast; they do not require multimodal MRI pulse sequences; and they do not require labeled training data. Even if no ground truth ICV is available for the reference atlas, the ICV can still be estimated up to a constant scaling factor, which has no impact on the subsequent ICV correction. The main drawback of registration-based ICV estimation is that it is very sensitive to registration errors, which reflect directly on the ICV estimates through the determinant of the transformation matrix.

Motivation: limitations in ICV computation for developing brain and animal models. The literature on ICV estimation in the developing human brain is very sparse, but it is possible to use registration-based methods based on existing atlases, such as those described in [12, 13]. In animals, however, the availability of atlases – particularly for species other than mouse, rat and monkey – is very limited, especially for the developing brain. A particularly interesting case is rabbit models, which are increasingly important in neuroscience. One of the main application of rabbit models is the study of preterm birth, a problem with large economic and social impact [14, 15], and which is difficult to study in humans [16]. The only available rabbit atlas [17] is for the adult brain, and has no ICV information, as it was created from *ex vivo* brains without the skull.

Contribution. In this paper, we address the problem of ICV estimation by computing the ICV of all subjects/cases in a study *simultaneously*. The method has two major advantages: it does not require any labeled data, which is costly to collect (which is why many methods rely on semiautomatically generated silver standards, e.g., [4, 11, 10]); and is agnostic to species. Therefore, the proposed method readily enables application to developing brain and animal studies.

More specifically, we propose a probabilistic framework, in which the true, underlying ICVs are assumed to be independent samples of a Gaussian distribution with unknown parameters, and in which pairwise registrations yield noisy measurements of the ratios between these ICVs. Within this framework, we use Bayesian inference to compute the most likely ICVs. The information in the registrations enables estimation up to a scaling factor, which is disambiguated by the model hyperparameters. The proposed method preserves the advantages of registration-based algorithms described above, while being more robust to registration errors. Such robustness translates into increased statistical power and reduced sample sizes in subsequent analyses, as shown in our experiments below.

$$\sigma^2 \sim IG(a,b) \quad = \frac{b^a}{\Gamma(a)} \left(\sigma^2\right)^{-a-1} \exp\left(-b/\sigma^2\right)$$

$$\mu \quad \sim \mathcal{N}(m,\sigma^2/n) = \frac{\sqrt{n}}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{n}{2\sigma^2}\left(\mu - m\right)^2\right]$$

$$v_i \quad \sim \mathcal{N}(\mu,\sigma^2) \quad = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(v_i - \mu)^2\right]$$

$$c \quad \sim IG(\alpha,\beta) \quad = \frac{\beta^\alpha}{\Gamma(\alpha)} c^{-\alpha-1} \exp(-\beta/c)$$

$$S_{ij} \sim \mathcal{L}(v_i - v_j, c) = \frac{1}{2c} \exp(-|v_i - v_j|/c)$$

(a)   (b)

Fig. 1. Graphical model (a) and corresponding equations (b). Circles represent random variables, boxes represent hyperparameters, shaded elements are observed, and plates indicate replication. $\mathcal{N}$ is the Gaussian distribution, and $\mathcal{L}$ is the Laplace distribution.

2 Methods

2.1 Probabilistic framework

The graphical model of our probabilistic framework is shown in Figure 1. Let $\boldsymbol{v} = [v_1, \ldots, v_N]^t$ be a vector of log-transformed ICVs from N subjects; since we work with ratios, the logarithmic domain is more appropriate. These ICVs are assumed to be independent samples of a Gaussian distribution with unknown mean and variance (μ, σ^2). Prior knowledge on these two parameters is encoded in the hyperparameters of their prior distribution, which we choose to be a Normal Inverse Gamma (NIG), parameterized by $[m, n, a, b]^t$. This is the conjugate prior of a Gaussian with unknown mean and variance, and can be decomposed into an Inverse Gamma (IG) distribution on σ^2 (parameterized by $[a, b]^t$) and a Gaussian distribution on μ, with mean m and variance σ^2/n.

