



Automatic skull segmentation from MR images for realistic volume conductor models of the head: Assessment of the state-of-the-art

Nielsen, Jesper Duemose; Madsen, Kristoffer Hougaard; Puonti, Oula; Siebner, Hartwig R.; Bauer, Christian; Madsen, Camilla Gøbel; Saturnino, Guilherme B.; Thielscher, Axel

Published in:
NeuroImage

Link to article, DOI:
[10.1016/j.neuroimage.2018.03.001](https://doi.org/10.1016/j.neuroimage.2018.03.001)

Publication date:
2018

Document Version
Peer reviewed version

[Link back to DTU Orbit](#)

Citation (APA):
Nielsen, J. D., Madsen, K. H., Puonti, O., Siebner, H. R., Bauer, C., Madsen, C. G., Saturnino, G. B., & Thielscher, A. (2018). Automatic skull segmentation from MR images for realistic volume conductor models of the head: Assessment of the state-of-the-art. *NeuroImage*, 174, 587-598.
<https://doi.org/10.1016/j.neuroimage.2018.03.001>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

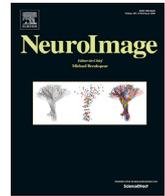
- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Automatic skull segmentation from MR images for realistic volume conductor models of the head: Assessment of the state-of-the-art

Jesper D. Nielsen^{a,b,e}, Kristoffer H. Madsen^{a,b,e}, Oula Puonti^{a,c}, Hartwig R. Siebner^{a,d},
Christian Bauer^a, Camilla Gøbel Madsen^a, Guilherme B. Saturnino^{a,c}, Axel Thielscher^{a,c,*}

^a Danish Research Centre for Magnetic Resonance, Centre for Functional and Diagnostic Imaging and Research, Copenhagen University Hospital Hvidovre, Hvidovre, Denmark

^b Department of Applied Mathematics and Computer Science, Technical University of Denmark, Kgs. Lyngby, Denmark

^c Center for Magnetic Resonance, Department of Electrical Engineering, Technical University of Denmark, Kgs Lyngby, Denmark

^d Department of Neurology, Copenhagen University Hospital Bispebjerg, Copenhagen, Denmark

^e Sino-Danish College, University of Chinese Academy of Sciences, Beijing, 100190, PR China

ARTICLE INFO

Keywords:

Skull segmentation
Volume conductor model
Forward modeling
Transcranial brain stimulation
Electroencephalography

ABSTRACT

Anatomically realistic volume conductor models of the human head are important for accurate forward modeling of the electric field during transcranial brain stimulation (TBS), electro- (EEG) and magnetoencephalography (MEG). In particular, the skull compartment exerts a strong influence on the field distribution due to its low conductivity, suggesting the need to represent its geometry accurately. However, automatic skull reconstruction from structural magnetic resonance (MR) images is difficult, as compact bone has a very low signal in magnetic resonance imaging (MRI). Here, we evaluate three methods for skull segmentation, namely FSL *BET2*, the *unified segmentation* routine of SPM12 with extended spatial tissue priors, and the *skullfinder* tool of BrainSuite. To our knowledge, this study is the first to rigorously assess the accuracy of these state-of-the-art tools by comparison with CT-based skull segmentations on a group of ten subjects. We demonstrate several key factors that improve the segmentation quality, including the use of multi-contrast MRI data, the optimization of the MR sequences and the adaptation of the parameters of the segmentation methods. We conclude that FSL and SPM12 achieve better skull segmentations than BrainSuite. The former methods obtain reasonable results for the upper part of the skull when a combination of T1- and T2-weighted images is used as input. The SPM12-based results can be improved slightly further by means of simple morphological operations to fix local defects. In contrast to FSL *BET2*, the SPM12-based segmentation with extended spatial tissue priors and the BrainSuite-based segmentation provide coarse reconstructions of the vertebrae, enabling the construction of volume conductor models that include the neck. We exemplarily demonstrate that the extended models enable a more accurate estimation of the electric field distribution during transcranial direct current stimulation (tDCS) for montages that involve extraencephalic electrodes. The methods provided by FSL and SPM12 are integrated into pipelines for the automatic generation of realistic head models based on tetrahedral meshes, which are distributed as part of the open-source software package SimNIBS for field calculations for transcranial brain stimulation.

Introduction

Volume conductor models of the head are key components of several neuroscientific methods such as electric field simulations for transcranial brain stimulation (TBS) and source localization in electro- (EEG) and magnetoencephalography (MEG). The anatomical accuracy of the head models has a strong influence on the accuracy of the calculated field

distributions (Cho et al., 2015; Dannhauer et al., 2011; Eichelbaum et al., 2014; Lanfer et al., 2012; Montes-Restrepo et al., 2014; Wolters et al., 2006) and attempts to use individualized models based on structural magnetic resonance (MR) images are gaining momentum (Vorwerk et al., 2014). Recently available open-source software, including FSL (Smith et al., 2004), BrainSuite (Shattuck and Leahy, 2002), and SPM12 (<http://www.fil.ion.ucl.ac.uk/spm/>), facilitates the adoption of this approach by

* Corresponding author. Danish Research Centre for Magnetic Resonance, Centre for Functional and Diagnostic Imaging and Research, Copenhagen University Hospital Hvidovre, Section 714, Kettegaard Allé 30, 2650 Hvidovre, Denmark.

E-mail address: axelt@drcmr.dk (A. Thielscher).

<https://doi.org/10.1016/j.neuroimage.2018.03.001>

Received 17 August 2017; Received in revised form 27 February 2018; Accepted 1 March 2018

Available online xxx

1053-8119/© 2018 Elsevier Inc. All rights reserved.

offering automatic segmentation procedures for the head. These tools have been integrated into software pipelines for the forward modeling of electric fields for TBS (e.g., SimNIBS; Thielscher et al., 2015) and EEG/MEG (e.g., FieldTrip; Oostenveld et al., 2011 and Brainstorm; Tadel et al., 2011). Accurate modeling of the skull compartment is an important aspect of individualized head models as the skull strongly shapes the forward solution due to its low conductivity (Dannhauer et al., 2011; Indahlstari et al., 2016; Lanfer et al., 2012; Montes-Restrepo et al., 2014; Stenroos et al., 2014). However, its automatic segmentation is still a major challenge, as the compact bone parts have a very low signal in conventional magnetic resonance imaging (MRI) sequences.

While the performance of most software packages in segmenting the brain have been thoroughly validated, similar tests are scarce for the skull. Thus, in this study we investigate the performance of three widely used neuroimaging software packages, FSL (Smith et al., 2004), BrainSuite (Shattuck and Leahy, 2002), and SPM12 (<http://www.fil.ion.ucl.ac.uk/spm/>). Specifically, we assess FSL *BET2* which includes the *BET* and *betsurf* tools (Pechaud et al., 2006), BrainSuite *skullfinder* (Dogdas et al., 2005), and the *unified segmentation* routine (Ashburner and Friston, 2005) implemented in SPM12. The latter was tested with spatially extended tissue priors in order to avoid clipping of the lower parts of the head (Huang et al., 2013). In contrast to BrainSuite, FSL and SPM12 support the use of multi-spectral MRI for segmentation. We therefore also compare the results when basing the segmentations on a single, high-resolution T1-weighted structural MR image, as often acquired in neuroimaging studies and used in clinical standard of care, versus a combination of high-resolution T1- and T2-weighted MR images. In addition, for the SPM12-based segmentations, we assess to which extent the results can be improved when applying morphological operations to “clean up” the raw segmentations. We test the quality of the segmentations by systematic comparisons against skull segmentations from computed tomography (CT) scans of the same subjects. To the best of our knowledge, this study is the first to rigorously assess the performance of these tools on skull segmentation and thus serves as important evaluation of the state-of-the-art on this topic.

Whereas the main focus of the paper is on skull segmentation, we further compare the accuracy of the reconstructed brain surfaces derived from SPM12-based segmentations with surfaces obtained using FreeSurfer 5.3.0 (Dale et al., 1999; Fischl et al., 1999). Finally, we exemplarily demonstrate the importance of selecting adequate MRI sequence parameters and adjusting the parameters of the SPM12 segmentation routine to the properties of the MR images in order to achieve robust and accurate results, particularly in non-brain regions. As such, our study gives useful guidelines for the adoption of individualized volume conductor models in neuroscientific research.

Material and methods

Subjects

Ten healthy subjects (five Caucasians [three males], five East Asians [two males], 20–50 years old; 28.9 ± 9.3 [mean age \pm SD]) were included in this study. They had no previous history of neurological or psychiatric disorders and were screened for contraindications to MRI and CT. In addition, the structural MR images were checked by a radiologist. Written informed consent was obtained from all participants prior to the scans. The study was approved by the Ethical Committee of the Capital Region of Denmark.

Data acquisition

The structural MR images were acquired on a 3.0 T Philips Achieva MRI scanner using a 32-channel head coil. A high-resolution T1-weighted scan (T1w; repetition time = 6.0 ms; echo time = 2.7 ms; flip angle = 8°; 245 sagittal slices; matrix = 288×288 ; field of view = $245 \times 245 \times 208$ mm³; voxel size = 0.85 mm³; bandwidth = 299.3 Hz; selective water excitation;

SENSE factor 2 along AP direction) and a high-resolution T2-weighted scan (T2w; repetition time = 2500 ms; echo time = 272 ms; flip angle = 90°; 224 sagittal slices; matrix = 288×288 ; field of view = $245 \times 245 \times 190$ mm³; voxel size = 0.85 mm³; bandwidth = 880.6 Hz, SENSE factor 2 along AP and 1.8 along RL) were performed. These sequence parameters were chosen to give good results for the skull segmentation using FSL *BET2* and the FreeSurfer-based brain segmentation, and are based on our prior experience with the tools (Windhoff et al., 2013). Specifically, the readout bandwidth of the T1w image was chosen low enough to give a good signal-to-noise in the brain region. In addition, the parameters were selected to give a good contrast between gray matter (GM) and white matter (WM). Selective water excitation was chosen to ensure that most of the signal of the fatty spongy bone of the skull was suppressed. Otherwise, the signal of the spongy bone would have been merged with that of the pial surface of the brain at some positions due to the increased fat shift at low bandwidth, rendering an accurate segmentation of the inner skull boundary difficult. The fat shift refers to a systematic displacement of the spatial position of fat along the frequency encoding direction in the MR image that is caused by slightly different resonance frequencies of water and fat (King, 2004). The T2w image had a sufficiently high readout bandwidth to ensure a small fat shift that allowed for a good separation between the scalp, the compact and spongy bone, and the cerebro-spinal fluid (CSF). It is reasonable to assume that the above strategy to ensure a good spatial separation and a good contrast between the tissue classes was at least not detrimental for the SPM12- and BrainSuite-based segmentations even though we did not specifically optimize the MRI parameters for these.

The low-dose CT scans were acquired on a Siemens Biograph mCT (PET-CT) with axial slices having a voxel size of 0.42×0.42 mm² and a field of view of 215×215 mm². The resolution along the Z-direction was 0.60 mm³. The extent in Z-direction was individually adjusted to cover the complete neck while minimizing the radiation dose, and was on average 236 mm³. The tube current-time product (number of electrons per helical rotation) was 115 mAs, the tube potential (intensity) was 80 KeV, and the maximum effective dose was below 0.35 mSv for all participants.

Segmentation and preparation of CT images

An initial segmentation of the CT images was provided by thresholding the image at an intensity of 450 Hounsfield units (HU), a value, which seemed a reasonable compromise between retaining signal from the bone parts and suppressing noise. In order to achieve a robust segmentation of the bone regions while keeping the details as good as possible, the Chan-Vese level set method (Chan and Vese, 2001) was used, as implemented in the 2D/3D image segmentation toolbox (Zhang et al., 2008) for MATLAB. This method minimizes an energy function, which combines an image term penalizing large intensity variations within the segmented area and a curvature term penalizing the curve length resulting in smooth segmentations. As a final step, a few morphological operations were performed to fix minor inaccuracies in the segmentations (e.g., remove small holes). Visual inspection of the resulting bone masks confirmed the very good quality of the final segmentations, suggesting that they were suitable as “ground truth” for validating the MRI-based segmentations.

Segmentation of MR images

Three different methods for obtaining skull segmentations from MR images were tested. Their choice was based on own prior work (Windhoff et al., 2013) and results reported in related studies (Perdue and Diamond, 2014; Huang et al., 2013). The first method is based on tools provided by FSL 5.0.9 (<https://fsl.fmrib.ox.ac.uk/fsl/fslwiki>). Specifically, *BET* was employed to generate a brain mask, which was then used by *betsurf* as initialization to detect the inner and outer skull boundaries (Pechaud et al., 2006). *Betsurf* uses local intensity profiles along vectors perpendicular to the brain surface to identify the boundaries, based on the

assumption that the intensity variations along this normal follow a specific pattern, i.e. low (compact bone), high (spongy bone), low (compact bone). This approach recovers the skull regions directly surrounding the brain but fails to segment the skull parts belonging to the facial region and the vertebrae. In case both T1w and T2w images were used for segmentation, the T2w image was initially coregistered to the T1w image using the *FLIRT* registration tool (Jenkinson et al., 2002; Jenkinson and Smith, 2001). As the tools were integrated into our pipeline for the automatic construction of volume head meshes (details are stated below), an additional post-processing step was used to ensure a minimal distance between the inner and outer skull boundaries of 1 mm in case *betsurf* had failed to accurately detect the boundaries. We also ensured that there is at least a one voxel thick layer of CSF between the skull and gray matter.

The second method, *skullfinder* from BrainSuite 17a (<http://brainsuite.org/>), consists of a series of morphological operations (Dogdas et al., 2005) following an initial skull-stripping step (Shattuck et al., 2001). *Skullfinder* automatically computes intensity thresholds used for initial skull and scalp segmentation, which is a starting point for the morphological operations. However, as these thresholds yielded poor skull segmentations on our dataset, we manually optimized the thresholds using the T1w images of three subjects and then applied those thresholds to all data.

The third method for skull segmentation was based on the *unified segmentation* routine (Ashburner and Friston, 2005) implemented in SPM12 (<http://www.fil.ion.ucl.ac.uk/spm/>) for MATLAB which combines spatial normalization to MNI space, intensity inhomogeneity correction, and tissue segmentation into one model. It uses a Gaussian mixture model for modeling tissue intensities and so fits one (or several) Gaussians to the intensity histogram of each tissue class. This is guided by a spatial prior on the probability of different tissues at a given position. Finally, SPM12 employs light Markov random field post-processing (default setting) to clean up the segmentations. The default spatial prior in SPM12 does not include the neck region. Thus, in order to avoid clipping of the lower parts of the head, we employed the spatially extended tissue priors introduced by (Huang et al., 2013). SPM12 outputs probability maps for each tissue, which were binarized by assigning the tissue type with the largest probability to each voxel. Since several local defects were typically still present in these binarized segmentations, we applied additional post-processing steps to correct the masks. The post-processing consists of simple morphological operations, similar to the strategy suggested in (Huang et al., 2013), including *closing* (dilation followed by erosion) of the bone structure to remove small holes, *eroding* the skin compartment, keeping only the largest connected component, and *dilating* it again to recover its original size (see the Appendix for details). These steps are automatically applied in our pipeline and help to improve the skull reconstruction, mostly by correcting the false labeling of spongy bone as skin. We assessed the quality of the segmentations both before and after these operations, and further tested the approach with only a T1w image or both T1w and T2w images as input.

In addition to testing the accuracy of the skull segmentations, further aspects of the SPM12-based segmentations were assessed:

- Binarized GM masks were extracted from the SPM12-based segmentation of the T1w images and compared with GM reconstructions obtained via FreeSurfer 5.3.0. It was also tested how much using the Computational Anatomy Toolbox 12 (CAT12) toolbox (<http://www.neuro.uni-jena.de/cat>) improved the SPM12-based GM segmentations. FreeSurfer has been shown to achieve reliable reconstructions of the GM and WM surfaces (Eggert et al., 2012; Han et al., 2006). We therefore chose it as suitable tool to assess the quality of the SPM12- and CAT12-based segmentations, but do not claim that the FreeSurfer-based segmentations should be considered as “ground truth”. In order to allow for a comparison of the results, binarized GM masks were obtained from the surface reconstructions provided by FreeSurfer by filling in the region between the GM and WM surfaces, with voxel size defined by the MR image, using custom-written

software. A similar strategy was employed for the results of CAT12, which provides a surface that delineates the middle of the GM sheet and additionally an estimate of the local GM thickness as output. Custom-written software implemented in Python 2.7 was used to reconstruct the GM and WM surfaces from the CAT12 results and then filling in the regions between the two surfaces. The WM surface was constructed from the WM segmentation provided by CAT12 using the marching cubes algorithm (Lewiner et al., 2003; van der Walt et al., 2014). The GM surface was constructed by moving the nodes of the central GM estimate outwards by half of the GM thickness to generate the pial GM surface. In the latter case, the movement was locally stopped whenever necessary to prevent self-intersections.

- MR images suffer from inhomogeneous intensity profiles due to imperfections in the radio-frequency (RF) excitation and receive coils so that a given tissue type will have varying intensities across the image which can cause errors in the resulting segmentations if not accounted for (Van Leemput et al., 1999). SPM12 deals with this problem by estimating the MRI-related intensity inhomogeneities, typically referred to as “the bias field” (Wells et al., 1996), during the segmentation process. The estimated bias field is then used to correct the intensity profile of the MR scan. Estimation of the bias field can, however, be problematic when performed across the whole head rather than only the brain as is usually done in neuroimaging. We found that careful adjustment of the regularization parameter, which controls how quickly the estimated bias field is allowed to change across space, helped to improve the segmentation in particular of the lower parts of the head. Given the practical relevance of this topic, we exemplarily demonstrate the impact of the regularization parameter on the segmentation results.
- The sequence parameters were chosen based on prior experience (Windhoff et al., 2013), and the rationale behind this choice is outlined in the discussion. To demonstrate the importance of careful selection and testing of the employed MRI sequence, we show an example using a T1w image from a publicly accessible dataset (obtained from the OpenfMRI database, accession number ds000171).

Validation and comparison of the MRI-based segmentations

We validated the skull reconstructions of all three methods against those obtained from CT scans. Images were downsampled to 1mm^3 voxels and the CT scans were coregistered to the T1w images using *FLIRT*. Subsequently, an additional, nonlinear coregistration step was performed using the Elastix software package (Klein et al., 2010) to properly coregister the neck. After applying *FLIRT* and Elastix, careful visual inspection was performed to ensure the quality of the coregistrations. In addition, we compared the SPM12-based GM segmentations (both with and without CAT12) against FreeSurfer-based GM segmentations. The full set of analyses is listed in Table 1. For the skull analysis, neither changing the downsampling parameter in SPM12 from threefold to none, nor using CAT12 affected the result of the SPM12-based segmentations (both with and without post-processing using morphological operations); hence, we report results only for the default settings. Also for the GM analysis, changing the downsampling parameter in SPM12 did not affect the result, so that we report the results only for the default settings. The binarized posteriors were used for all comparisons.