Now, we assume that we have a set $\mathcal{S}$ of subject pairs (i, j) for which pairwise scaling factors S_{ij} have been computed. Factor S_{ij} corresponds to a noisy estimate of the difference in log-ICVs between subjects i and j, i.e., $v_i - v_j$. The set of measurements does not need to exhaustively cover every (i, j), but must ensure that the adjacency matrix $A_{ij} = \delta(S_{ij} \neq 0)$ corresponds to a fully connected graph, such that there is always a path of scaling factors available between any two subjects. We use the matrix $\boldsymbol{S}$ to represent the measured scaling factors, with diagonal $S_{ii} = 0, \forall i$, and $S_{ij} = \text{NaN}$ if the scaling factor is not available, i.e., $(i, j) \notin \mathcal{S}$. These measurements, which are computed with a linear registration algorithm, correspond to the logarithm of the determinant of the estimated transform matrices. Here we assume that the registration method is symmetric, which yields an antisymmetric $\boldsymbol{S}$, i.e., $S_{ij} = -S_{ji}$. In order to make the algorithm robust against outliers, we further assume that the measurement errors are independent samples of a Laplace distribution (based on the robust ℓ_1 norm, rather than ℓ_2 that the Gaussian distribution relies on) with zero location and scale parameter c. Assuming zero location is appropriate, since the registration is symmetric. The scale parameter is unknown, but we place a conjugate prior distribution on it – the IG distribution, with hyperparameters α and β.

2.2   Bayesian inference

Our goal is to find the value $\hat{\boldsymbol{v}}$ that maximizes the posterior probability of the ICVs $\boldsymbol{v}$, given the pairwise scaling measurements and the hyperparameters:

$$
\hat{\boldsymbol{v}} = \underset{\boldsymbol{v}}{\mathrm{argmax}}\, p(\boldsymbol{v}|\boldsymbol{S}, m, n, a, b, \alpha, \beta) = \underset{\boldsymbol{v}}{\mathrm{argmax}}\, p(\boldsymbol{S}|\boldsymbol{v}, \alpha, \beta)p(\boldsymbol{v}|m, n, a, b)
$$

$$
= \underset{\boldsymbol{v}}{\mathrm{argmax}} \int_c p(\boldsymbol{S}|c, \boldsymbol{v})p(c|\alpha, \beta)dc \int_\mu \int_{\sigma^2} p(\boldsymbol{v}|\mu, \sigma^2)p(\mu, \sigma^2|m, n, a, b)d\mu d\sigma^2. \quad (1)
$$

Thanks to the conjugate priors, the two integrals in Equation 1 have closed-form solutions, so we can easily consider all possible values of the model parameters – weighted by their probabilities – in the estimation. The negated logarithm of this expression is the cost function to minimize ($\mathcal{C}$). In the appendix, we show that $\mathcal{C}$ is equal to:

$$
\mathcal{C}(\boldsymbol{v}; \boldsymbol{S}, m, n, a, b, \alpha, \beta) = (\alpha + |\mathcal{S}|) \log \left[\beta + \sum_{(i,j)\in\mathcal{S}} |S_{ij} - v_i + v_j| \right] + \ldots
$$

$$
\frac{2a + N}{2} \log \left[b + \sum_{i=1}^{N} \frac{(v_i - \bar{v})^2}{2} + \frac{Nn(\bar{v} - m)^2}{2(N + n)} \right] + Z(\alpha, \beta, |\mathcal{S}|, n, a, b, N), \quad (2)
$$

where Z is a term independent of $\boldsymbol{v}$, and $\bar{v} = (1/N)\sum_{i=1}^{N} v_i$ is the sample mean of $\boldsymbol{v}$. The final optimization problem is hence: $\mathrm{argmin}_{\boldsymbol{v}}\, C(\boldsymbol{v}; \boldsymbol{S}, m, n, a, b, \alpha, \beta)$. This is an unconstrained problem, which can be efficiently solved with standard algorithms; we used conjugate gradient [18] initialized with $v_i = m, \forall i$. The final solution is obtained by exponentiating $\hat{\boldsymbol{v}}$ to bring it back to the natural domain.

3   Experiments and results

3.1   Data

We used two brain MRI datasets in this study: one of rabbits, and one of humans. The rabbit dataset was acquired as part of a study seeking to understand the effects of steroids on fetuses. Scans from 7 rabbits (5 preterm born, 2 term) were acquired *in vivo* on a Bruker 9.4T animal scanner using a RARE T2 sequence (TR=42ms, TE=1000ms, 0.15mm resolution isotropic). The intracranial cavity was manually delineated by S.F. on the images, providing a ground truth for the ICVs. Since the size of the rabbit dataset is limited, we also performed experiments on a larger, more conventional human dataset consisting of T1 scans of 150 subjects from the Alzheimer's Disease Neuroimaging Initiative (ADNI): 77 Alzheimer's disease (AD) patients and 73 age-matched controls (EC). No ground truth ICVs was available for this dataset, so we used indirect validation techniques. While direct validation is generally preferable, indirect methods make the evaluation independent of segmentation errors, and do not require manual delineation, which is prohibitive for large datasets.
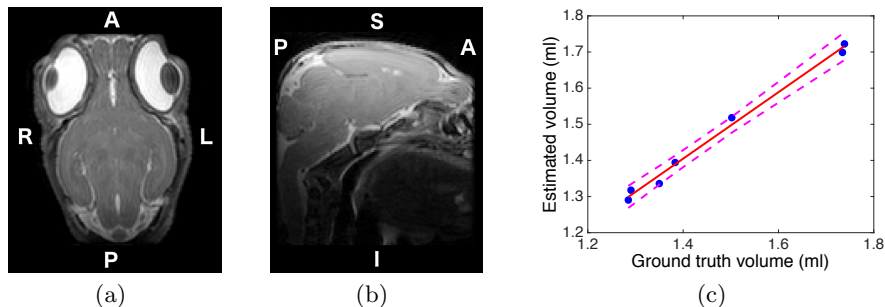
Fig. 2. (a) Sample axial and (b) sagittal slices from rabbit scans in our dataset; letters indicate orientation. (c) Scatter plot of ICVs: ground truth vs. estimated with our method, along with regression line (solid) and its 95% confidence interval (dashed).

3.2   Experiments on rabbit dataset

This experiment assesses the performance of the method in pediatric animal brain MRI, in which the availability of labeled data is extremely limited. We used correlation with ground truth volumes rather than absolute errors because the errors depend on m, whereas correlations do not, and also because ICV correction is based on correlation. We set the hyperparameters to $n = a = \alpha = 0.001$, $b = \beta = 0.1$, which represents very weak priors, such that the posterior distribution of the model parameters is mostly driven by the data. We set m to the mean value of the ground truth volumes, which has no effect on the correlation or ICV correction. For registration, we used a symmetric linear method based on block matching (NiftyReg [19]), applied to images downsampled by a factor of 4 in each dimension (for efficiency). Figure 2 shows sample MRI slices and the scatter plot for the ground truth and estimated ICVs, along with their linear regression. The correlation is very strong: $\rho = 0.996$, with p $\approx 10^{-6}$. These results are encouraging, but further validation is needed, given the small dataset.

3.3   Experiments on human dataset

We further evaluated our method indirectly with three experiments on the human dataset: we tested the strength of the correlation of hippocampal volume with age and ICV; we evaluated the ability of the ICV-corrected hippocampal volumes to discriminate EC from AD; and we tested the dependence of the performance on the set size $|\mathcal{S}|$.

We compared our proposed approach ("**PROP**") against performing no correction ("**NOCORR**") and four different methods: 1. FreeSurfer v5.3 [4] ("**FS**"): based on registration to a single template (MNI305) using cross-correlation; 2. Single atlas ("**SINGAT**"): a reimplementation of FS with NiftyReg at full resolution; 3. **SPM** [7]: we compute the ICV by summing the volumes of the gray matter, white matter and CSF, computed with SPM12 (default parameters); and 4. Non-linear ("**PROPNL**"): PROP with nonlinear registration [20]

Table 1. Correlation coefficients (ρ) between hippocampal volume and age/ICV, with 95% confidence intervals and p-values (null hypothesis: $\rho = 0$). For age, we have included the required sample size to detect the effect of age on hippocampal volume, with significance level 0.01 and power 0.99. Bold font indicates the top performing method.