For validating the bone reconstructions against the CT scans, we used the Dice coefficient and the modified Hausdorff distance (Dubuisson and Jain, 1994) to provide summary scores of the overall quality of the segmentations. The Dice coefficient measures the similarity of sets, A and B , by their overlap and is defined as

$$D = \frac{2|A \cap B|}{|A| + |B|}$$

where $|X|$ denotes the number of points in set X . Thus, $0 \leq D \leq 1$ with perfect overlap resulting in $D = 1$. The modified Hausdorff distance

$H(A, B)$ measures the average of two directed Hausdorff distances $h_m(A, B)$ and $h_m(B, A)$ between the borders of sets A and B , and is defined as

$$H(A, B) = \text{mean}\{h_m(A, B), h_m(B, A)\}$$

with the directed Hausdorff distance

$$h_m(A, B) = \frac{1}{|A|} \sum_{a \in A} \min_{b \in B} \{d(a, b)\}$$

being the average minimum distance from set A to set B for some suitable distance measure d (Dubuisson and Jain, 1994). In our case, d denotes the Euclidean distance and the boundary voxels (surface) were obtained by subtraction with the eroded segmentation. The reason for reporting both the Dice score and the modified Hausdorff distance is that the Dice score is dominated by large clusters of voxels and is not very sensitive to differences along the surfaces of the segmentations (which are often the most difficult parts to properly model).

In addition to providing the summary scores, we also produced spatial maps of false positive and negative rates of the correspondence between the MRI and CT segmentations to document how the segmentation accuracy varied between different parts of the head. These maps were obtained by labeling a voxel that was incorrectly classified as bone in the MRI-based segmentations as a false positive and a missed bone voxel correspondingly as a false negative. Finally, the subject-specific false positive and negative maps were transformed to the MNI space and averaged, resulting in false positive and negative rates across subjects for each anatomical location. We produced similar maps for the accuracy of the SPM12- and CAT12-based GM segmentations by comparing them to the FreeSurfer GM segmentations.

For the SPM12- and BrainSuite-based skull segmentations, we report the Dice coefficients and modified Hausdorff distances for the complete head, including the vertebrae, and in addition for the upper part of the head only. We included the latter results to allow for a better comparison with the FSL-based skull segmentations, which do not include the vertebrae.

Generation of head volume meshes and TDCS field simulations

The FSL- and SPM12-based methods for skull segmentation were integrated into software pipelines for the automatic generation of individualized volume conductor models of the head. These head models can be employed in our open-source software SimNIBS 2 (www.simnibs.org; Thielscher et al., 2015) that uses the finite element method (FEM) for the calculation of the electric fields generated by transcranial magnetic stimulation (TMS) and transcranial direct current stimulation (TDCS). Their usefulness has further been demonstrated in field calculations for electro- and magnetoencephalography (EEG and MEG) based on the boundary element method (Stenroos and Nummenmaa, 2016), as well as for FEM-based forward models for EEG (Ziegler et al., 2014).

The software pipeline *mri2mesh* has already been released as part of SimNIBS and uses FSL *BET2* for skull segmentation and FreeSurfer 5.3.0

Table 1

Summary of the tested software configurations.

Segmentation	Postprocessing of segmentations
Skull segmentations (compared against CT; input: T1w only, or T1w&T2w)	
FSL <i>BET2</i>	ensure 1 mm thickness
SPM12 unified segmentation	MRF
SPM12 unified segmentation	MRF + morph
GM segmentations (compared against FreeSurfer; input: T1w only)	
SPM12 unified segmentation	MRF
SPM12&CAT12	–

MRF denotes the Markov random field procedure integrated in SPM12, “morph” refers to a number of morphological operations to clean up the binarized skull segmentations

for brain segmentation. The methods used by *mri2mesh* to convert the voxel segmentations into a volume mesh based on tetrahedral elements are described in detail in (Windhoff et al., 2013). Briefly, voxel segmentations of the inner and outer skull boundaries (and the scalp) are converted into surfaces based on triangle elements. These surfaces and the GM and WM surfaces reconstructed by FreeSurfer then undergo cleaning steps to ensure a good triangle quality and to avoid mutual intersections. Finally, volume meshes are constructed by filling the spaces between the surfaces by tetrahedra using Gmsh (Geuzaine and Remacle, 2009). The meshes contain around 3–4 million tetrahedra and distinguish between five tissue types, namely WM, GM, CSF, skull and scalp. Due to the chosen meshing approach, the tissue surfaces need to be nested and thus we ensure that there is at least one voxel layer of CSF between skull and gray matter as described in section *Segmentation of MR images*. This artificial restriction might not be exactly anatomically accurate, and is a limitation enforced solely by the current meshing approach.

Our new software pipeline, *headreco*, uses the voxel segmentations generated by SPM12 cleaned using morphological operations and, if available, the reconstruction of the middle of the GM sheet supplied by CAT12 to build individualized head models. It will be added to the next release of SimNIBS 2. Similar methods as described in (Windhoff et al., 2013) are used to create triangle surfaces from the voxel segmentations, which are then cleaned and finally used as input for the generation of tetrahedral volume meshes. As part of this process, the GM surface was constructed by moving the surface representing the middle of the GM sheet outwards as described above. The final head meshes produced by *headreco* extend the head meshes from *mri2mesh* by also including the vitreous bodies of the eyes and, in addition, modeling the paranasal sinuses, parts of the mouth and the throat as air cavities.

The numerical accuracy and stability of the FEM calculations relies on the geometrical quality of the tetrahedra (i.e., their “well-shapedness”). We therefore compared the tetrahedral quality of the head meshes generated with the new pipeline with those reported in (Windhoff et al., 2013) for *mri2mesh*, using three metrics, ρ , η , and γ which relate to the shape of the tetrahedra (Liu and Joe, 1994; Zhang et al., 2005)

$$\rho = C_1 \frac{\min_i e_i}{\max_i e_i}$$

$$\eta = C_2 \frac{V^{2/3}}{\sum_i e_i^2}$$

$$\gamma = C_3 \frac{V}{\max_i \{e_i\} \sum_i A_i}$$

where e_i is the length of the i th edge, V is the volume of a particular tetrahedron and A_i is the area of the i th face of a particular tetrahedron. Constants C_1 , C_2 and C_3 were chosen so that all of these metrics are bounded between 0 (worst quality) and 1 (best quality). For example, ρ is the ratio of the lengths of the shortest and longest edges and thus penalizes long, thin tetrahedra. The metric η assesses the ratio between the tetrahedral volume and the edge lengths, again favoring tetrahedra, which achieve a given volume V with on average shorter edges compared to tetrahedra requiring on average longer edges. Metric γ works similarly by relating the tetrahedral volume to the surface area. The metrics show different sensitivities to various possible deviations from an ideal symmetric pyramid shape, rendering an assessment of all three metrics useful. All metrics are calculated per tetrahedron, and then average statistics are extracted. We report the minimal values occurring per head mesh (averaged across the 10 subjects), the mean values, and the percentage of tetrahedra having a value lower than 0.1. The latter value of 0.1 was selected as lower threshold for a “good” tetrahedral quality based on our subjective experience with the stability of the FEM results.

In order to illustrate the relevance of using extended head models

that include the neck region when calculating the electric field distribution for TDCS, we exemplarily show the calculated fields from two tDCS montages; a standard montage for motor cortex stimulation with a return electrode on the contralateral supraorbital region and a montage with an extraencephalic return electrode. Both montages were simulated using a head model restricted to the upper part of the head (as generated by *mri2mesh*) and an extended head model (generated by *headreco*). In case of the restricted head model, the extraencephalic return electrode was positioned on the bottom surface to mimic the correct position “as good as possible”. We give two further examples to show how a misestimation of the skull thickness or a coarse segmentation of the gyral folding pattern can affect the electric field in the brain, both times based on the standard montage. All tissues, including WM, were modelled as isotropic conductors, and the assigned conductivity values were 0.126 S/m (WM), 0.275 S/m (GM), 1.79 S/m (CSF), 0.01 S/m (bone), 0.50 S/m (vitreous bodies) and 0.465 S/m (scalp) (Baumann et al., 1997; Saturnino et al., 2015). The electrodes were modelled as rubber pads ($40 \times 40 \text{ mm}^2$, 1 mm thickness, 0.1 S/m) embedded in saline-soaked sponges ($70 \times 50 \text{ mm}^2$, 5 mm thickness, 1 S/m), with the connectors placed centrally on the rubber pads. A current strength of 1 mA was simulated.

Results

MRI-based skull reconstructions

Comparison of the MRI- with the CT-based segmentations reveals better results for the segmentations based on combined T1w- and T2w-images versus those using only a T1w image (Fig. 1), consistently for FSL *BET2* and the *unified segmentation* routine of SPM12. Inclusion of a T2w image generally serves to improve the segmentations and stabilizes the results (i.e., decrease the variance across subjects). Importantly, outliers with very bad segmentations are mostly prevented. This is likely due to the better contrast between skull and CSF, which is mostly absent in the T1w images. Comparison between SPM12 and FSL *BET2*, using the reduced field-of-view (FoV) and a T1w image as input, shows that both perform comparably in terms of Dice score but that SPM12 obtains a lower Hausdorff distance, which is likely due to a more accurate segmentation of the boundary of the skull. When both T1w and T2w images are given as input, SPM12 performs better in both metrics compared to FSL *BET2*. Finally, BrainSuite consistently performed worse than the other methods irrespective of FoV.

Post-processing the SPM12-based voxel segmentations using simple morphological operations improves the results moderately, but consistently (bold vs. patterned boxes of same colors in Fig. 1). In addition, we note improved accuracy of the SPM12 segmentations in the upper part of the head (covering the skull) compared to the lower part (covering the

neck) as indicated by better results for the reduced vs. full FoVs.

The false positive and negative maps shown in Fig. 2A allow for a more fine-grained evaluation of the segmentation accuracy for the different parts of the skull and the vertebrae. Generally, most errors occur at the boundaries, particularly at the inner skull boundary if a T2w image is not used. While FSL *BET2* both over- and underestimates the inner skull boundary in this case, BrainSuite and SPM12 systematically underestimate the skull thickness (the results after applying the morphological cleanup operations are shown for the SPM12 segmentations). Using also a T2w image, FSL *BET2* slightly, but consistently, underestimates the skull thickness, while the SPM12-based results are quite accurate in the upper part of the skull. In general, air cavities such as the frontal and paranasal sinuses tend to be underestimated by all methods and be represented as bone (FSL *BET2* covers only the region of the frontal sinuses). FSL *BET2* and BrainSuite oversimplify the shape of the skull base. SPM12 recovers this part better in case both T1w and T2w images are given as input, but generally underestimates the vertebrae.

MRI-based gray matter reconstructions

Fig. 2B shows the map of false positive and negative rates across the subjects for the GM segmentations. Comparing the SPM12-based GM segmentations with those of FreeSurfer (upper row in Fig. 2B and first row in Table 2) reveals that SPM12 tends to miss CSF in thin and deep sulci, which are instead labeled as GM (visible as yellow false positives and higher average Hausdorff distance compared to CAT12 in Table 2). In contrast, CAT12 generates a better segmentation of these sulci, resulting in lower rates of false positives and negatives when compared to FreeSurfer (lower row in Fig. 2B and second row in Table 2). This effect is also clearly visible in pial surfaces reconstructed via the three different methods, as exemplarily demonstrated on a single subject in Fig. 2C.

Effect of regularization of bias field correction and of MR image parameters

While helpful for a robust segmentation of the CSF-bone border, using a T2w image in addition to the T1w image in the SPM12-based segmentation sometimes had detrimental effects on the segmentation of the spinal cord in several subjects. This was caused by an incorrect estimate of the bias field, which could be resolved by adjusting the regularization parameter that controls the flexibility of the estimated bias field. Fig. 3A shows an example of the effect of the regularization parameter on the estimated probability map of white matter (which is binarized to obtain the segmentation). Note the spinal cord which is missed when using too large (0.1) or too low (0.001) values of the regularization parameter.

In addition, the quality of the segmentation depends clearly on the chosen MRI sequence parameters. In Fig. 3B, this is exemplarily demonstrated for the SPM12-based skull segmentation using a T1w

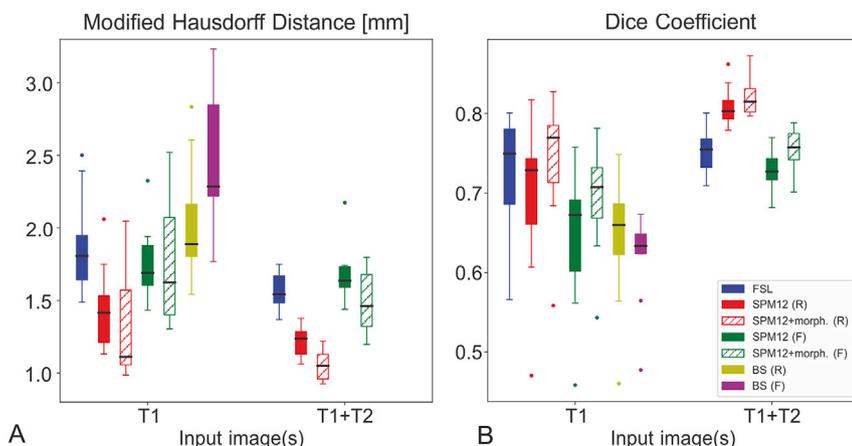


Fig. 1. Global accuracy measures for the skull segmentation. **A)** Modified Hausdorff distance. **B)** Dice coefficient. The black lines indicate the medians across the 10 subjects. The boxes indicate the 25 and 75 percentiles, and the error bars indicate the most extreme datum within 150% of this range. Data points outside the error bars are treated as outliers and are shown as small circles. Analyses using a reduced FoV restricted to the upper part of the head are marked by (R) in the legend. Analyses including the full FoV including the neck and vertebrae are marked by (F). Solid blue: FSL *BET2* – results assessed in the *reduced* FoV; Solid red: SPM12–*reduced* FoV; Striped red: SPM12 + morphological operations – *reduced* FoV; Solid green: SPM12–*full* FoV; Striped green: SPM12 + morphological operations – *full* FoV; Solid yellow: BrainSuite – *reduced* FoV; Solid magenta: BrainSuite – *full* FoV. To aid visualization in A, outliers at 4.25 mm and 4.42 mm for SPM12 (T1+T2, full FoV) and BrainSuite (full FoV), respectively, are not shown.

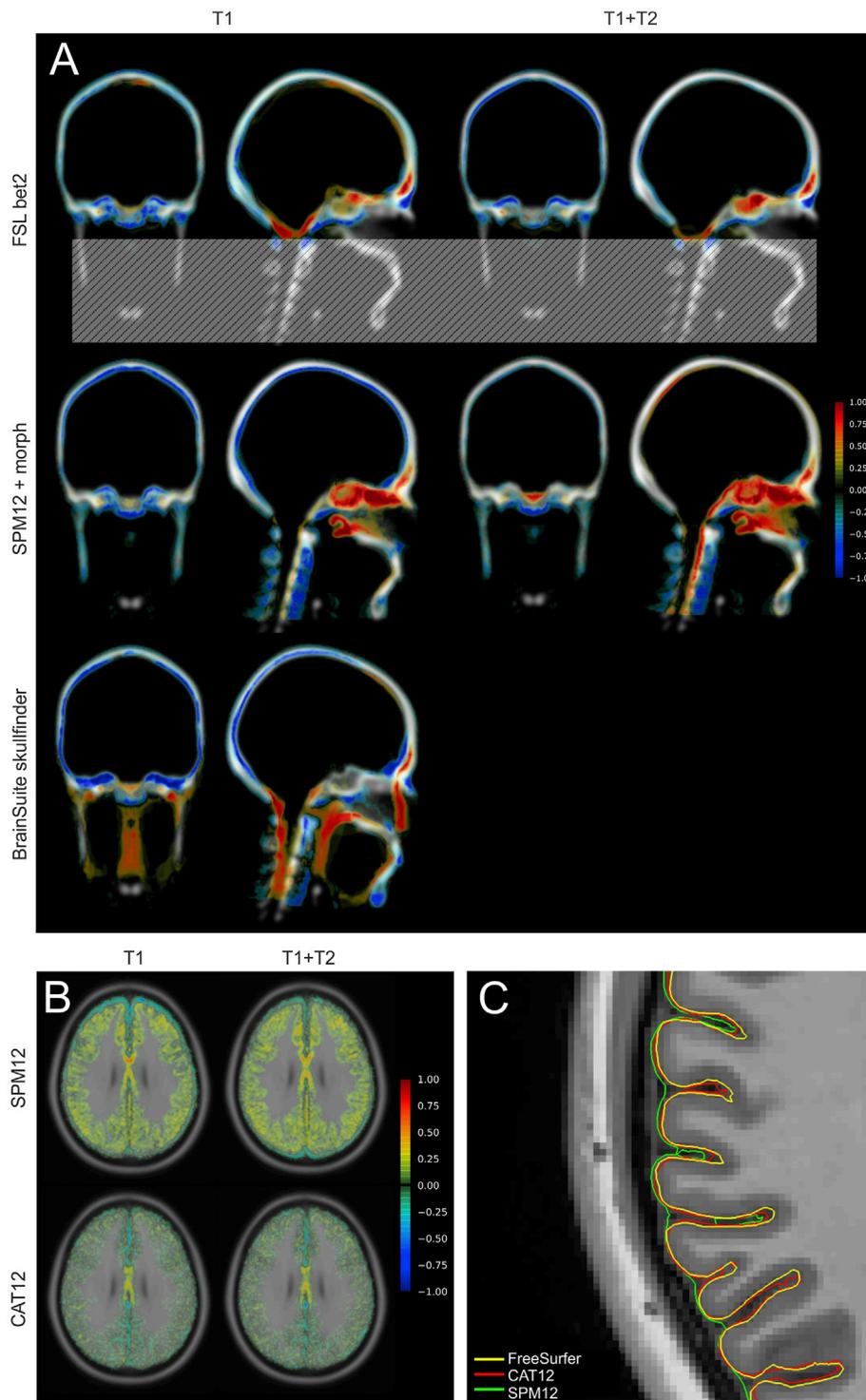


Fig. 2. A) Local accuracy of bone reconstruction assessed by false positive and false negative rates across subjects overlaid on the bone prior of the extended atlas. Red denotes false positives with a ratio of one corresponding to a false positive labeling of non-bone tissue as bone in all subjects. Blue denotes false negatives with a one indicating a labeling of bone as non-bone in all subjects. Shown are exemplary coronal and sagittal slices. The gray box in the upper row indicates the neck region that is ignored by FSL *BET2*. B) Local comparison of the gray matter reconstruction of SPM12 (upper row) and CAT12 (lower row) with FreeSurfer assessed by false positive and false negative rates across subjects overlaid on the MNI T1 template. Red and yellow colors indicate false positives, blue denotes false negatives. C) Representative gray matter surface reconstruction on a single subject based on the GM segmentations of SPM12 (green) and CAT12 (red), and the pial surface reconstructed by FreeSurfer (yellow).

image from an openly available dataset as input. This T1w image was acquired without fat suppression, resulting in strong signal from subdural fat belonging to the scalp and the neighboring spongy parts of the bone. As a result, large parts of the skull were incorrectly segmented as scalp and the skull was strongly underestimated. Interestingly, facial bones were modelled in the defaced region based on the skull priors.