Method	ρ_{age} (95% C.I.)	p-value	Sample Size	ρ_{icv} (95% C.I.)	p-value
NOCORR	-0.23 ([-0.44,-0.01])	0.0453	13	N/A	N/A
FS	-0.32 ([-0.51,-0.10])	0.0076	10	0.31 ([0.09,0.51])	0.009312
SINGAT	-0.38 ([-0.56,-0.16])	0.0014	9	0.47 ([0.27,0.63])	0.000042
SPM	-0.37 ([-0.55,-0.15])	0.0021	9	0.45 ([0.24,0.61])	0.000109
PROPNL	-0.32 ([-0.51,-0.10])	0.0067	10	0.40 ([0.19,0.58])	0.000615
PROP	**-0.40 ([-0.58,-0.18])**	**0.0008**	**8**	**0.51 ([0.32,0.67])**	**0.000007**
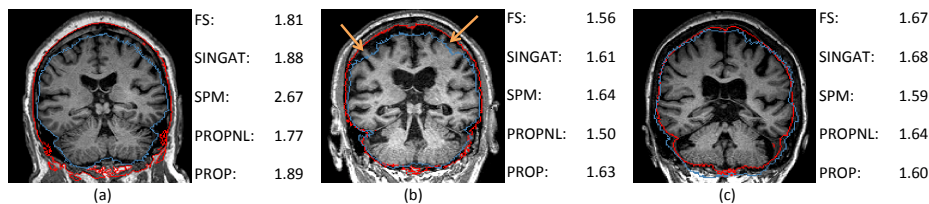
Fig. 3. Coronal slices of three subjects, and estimated ICVs (in liters). The intracranial cavity mask from SPM is contoured in red, and the one propagated from the reference atlas in blue. (a) Oversegmentation by SPM. (b) Good SPM segmentation, poor mask from reference (see areas pointed by arrows). (c) Good SPM segmentation; poor registration to the atlas negatively impacts the performance of all methods, except ours.

– we manually segmented the intracranial cavity in the atlas; nonlinearly propagated this mask to all subjects to create a silver standard; computed pairwise nonlinear registrations between subjects; propagated the silver standard masks; and computed S_{ij} as the difference in mask volume before and after registration.

By matching registration algorithms, SINGAT isolates the contribution of our framework to the improvement achieved over FS. SPM represents a much more complex, segmentation-based algorithm. PROPNL enables us to assess the potential improvement yielded by a more precise registration, which in principle could avoid bias from brain atrophy and disregard the contribution of extracranial regions. We used the same values of n, a, α, β as in the rabbit dataset, and set m to the mean ICV computed by FreeSurfer.

Effect of aging on hippocampal volume: Using only the healthy subjects, we first inspect the partial correlation between hippocampal volumes (computed with FreeSurfer, left-right averaged) and age/ICV, i.e., correcting for each other. For our method, we used all pairwise registrations. The results are shown in Table 1, and sample outputs in Figure 3. All methods increase the correlations between hippocampal volume and age. SINGAT outperforms FreeSurfer, thanks to the more robust registration. Despite being a more complex method, the performance of SPM is on par with that of SINGAT, because SPM sometimes includes large portions of bone in the segmentation due to poor contrast in T1

Table 2. AD/EC classification: AUROC, accuracy at elbow and Delong's p for comparison of the AUROC with that of the method in the corresponding column.

Method	AUROC	Acc. Elbow	Vs.	FS	SINGAT	SPM	PROPNL	PROP
NOCORR	0.905	0.847		0.0311	0.0077	0.0052	0.0031	0.0008
FS	0.911	0.840		*	0.0193	0.0148	0.0045	0.0005
SINGAT	0.915	0.847	DeLong p	*	*	0.0716	0.0461	0.0004
SPM	0.921	0.873		*	*	*	0.9962	0.1052
PROPNL	0.921	**0.880**		*	*	*	*	0.0468
PROP	**0.927**	0.873		*	*	*	*	*
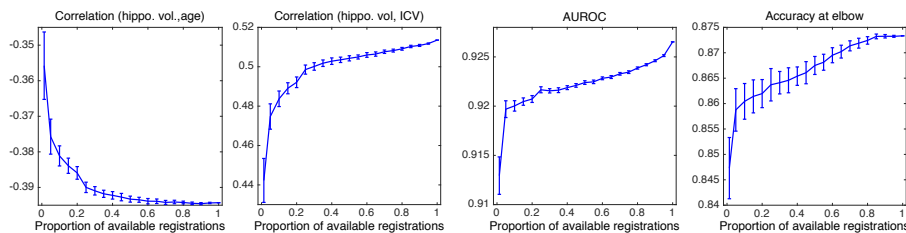
Fig. 4. Performance vs. % of maximum set size; error bars span two standard deviations