Mesh quality

The quality of the head meshes created by *mri2mesh* (which employs

FSL *BET2* and FreeSurfer for segmentation) and *headreco* (which uses SPM12 and CAT12) is comparable (Tables 3–5). This is expected as similar methods are used for generation and optimization of the triangle surfaces and for the subsequent volume meshing. A more accurate modeling of fine sulcal structures slightly increases the number of low-quality tetrahedra, as revealed by a comparison of the results for *mri2mesh* and *headreco* with CAT12 versus *headreco* using only SPM12 in Tables 3–5. However, a good mesh quality is reached in all cases. The number of low-quality tetrahedra is lower than originally reported in Table 2 in (Windhoff et al., 2013). This is most likely due to improvements in the volume meshing routines in the newer version of Gmsh

Table 2

Similarity of the gray matter segmentation to the FreeSurfer result for default SPM12 and when using CAT12. We report the mean and standard deviation of the modified Hausdorff distance for the 10 subjects.

Method/Score	Modified Hausdorff distance in mm (mean/std)
SPM12	
T1	0.7168/0.057
T1+T2	0.7242/0.061
SPM12&CAT12	
T1	0.3740/0.047
T1+T2	0.3604/0.029

employed here.

Electric field simulations for TDCS

Exemplary volume conductor models, as generated by the different methods, can be seen in Fig. 4A&B. Compared to *mri2mesh*, *headreco* offers an extended spatial coverage including a (coarse) modeling of the vertebrae, a finer modeling of the skull base and of the air cavities, and the modeling of the vitreous bodies of the eyes as a separate class. Interestingly, these differences only seem to play a minor role for a standard TDCS montage targeting the motor cortex, which places both sponge electrodes on the upper part of the head (electric field images in Fig. 4A&B). In contrast, very clear differences are seen when modeling an extracranial return electrode (Fig. 4C&D). While the field distribution underneath the orange stimulation electrode is similar for the two head models (image not shown), very clear differences occur in the

Table 3

Mesh quality as quantified by η . Summary statistics include minimum quality tetrahedron, percentage of tetrahedra below $\eta = 0.1$, and mean quality of tetrahedra. Values reflect mean (standard deviation) across all subjects.

	η		
	min (std)	% < 0.1 (std)	mean (std)
mri2mesh			
T1	0.034 (0.023)	7.5e-04 (3.8e-04)	0.768 (0.004)
T1+T2	0.033 (0.020)	5.5e-04 (2.6e-04)	0.763 (0.005)
headreco			
T1	0.086 (0.035)	2.3e-05 (2.5e-05)	0.784 (0.005)
T1+T2	0.082 (0.047)	1.2e-04 (2.4e-04)	0.785 (0.005)
T1 (CAT)	0.025 (0.014)	7.0e-04 (4.0e-04)	0.780 (0.004)
T1+T2 (CAT)	0.013 (0.007)	1.1e-03 (4.8e-04)	0.780 (0.005)

cerebellum and the temporal lobe of the right hemisphere.

The effects of a gross underestimation of the average skull thickness on the electric field in the brain are exemplarily shown in Fig. 5A. The inaccurate skull segmentation (similar in quality to the one shown in Fig. 3B) results in less blurring of the electric field on the brain surface and the occurrence of localized field peaks. Interestingly, also a coarse segmentation of the gyral folding pattern clearly affects the field estimates in the brain (Fig. 5B). Not modeling of the CSF in thin and deep sulci increases the field strength in large parts of gray matter, presumably due to reduced shunting of currents through well-conductive CSF pathways.

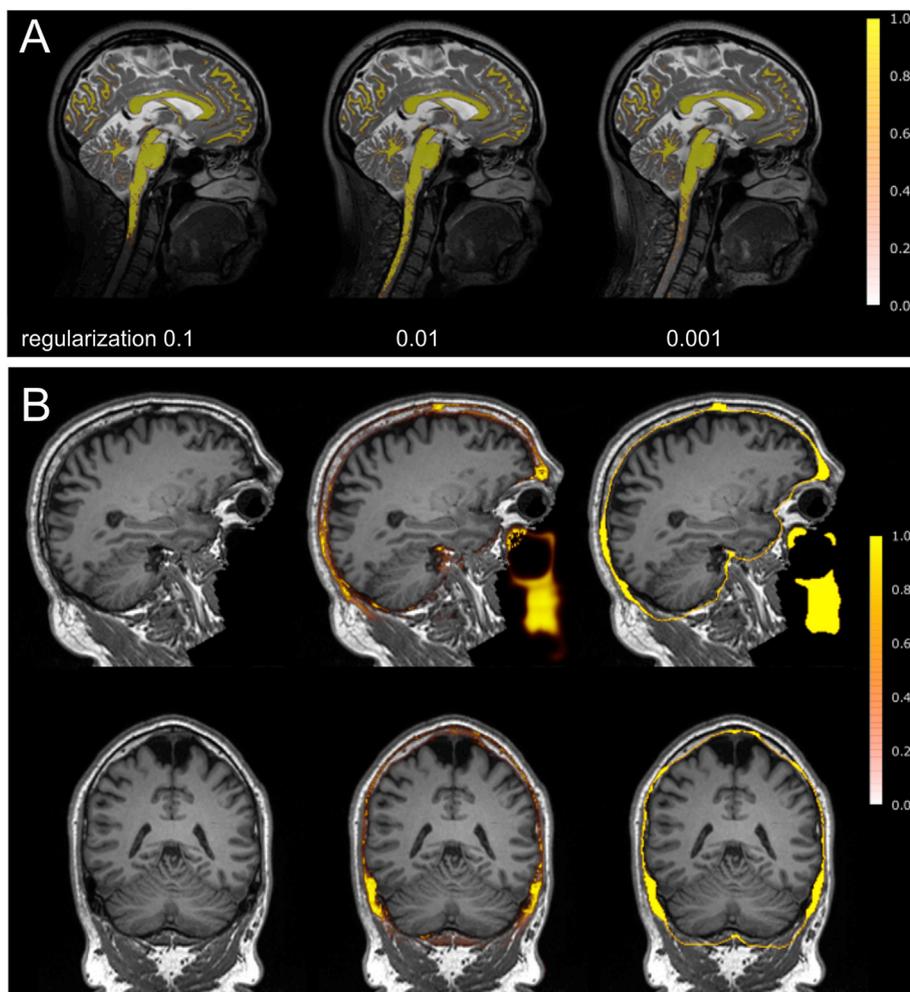


Fig. 3. A) Probability map of white matter from SPM12 for different strengths of the regularization of the bias field correction for the T2 image. The regularization strength of 0.001 used in the right column is the default value in SPM12. The results have been overlaid on the bias corrected T2 image. B) Exemplary demonstration of the impact of the MRI parameter settings on the results of the skull segmentations. A T1w image from an open-source dataset was used as input. Please note that the images of this dataset were defaced to protect the privacy of the subjects. The left column shows sagittal and coronal slices through the T1w image; the middle column shows the posterior probability given by the SPM12 *unified segmentation* routine; the right column shows the final skull segmentation after the morphological operations.

Table 4

Mesh quality as quantified by γ . Summary statistics include minimum quality tetrahedron, percentage of tetrahedra below $\gamma = 0.1$, and mean quality of tetrahedra. Values reflect mean (standard deviation) across all subjects.

	γ		
	min (std)	% < 0.1 (std)	mean (std)
mri2mesh			
T1	0.031 (0.023)	1.0e-03 (4.5e-04)	0.630 (0.004)
T1+T2	0.032 (0.020)	8.0e-04 (4.0e-04)	0.625 (0.005)
headreco			
T1	0.067 (0.034)	4.2e-05 (4.2e-05)	0.646 (0.005)
T1+T2	0.068 (0.039)	1.6e-04 (2.9e-04)	0.647 (0.005)
T1 (CAT)	0.023 (0.014)	1.0e-03 (4.8e-04)	0.642 (0.004)
T1+T2 (CAT)	0.011 (0.007)	1.4e-03 (6.0e-04)	0.643 (0.005)

Table 5

Mesh quality as quantified by ρ . Summary statistics include minimum quality tetrahedron, percentage of tetrahedra below $\rho = 0.1$, and mean quality of tetrahedra. Values reflect mean (standard deviation) across all subjects.

	ρ		
	min (std)	% < 0.1 (std)	mean (std)
mri2mesh			
T1	0.043 (0.011)	1.4e-03 (6.8e-04)	0.548 (0.005)
T1+T2	0.057 (0.010)	1.1e-03 (4.2e-04)	0.543 (0.005)
headreco			
T1	0.104 (0.015)	3.2e-05 (4.5e-05)	0.565 (0.005)
T1+T2	0.102 (0.022)	3.9e-05 (1.0e-04)	0.566 (0.005)
T1 (CAT)	0.066 (0.019)	4.4e-04 (3.1e-04)	0.561 (0.004)
T1+T2 (CAT)	0.057 (0.016)	6.8e-04 (3.4e-04)	0.562 (0.005)

Discussion

We have validated the accuracy of skull segmentations obtained by three methods (FSL *BET2*, BrainSuite *skullfinder*, and the SPM12 *unified segmentation* routine) by comparing against CT-based skull segmentations in ten subjects. Both FSL and SPM12 give reasonable results for the upper part of the skull, in particular when both a T1w and T2w image are used as input. The results of BrainSuite are less accurate. For FSL and SPM12, including a T2w image strongly reduces the variability of the segmentation accuracy across subjects, with the SPM12-based being moderately more accurate than FSL *BET2* in that case. Post-processing of the SPM12-based segmentation using simple morphological operations increased the accuracy slightly. Visual inspection of the local accuracy maps demonstrated the remaining specific weaknesses of all methods – even when using both T1w and T2w images.