MR (see Figure 3a). We observed a similar effect in PROPNL, exacerbated by the fact that errors propagate along two registrations (see Figure 3b). Our method produces the highest correlation, with a mild, borderline significant (Steiger's test [21]: p = 0.05) improvement over the second-to-best method (SINGAT). Compared with the widely used FreeSurfer, the improvement is noticeable: $\Delta\rho = 0.08$ (Steiger's p = 0.01), and sample size reduced from 10 to 8. A similar trend can be observed for the correlation between ICV and hippocampal volume.

Alzheimer's disease classification: We computed the area under the ROC curve (AUROC) and the accuracy at its elbow for classifiers based on thresholding ICV/age corrected hippocampal volumes. We used DeLong's test [22] to compare AUROCs. The results are shown in Table 2. Our method provides the highest AUROC, with significant improvement with respect to all others, except SPM. It also provides the second-to-best accuracy at elbow, after PROPNL.

Performance as a function of the number of available registrations: Since the number of registrations increases quickly with N, it is useful to test how robust our algorithm is against missing scaling factors, to see if computational cost can be reduced by computing only a subset of the registrations, without significant loss of accuracy. For a number of set sizes, we drew 100 random samples (with rejection to ensure fully connected graphs); computed the partial correlations, AUROCs and accuracies; and calculated their averages across the samples. The results are shown in Figure 4. The correlations increase quickly in the beginning, and plateau at around 40-50% of the maximum set size. The AUROC and accuracy improve slowly until 100%, as they are more sensitive to volume estimate changes in samples closer to the decision boundary.

4   Discussion and conclusion

We propose an ICV estimation method that does not require labeled data and is agnostic to the imaged species, and which inherits the advantages of registration-based ICV estimation – while being more robust against registration errors. This is despite using linear registration, which could be biased by extracranial and intracranial changes (e.g., neck size, atrophy); such bias was not observed in our experiments. Our approach can be combined with any registration method; we used a symmetric algorithm to save half of the registrations when building $\boldsymbol{S}$.

Despite the high number of required registrations, our method is not too computationally expensive, as low-resolution registrations run in 3-4 seconds on a single core. Moreover, $\boldsymbol{S}$ can be precomputed such that, when a new scan arrives, only N new registrations are required. We also tested a non-linear version, but the increased flexibility did not compensate for the introduced registration errors and increased computational cost. The inference algorithm takes just a few seconds, which is negligible compared with the running time of the registrations.

The experiments in this paper have shown that our method outperforms single atlas ICV estimation in analyses like effect of age and AD classification. Moreover, it also outperforms the much more complex SPM, in spite of using linear registrations. Even though the contribution of the raw hippocampal volume is larger that that of the ICV to such analyses, our method can still provide a moderate, statistically significant improvement. Future work will consider more complex distributions for the ICVs, e.g., conditioned on sex and gestational age.

Appendix: Details of the inference algorithm

Replacing $p(\boldsymbol{S}|c,\boldsymbol{v})$ and $p(c|\alpha,\beta)$ (Figure 1b) in the first integral of Equation 1:

$$\int_c (2c)^{-|\mathcal{S}|} \exp\left(-\frac{1}{c}\sum_{(i,j)\in\mathcal{S}}|S_{ij}-v_i+v_j|\right)\frac{\beta^\alpha}{\Gamma(\alpha)}c^{-\alpha-1}\exp(-\beta/c)dc. \qquad (3)$$

Defining $\alpha'=\alpha+|\mathcal{S}|$ and $\beta'=\beta+\sum_{(i,j)\in\mathcal{S}}|S_{ij}-v_i+v_j|$, Equation 3 becomes:

$$\frac{1}{2^{|\mathcal{S}|}}\frac{\Gamma(\alpha')}{\Gamma(\alpha)}\frac{\beta^\alpha}{\beta'^{\alpha'}}\int_c\frac{\beta'^{\alpha'}}{\Gamma(\alpha')}c^{-\alpha'-1}\exp(-\beta'/c)dc=\frac{1}{2^{|\mathcal{S}|}}\frac{\Gamma(\alpha')}{\Gamma(\alpha)}\frac{\beta^\alpha}{\beta'^{\alpha'}}, \qquad (4)$$

as the integral is over the probability density of $IG(\alpha',\beta')$ and thus equal to 1.