BrainSuite *skullfinder* (which uses only a T1w image) was constructed primarily with the purpose of segmenting the upper part of the skull (Dogdas et al., 2005) explaining its relatively low accuracy in segmenting the jaw and neck areas. We see, however, that it generally underestimates the inner skull border, i.e., the resulting skull segmentation is thinner compared to the CT-based ground truth. The underlying reason seems to be the use of too large brain masks that are used as starting points of the inner skull estimates. Consequently, the manual optimization of the threshold parameters (as we did in this study) helps to improve the outer but not inner skull estimate. The reconstructed vertebrae seem to be spatially misplaced, resulting in a mixture of false positives and false negatives in the lower part of the segmentations.

FSL *BET2* relatively coarsely approximates the shape of the skull base, resulting in both false positives and negatives in this region. This approximation is caused by *betsurf*, which assumes a particular pattern of intensity variations along vectors from the brain to the outer scalp and identifies the skull boundaries as minima along this intensity profile. This strategy works reasonably well for the upper part of the skull, but not for the lower parts with more complicated bone structures.

In contrast to the results of FSL *BET2*, the SPM12-based segmentation

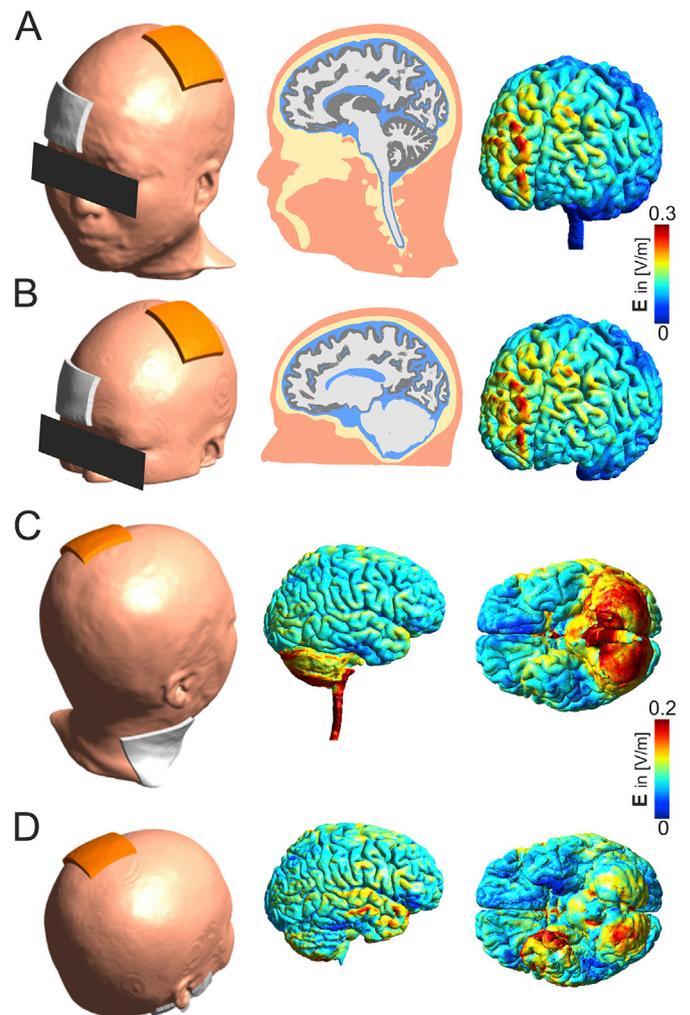


Fig. 4. A) Extended volume conductor model including the neck of an exemplary subject, created by the *headreco* script using SPM12 and CAT12. A “standard” montage for left motor cortex stimulation is simulated, with one electrode positioned above the left motor cortex and the return electrode placed above the right supraorbital region (the black bars were added to camouflage the facial features). The middle column shows a sagittal slice through the volume mesh. The right column shows the induced field over the target area. B) Volume conductor model with limited coverage of the same subject, created by *mri2mesh* using FSL *BET2* and FreeSurfer. The injected electric field in GM is very similar to the one calculated with the extended head model. C) Norm of the electric field for a montage targeting the left motor cortex, with one electrode placed above the motor cortex and an extraencephalic return electrode on the right side of the neck. The extended volume conductor model was used. D) Norm of the electric field for the same montage as in C, but adapted for the volume conductor model covering only the upper part of the head. The position of the extraencephalic return electrode was mimicked as good as possible by placing it on the right bottom surface. The estimated field clearly deviates from that in C. While the field strength in the cerebellum is underestimated, the field in the right temporal lobe is overestimated. The rightmost view in subfigures C and D is from below the brain.

also covers facial bone structures and the vertebrae. Remaining inaccuracies compared to the CT-based segmentations manifest themselves mainly as false positives around the air cavities and false negatives around the vertebrae. Both air and compact bone do not give signal in conventional MR images, so that the *unified segmentation* routine of SPM12 has to rely on the spatial tissue priors in order to distinguish both. The employed spatial priors did not enforce the shapes of the air cavities strongly enough, resulting in inaccurate segmentations. The vertebrae were consistently underestimated, which was likely a result of

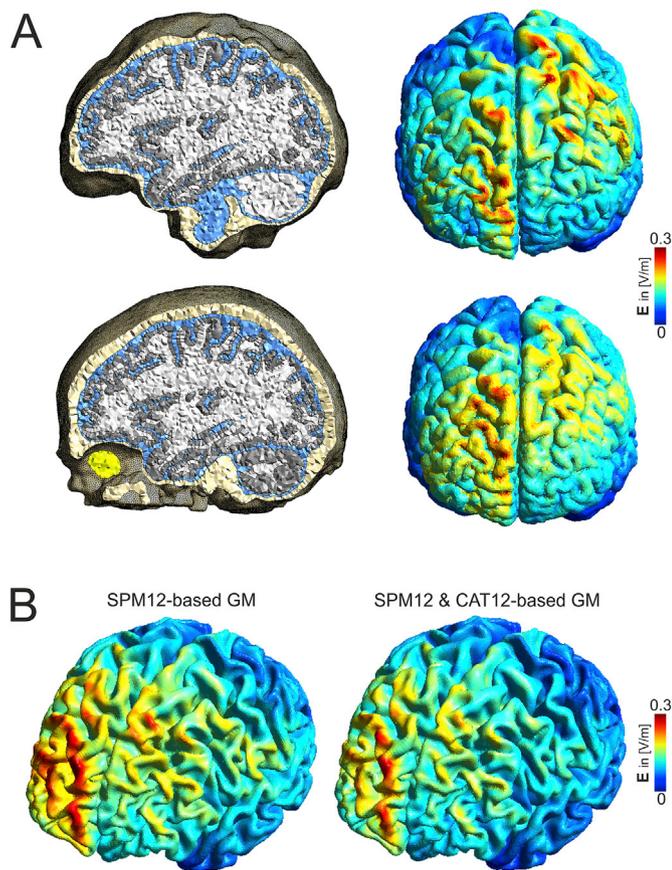


Fig. 5. A) Demonstration of the effect of inaccurate (top row) and accurate (bottom row) skull segmentation on the simulated electric fields. The inaccurate skull model is the result of an automatic segmentation of a T1w image using FSL BET2. The accurate model used SPM12 with standard tissue probability maps for the initial segmentation of the same image, followed by manual corrections. The norm of the electric field is shown on the FreeSurfer cortical (pial) surface, where the electric field strength was read out on the mid layer between the white matter and pial surfaces. B) Demonstration of the effect of SPM12 (left) and SPM12&CAT12 (right) gray matter segmentations on the simulated electric fields. For both models, the electric field strength was read out on the mid layer of the gray matter sheet, as reconstructed by CAT12. In both subfigures, a “standard” montage for left motor cortex stimulation was used, with one electrode positioned above the left motor cortex and the return electrode placed above the right supraorbital region.

insufficient contrast between different tissue types in the MR images together with smooth spatial priors missing the details of the vertebrae structures. For example, in the T2w images, the spinal cord, the spongy bone of the vertebrae, and subdural fat show similar intensities and occur in close proximity, so that recovering the vertebrae structures would necessarily require strong prior information. Furthermore, the spatial priors do not account for abnormalities, such as tumors, stroke or surgical intervention, as occurring in patients. Large deviations from normal anatomy can lead to significant degradation of the resulting head models (Biro et al., 2014; Minjoli et al., 2017) and need to be accounted for explicitly to ensure good quality of the head models in these cases. This can be done either by manual editing (Datta et al., 2011; Minjoli et al., 2017) or by tailored automated segmentation tools (Kamnitsas et al., 2017; Menze et al., 2015).

Commercial software packages were excluded from this study. We are aware of two tools: BESA MRI (<http://www.besa.de/products/besa-mri/besa-mri-overview/>) and EGI's GeoSource 3 (<https://www.egi.com/research-division/electrical-source-imaging/geosource>), which both provide automated head model generation, including skull segmentation, from MR scans. The skull segmentation approach in BESA MRI combines

tissue probability atlases with a spatial Markov random field model (Lanfer et al., 2014), which on a general level is closely related to what SPM12 does, whereas GeoSource 3 uses landmark-based nonlinear registration to warp a skull template to the target MR scan to produce a skull segmentation (Li et al., 2016). However, quantitative comparison of these methods was not possible, as we do not have access to either of the software packages.