For the second integral in Equation 1 (over μ and σ^2), substitution of the expressions for the probabilities (again, see Figure 1b in the paper) yields:

$$\int_\mu\int_{\sigma^2}\frac{1}{(2\pi\sigma^2)^{N/2}}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^N(v_i-\mu)^2\right)\cdots$$

$$\times\frac{b^a}{\Gamma(a)}\left(\sigma^2\right)^{-a-1}\exp\left(-b/\sigma^2\right)\frac{\sqrt{n}}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{n}{2\sigma^2}\left(\mu-m\right)^2\right]d\mu d\sigma^2. \qquad (5)$$

We now define $m' = (nm + N\bar{v})/(n + N)$, $n' = n + N$, $a' = a + \frac{N}{2}$, and:

$$b' = b + \frac{1}{2}\sum_{i=1}^{N}(v_i - \bar{v})^2 + \frac{nN}{n+N}\frac{(\bar{v}-m)^2}{2},$$

where $\bar{v}$ is the average of $\boldsymbol{v}$. Then, Equation 5 becomes:

$$\int_\mu \int_{\sigma^2} \frac{b'^{a'}}{\Gamma(a')}\left(\sigma^2\right)^{-a'-1}\exp\left(-b'/\sigma^2\right)\frac{\sqrt{n'}}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{n'}{2\sigma^2}\left(\mu - m'\right)^2\right]d\mu d\sigma^2 \dots$$

$$\times \frac{1}{(2\pi)^{N/2}}\sqrt{\frac{n}{n'}}\frac{b^a}{b'^{a'}}\frac{\Gamma(a')}{\Gamma(a)} = \frac{1}{(2\pi)^{N/2}}\sqrt{\frac{n}{n'}}\frac{b^a}{b'^{a'}}\frac{\Gamma(a')}{\Gamma(a)}, \tag{6}$$

since the integral is over the probability density function of $NIG(m', n', a', b')$ and hence equal to 1.

Combining Equations 4 and 6, the problem in Equation 1 becomes:

$$\hat{\boldsymbol{v}} = \underset{\boldsymbol{v}}{\mathrm{argmax}}\ \frac{1}{2^{|\mathcal{S}|}}\frac{\Gamma(\alpha')}{\Gamma(\alpha)}\frac{\beta^\alpha}{\beta'^{\alpha'}}\frac{1}{(2\pi)^{N/2}}\sqrt{\frac{n}{n'}}\frac{b^a}{b'^{a'}}\frac{\Gamma(a')}{\Gamma(a)}$$

$$= \underset{\boldsymbol{v}}{\mathrm{argmax}}\left(\beta'^{\alpha'}b'^{a'}z(\alpha, \beta, |\mathcal{S}|, n, a, b, N)\right)^{-1},$$

where z groups the terms independent of $\boldsymbol{v}$. Taking the negated logarithm:

$$\mathcal{C} = \alpha'\log\beta' + a'\log b' + \log z.$$

Substituting a', b', α' and β' into this equation, and defining $Z = \log z$, we finally obtain the cost function in Equation 2.

References

1. Sanfilipo, M.P., Benedict, R.H., Zivadinov, R., Bakshi, R.: Correction for intracranial volume in analysis of whole brain atrophy in multiple sclerosis: the proportion vs. residual method. Neuroimage **22**(4) (2004) 1732–1743
2. Lemieux, L., Hagemann, G., Krakow, K., Woermann, F.G.: Fast, accurate, and reproducible automatic segmentation of the brain in T1-weighted volume MRI data. Magnetic Resonance in Medicine **42**(1) (1999) 127–135
3. Smith, S.M., Zhang, Y., Jenkinson, M., Chen, J., Matthews, P., Federico, A., De Stefano, N.: Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. Neuroimage **17**(1) (2002) 479–489
4. Buckner, R.L., Head, D., Parker, J., Fotenos, A.F., Marcus, D., Morris, J.C., Snyder, A.Z.: A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: reliability and validation against manual measurement of total intracranial volume. Neuroimage **23**(2) (2004) 724–738
5. Fischl, B.: Freesurfer. Neuroimage **62**(2) (2012) 774–781
6. Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E., et al.: Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage **23** (2004) S208–S219

7. Ashburner, J., Friston, K.J.: Unified segmentation. Neuroimage **26**(3) (2005) 839–851
8. Penny, W.D., Friston, K.J., Ashburner, J.T., Kiebel, S.J., Nichols, T.E.: Statistical parametric mapping: the analysis of functional brain images. Academic Press (2011)
9. Keihaninejad, S., Heckemann, R.A., Fagiolo, G., Symms, M.R., Hajnal, J.V., Hammers, A.: A robust method to estimate the intracranial volume across MRI field strengths (1.5T and 3T). Neuroimage **50**(4) (2010) 1427–1437
10. Huo, Y., Asman, A.J., Plassard, A.J., Landman, B.A.: Simultaneous total intracranial volume and posterior fossa volume estimation using multi-atlas label fusion. Human Brain Mapping **2** (2017) 599–616
11. Manjón, J.V., Eskildsen, S.F., Coupé, P., Romero, J.E., Collins, D.L., Robles, M.: Nonlocal intracranial cavity extraction. Journal of Biomedical Imaging, Article ID 820205 (2014)
12. Kuklisova-Murgasova, M., Aljabar, P., Srinivasan, L., Counsell, S.J., Doria, V., Serag, A., Gousias, I.S., Boardman, J.P., Rutherford, M.A., Edwards, A.D., et al.: A dynamic 4d probabilistic atlas of the developing brain. NeuroImage **54**(4) (2011) 2750–2763
13. Licandro, R., Langs, G., Kasprian, G., Sablatnig, R., Prayer, D., Schwartz, E.: Longitudinal diffeomorphic atlas learning for fetal brain tissue labeling using geodesic regression. Brain **2011** (2011)
14. Tyson, J.E., Parikh, N.A., Langer, J., Green, C., Higgins, R.D.: Intensive care for extreme prematuritymoving beyond gestational age. New England Journal of Medicine **358**(16) (2008) 1672–1681
15. Blencowe, H., Cousens, S., Oestergaard, M.Z., Chou, D., Moller, A.B., Narwal, R., Adler, A., Garcia, C.V., Rohde, S., Say, L., et al.: National, regional, and worldwide estimates of preterm birth rates in the year 2010 with time trends since 1990 for selected countries: a systematic analysis and implications. The Lancet **379**(9832) (2012) 2162–2172
16. Liu, L., Johnson, H.L., Cousens, S., Perin, J., Scott, S., Lawn, J.E., Rudan, I., Campbell, H., Cibulskis, R., Li, M., et al.: Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. The Lancet **379**(9832) (2012) 2151–2161
17. Muñoz-Moreno, E., Arbat-Plana, A., Batalle, D., Soria, G., Illa, M., Prats-Galino, A., Eixarch, E., Gratacos, E.: A magnetic resonance image based atlas of the rabbit brain for automatic parcellation. PLoS One **8**(7) (2013) e67418
18. Shewchuk, J.R.: An introduction to the conjugate gradient method without the agonizing pain (1994)
19. Modat, M., Cash, D.M., Daga, P., Winston, G.P., Duncan, J.S., Ourselin, S.: Global image registration using a symmetric block-matching approach. Journal of Medical Imaging **1**(2) (2014) 024003–024003
20. Modat, M., Ridgway, G.R., Taylor, Z.A., Lehmann, M., Barnes, J., Hawkes, D.J., Fox, N.C., Ourselin, S.: Fast free-form deformation using graphics processing units. Computer methods and programs in biomedicine **98**(3) (2010) 278–284
21. Steiger, J.H.: Tests for comparing elements of a correlation matrix. Psychological bulletin **87**(2) (1980) 245–251
22. DeLong, E.R., DeLong, D.M., Clarke-Pearson, D.L.: Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics (1988) 837–845