While our results indicate that relatively good skull segmentations can be obtained from automatic procedures, the achieved quality depended on the MRI sequence parameters and (for SPM12 and BrainSuite) also the input parameters and thresholds for the segmentation routines. FSL BET2 relies on clearly detectable intensity variations between brain, compact and spongy bone and scalp along the sampled vectors, which is usually well achieved in T2w, but not in T1w images. In the latter, CSF is dark, rendering it similar to compact bone. In addition, the fat-water shift due to a rather low readout bandwidth can camouflage the boundaries between scalp, spongy bone and brain. During prior testing, this led us to combine a T1w image with selective water excitation (in which compact and, to a large extent, spongy bone gives homogeneously low intensity) with a non-fat suppressed T2w image. While we did not separately optimize the MR image parameters for use with SPM12 or BrainSuite, our results indicate that they are also a reasonable choice for these methods. Testing BrainSuite with a T1w image without fat suppression, we observed a similar pattern as in our previous tests, namely that the outer skull border seemed to be well delineated whereas the inner skull border was recovered less successfully. Testing SPM12 on this data gave far worse results than obtained for our chosen MRI parameters, even when compared to the results obtained without an additional T2w image. To confirm our prior experience, we also tested FSL BET2 on this data, which resulted in a segmentation where large parts of the spongy bone were mislabeled as skin. This highlights the importance of testing the segmentation performance on some pilot data sets prior to the start of a study. We do not claim that our chosen parameters are the only possible or optimal choice, but hope that they are a helpful guidance for the reader. Some of the problems related to bone modeling could be resolved by using a combination of CT and MR scans (Eichelbaum et al., 2014). The good bone contrast from CT allows to distinguish small details, such as small openings due to blood vessels, and would likely help with modeling the vertebrae. However, acquiring a CT scan in addition to the MRI requires access to additional equipment, and results in an increased workload and exposure to radiation. Moving to ultra high-field MRI (7T) could facilitate the modeling of small blood vessels or openings in the skull (Fiederer et al., 2016), but access to high-field scanners is often limited in practice.

We tested the impact of changes of the image downsampling factor (the default setting in SPM12 is downsampling the resolution of the input image by a factor of three) and the regularization parameter for the bias field correction in SPM12 on the segmentation accuracy. While the segmentations were robust to changes in the downsampling factor, optimization of the regularization parameter helped to ensure better segmentations of the spinal cord. This was only necessary for the T2w images as they suffered more from intensity inhomogeneities in the lower part than the T1w image. The high flip angles employed in T2w images likely render this image type more susceptible to spatially varying flip angles at the lower boundary of the RF excitation coil. Therefore, the incorrectly estimated bias field when using the standard regularization parameter setting caused the intensity profiles of the tissues in the bottom part (e.g., the spinal cord) to be different from that in the upper part (e.g., cortical white matter), so that spinal cord voxels were assigned a low probability of being WM, and consequently were classified as scalp. However, as the same MRI hardware and image parameters were used for all subjects, the selection of optimized regularization parameter had to be done only once by testing a few subjects.

Our results are broadly in line with those presented in (Perdue and Diamond, 2014) for the upper part of the head. In that study, a fair segmentation performance of FSL and SPM8 for the skull was

documented when using a T1w image as input. We further replicated their finding that BrainSuite offers a lower segmentation performance, which is in contrast to the original results for BrainSuite presented in (Dogdas et al., 2005). We also demonstrated the relevance of using a combination of T1w and T2w images to improve the segmentation accuracy. In addition, we based our validations on individual CT-based skull segmentations as ground truth, tested a new version of SPM and included spatially extended tissue priors that covered the neck.

Whereas automatic skull segmentation was the focus of this study, we further demonstrated that the CAT12 toolbox in SPM12 achieves GM segmentations similar to the ones produced by FreeSurfer 5.3.0, thus confirming the results of Dahnke et al., 2013. As a result, the segmentations based on SPM12 combined with CAT12 enabled automatic construction of volume conductor models that had good anatomical accuracy in the upper part of the head (in particular when a T2w image was used), and included the coarse features of the neck and vertebrae. Specifically, accurately modeling the skull and the CSF-GM and GM-WM boundaries have been suggested to be important for the realistic estimation of volume conduction effects in the head (Vorwerk et al., 2014). In addition, for EEG and MEG, the reconstructed cortical surface is often considered important for determining the orientation of the neural dipole sources and for spanning the source space. In practice, including Figs. 4 and 5, we define the cortical surface to be the mid layer between the white matter and pial surfaces. The placement of this surface, as long as it is not right at the white matter or pial boundary, should not have a large effect on the resulting field strength as the electric field is discontinuous only right at the tissue boundaries. We exemplarily demonstrated the importance of a spatially extended head model for electric field estimations of tDCS montages that involve extraencephalic electrodes. Similarly, for EEG source modeling, cutting the model right below the cerebellum (and artificially closing the bone compartment) has been shown to result in non-negligible errors in particular for deep sources (Indahlastari et al., 2016; Lanfer et al., 2012). Thus, extended volume conductor models including the neck should help to improve the accuracy of the calculated fields and sources.

Despite these advantages, the SPM12-based segmentations still have clear limitations. Large parts of the air cavities are wrongly segmented as bone. However, as both bone and air have a low conductivity, this might not have a strong impact on the accuracy of the resulting field calculations unless the erroneously labeled voxels are very close to the brain area of interest (Indahlastari et al., 2016; Lanfer et al., 2012). While modeling of the spinal cord and the vertebrae likely allows for a more accurate estimation of the current pathways in the cerebellum that is positioned mostly superior to these structures, underestimation of the vertebrae renders the calculated fields in the spinal cord itself inaccurate. None of the segmentation methods tested here currently distinguish between spongy and compact skull. Given their different conductivities, distinguishing these tissues might be beneficial to improve the accuracy of the field estimates (Dannhauer et al., 2011), though using an optimized bone conductivity may also suffice (Cho et al., 2015; Dannhauer et al., 2011; Huang et al., 2017; Vorwerk et al., 2014). Regarding the SPM12-based segmentations, the explicit modeling of spongy bone, including a separate spatial prior, might also help to improve the robustness of the skull segmentation by reducing the extent to which spongy bone is mislabeled as, for example, scalp. Finally, scalp is currently a lumped representation of skin, subdural fat and muscle tissue, rendering the reliable assignment of a single conductivity value to this tissue class difficult. In particular, this is the case for the segmented scalp

in the neck region, in which fat and muscle tissue are spatially intermingled. In contrast, the scalp in the upper part of the head mostly consists of a regularly layered structure of skin and subdural fat with only little muscle tissue. This again suggests that the field estimates in the upper part of the head model will likely be more accurate than in the neck region when a spatially extended head model is used. However, estimations of the size of the error caused by this simplification would require a systematic sensitivity analysis. It is worth noting in this respect that the anatomical accuracy of volume conductor models is only one, albeit important, aspect to obtain accurate forward models of the electric field distributions. In addition, the applied conductivity values are average values taken from a rather sparse literature, in which the reported values often vary considerably between studies (e.g., Gabriel et al., 1996; Geddes and Baker, 1967). This limitation adds to the uncertainty of the field estimates, and the inter-subject variabilities in the tissue conductivities which have been observed (Dabek et al., 2016; Huang et al., 2017) suggest the need for subject-specific calibrations, either through dedicated measurements (Dabek et al., 2016) or optimization procedures (Aydin et al., 2014; Huang et al., 2017).

Conclusion

In summary, our study demonstrates the current state-of-the-art of automatic skull segmentation from MR images, including the identification of remaining shortcomings, and introduces a novel, easily accessible and validated open-source tool for the automatic creation of volume meshes of the complete head. We have compared three methods (FSL *BET2*, BrainSuite *skullfinder*, and the *unified segmentation* routine of SPM12 with extended spatial tissue priors) to automatically segment the human skull. We conclude that FSL and SPM12 achieve reasonable results for the upper part of the skull when a combination of T1w and T2w images is used as input and suited MRI sequence parameters are chosen. In that case, the skull segmentation of SPM12 seems to be moderately better than that of FSL *BET2*, in particular when combined with a post-processing step based on simple morphological operations to clean up local defects. In contrast to FSL *BET2*, the SPM12-based segmentation also covers the lower part of the head including the neck and recovers details of the facial bones. For EEG and TDCS, the skull has a strong influence on the electric field distribution occurring in the brain and on the scalp surface due to its low conductivity. Employing an extended volume conductor model that builds upon the SPM12-based segmentation can thus help to improve the accuracy of the forward models. However, the reconstructed vertebrae suffer from a low anatomical accuracy so that the volume conductor models are still too coarse to allow for reliable estimation of the electric field in the spinal cord. The validated tools (FSL and SPM12) are integrated into our open-source pipeline SimNIBS (www.simnibs.org) that allows for the automatic generation of high-quality tetrahedral head meshes and their subsequent use in field calculations for brain stimulation. The released version supports head meshes using the FSL-based skull segmentation, while the SPM12-based segmentation will be added in the upcoming version.

Acknowledgements

This study was supported by the Lundbeck foundation (grant R118-A11308 to AT and grant R59-A5399 - PI Hartwig Siebner), the Novonordisk foundation (grant no. NNF14OC0011413) and a PhD stipend of the Sino-Danish Center to JDN.

APPENDIX

Table A1

List of the main post-processing operations applied to the different tissues. *Dilation* flips background voxels to foreground voxels if any of the neighboring voxels, defined by the kernel, is a foreground voxel. *Erosion* removes foreground voxels if any of the neighboring voxels, defined by the kernel, is background voxel. *Closing* is defined as a dilation followed by an erosion (Gonzalez and Woods, 2002).

Tissue	Operation	Kernel	Explanation
Bone	Closing	Six closest neighbors (6-connectivity)	Remove holes from the bone structure by closing.
Skin	Erosion	Six closest neighbors (6-connectivity)	Erode to remove small erroneously segmented skin areas.
Skin	Keep largest component	Six closest neighbors (6-connectivity)	Find the largest connected component, and discard the rest, which are considered noise.
Skin	Dilation	Six closest neighbors (6-connectivity)	Dilate the largest component to account for the initial erosion.
All tissues	Iterative Gaussian smoothing ($\sigma = 1$) and assignment of voxels.	N/A	After the post-processing steps, some of the voxels might be unassigned. Iteratively smooth the binary masks, and assign unassigned voxels.

References

- Ashburner, J., Friston, K.J., 2005. Unified segmentation. *Neuroimage* 26, 839–851. <https://doi.org/10.1016/j.neuroimage.2005.02.018>.
- Aydin, U., Vorwerk, J., Küpper, P., Heers, M., Kugel, H., Galka, A., Hamid, L., Wellmer, J., Kellinghaus, C., Rampp, S., Wolters, C.H., 2014. Combining EEG and MEG for the reconstruction of epileptic activity using a calibrated realistic volume conductor model. *PLoS One* 9, e93154. <https://doi.org/10.1371/journal.pone.0093154>.
- Baumann, S.B., Wozny, D.R., Kelly, S.K., Meno, F.M., 1997. The electrical conductivity of human cerebrospinal fluid at body temperature. *IEEE Trans. Biomed. Eng.* 44, 220–223. <https://doi.org/10.1109/10.554770>.
- Biront, G., Spinelli, L., Vulliémoz, S., Mégevand, P., Brunet, D., Seeck, M., Michel, C.M., 2014. Head model and electrical source imaging: a study of 38 epileptic patients. *NeuroImage Clin.* 5, 77–83. <https://doi.org/10.1016/j.nicl.2014.06.005>.
- Chan, T.F., Vese, L.A., 2001. Active contours without edges. *IEEE Trans. Image Process.* 10, 266–277. <https://doi.org/10.1109/83.902291>.
- Cho, J.H., Vorwerk, J., Wolters, C.H., Knösche, T.R., 2015. Influence of the head model on EEG and MEG source connectivity analyses. *Neuroimage* 110, 60–77. <https://doi.org/10.1016/j.neuroimage.2015.01.043>.
- Dabek, J., Kalogianni, K., Rotgans, E., van der Helm, F.C.T., Kwakkel, G., van Wegen, E.E.H., Daffertshofer, A., de Munck, J.C., 2016. Determination of head conductivity frequency response in vivo with optimized EIT-EEG. *Neuroimage* 127, 484–495. <https://doi.org/10.1016/j.neuroimage.2015.11.023>.
- Dahnke, R., Yotter, R.A., Gaser, C., 2013. Cortical thickness and central surface estimation. *Neuroimage* 65, 336–348. <https://doi.org/10.1016/j.neuroimage.2012.09.050>.
- Dale, A.M., Fischl, B., Sereno, M.I., 1999. Cortical surface-based analysis: I. Segmentation and surface reconstruction. *Neuroimage* 9, 179–194. <https://doi.org/10.1006/nimg.1998.0395>.
- Dannhauer, M., Lanfer, B., Wolters, C.H., Knösche, T.R., 2011. Modeling of the human skull in EEG source analysis. *Hum. Brain Mapp.* <https://doi.org/10.1002/hbm.21114>.
- Datta, A., Baker, J.M., Bikson, M., Fridriksson, J., 2011. Individualized model predicts brain current flow during transcranial direct-current stimulation treatment in responsive stroke patient. *Brain Stimul.* 4, 169–174. <https://doi.org/10.1016/j.brs.2010.11.001>.
- Dogdas, B., Shattuck, D.W., Leahy, R.M., 2005. Segmentation of skull and scalp in 3-D human MRI using mathematical morphology. *Hum. Brain Mapp.* 26, 273–285. <https://doi.org/10.1002/hbm.20159>.
- Dubuisson, M.-P., Jain, A.K., 1994. A modified Hausdorff distance for object matching. *Proc. 12th Int. Conf. Pattern Recognit.* 1, 566–568. <https://doi.org/10.1109/ICPR.1994.576361>.
- Eggert, L.D., Sommer, J., Jansen, A., Kircher, T., Konrad, C., 2012. Accuracy and reliability of automated gray matter segmentation pathways on real and simulated structural magnetic resonance images of the human brain. *PLoS One* 7. <https://doi.org/10.1371/journal.pone.0045081>.
- Eichelbaum, S., Dannhauer, M., Hlawitschka, M., Brooks, D., Knösche, T.R., Scheuermann, G., 2014. Visualizing simulated electrical fields from electroencephalography and transcranial electric brain stimulation: a comparative evaluation. *Neuroimage* 101, 513–530. <https://doi.org/10.1016/j.neuroimage.2014.04.085>.
- Fiederer, L.D.J., Vorwerk, J., Lucka, F., Dannhauer, M., Yang, S., Dümpelmann, M., Schulze-Bonhage, A., Aertsen, A., Speck, O., Wolters, C.H., Ball, T., 2016. The role of blood vessels in high-resolution volume conductor head modeling of EEG. *Neuroimage* 128, 193–208. <https://doi.org/10.1016/j.neuroimage.2015.12.041>.
- Fischl, B., Sereno, M.I., Dale, A.M., 1999. Cortical surface-based analysis: II: inflation, flattening, and a surface-based coordinate system. *Neuroimage* 9, 195–207. <https://doi.org/10.1006/nimg.1998.0396>.
- Gabriel, C., Gabriel, C., Gabriel, S., Gabriel, S., Corthout, E., Corthout, E., 1996. The dielectric properties of biological tissues: I. Literature survey. *Phys. Med. Biol.* 41, 2231–2249. <https://doi.org/10.1088/0031-9155/41/11/001>.
- Geddes, L.A., Baker, L.E., 1967. The specific resistance of biological material - a compendium of data for the biomedical engineer and physiologist. *Med. Biol. Eng.* 5, 271–293.
- Geuzaine, C., Remacle, J.-F., 2009. Gmsh: a three-dimensional finite element mesh generator with built-in pre- and post-processing facilities. *Int. J. Numer. Methods Eng.* 79, 1309–1331.
- Gonzalez, R., Woods, R., 2002. *Digital Image Processing*. Prentice Hall. [https://doi.org/10.1016/0734-189X\(90\)90171-Q](https://doi.org/10.1016/0734-189X(90)90171-Q).
- Han, X., Jovicich, J., Salat, D., van der Kouwe, A., Quinn, B., Czanner, S., Busa, E., Pacheco, J., Albert, M., Killiany, R., Maguire, P., Rosas, D., Makris, N., Dale, A., Dickerson, B., Fischl, B., 2006. Reliability of MRI-derived measurements of human cerebral cortical thickness: the effects of field strength, scanner upgrade and manufacturer. *Neuroimage* 32, 180–194. <https://doi.org/10.1016/j.neuroimage.2006.02.051>.
- Huang, Y., Dmochowski, J.P., Su, Y., Datta, A., Rorden, C., Parra, L.C., 2013. Automated MRI segmentation for individualized modeling of current flow in the human head. *J. Neural Eng.* 10. <https://doi.org/10.1088/1741-2560/10/6/066004>, 66004.
- Huang, Y., Liu, A.A., Lafon, B., Friedman, D., Dayan, M., Wang, X., Bikson, M., Doyle, W.K., Devinsky, O., Parra, L.C., 2017. Measurements and models of electric fields in the in vivo human brain during transcranial electric stimulation. *Elife* 6. <https://doi.org/10.7554/eLife.18834>.
- Indahlastari, A., Chauhan, M., Schwartz, B., Sadleir, R.J., 2016. Changing head model extent affects finite element predictions of transcranial direct current stimulation distributions. *J. Neural Eng.* 13. <https://doi.org/10.1088/1741-2560/13/6/066006>, 66006.
- Jenkinson, M., Bannister, P., Brady, M., Smith, S., 2002. Improved optimisation for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* 17, 825–841. [https://doi.org/10.1016/S1053-8119\(02\)91132-8](https://doi.org/10.1016/S1053-8119(02)91132-8).
- Jenkinson, M., Smith, S., 2001. A global optimisation method for robust affine registration of brain images. *Med. Image Anal.* 5, 143–156. [https://doi.org/10.1016/S1361-8415\(01\)00036-6](https://doi.org/10.1016/S1361-8415(01)00036-6).
- Kamnitsas, K., Ledig, C., Newcombe, V.F.J., Simpson, J.P., Kane, A.D., Menon, D.K., Rueckert, D., Glocker, B., 2017. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med. Image Anal.* 36, 61–78. <https://doi.org/10.1016/j.media.2016.10.004>.
- King, K.F., 2004. Operator parameter selection. In: BERNSTEIN, M.A., KING, K.F., ZHOU, X.J. (Eds.), *Handbook of MRI Pulse Sequences*. Elsevier Academic Press, Amsterdam, pp. 374–377.
- Klein, S., Staring, M., Murphy, K., Viergever, M.A., Pluim, J.P.W., 2010. Elastix: a toolbox for intensity-based medical image registration. *IEEE Trans. Med. Imaging* 29, 196–205. <https://doi.org/10.1109/TMI.2009.2035616>.
- Lanfer, B., Paul-Jordanov, I., Wolters, C., 2014. Individual FEM pipeline for EEG source analysis requiring minimal user intervention. In: *Proceedings of the Annual Meeting of the Organization for Human Brain Mapping*. Hamburg, Germany, p. 1657.
- Lanfer, B., Scherg, M., Dannhauer, M., Knösche, T.R., Burger, M., Wolters, C.H., 2012. Influences of skull segmentation inaccuracies on EEG source analysis. *Neuroimage* 62, 418–431. <https://doi.org/10.1016/j.neuroimage.2012.05.006>.
- Lewiner, T., Lopes, H., Vieira, A.W., Tavares, G., 2003. Efficient implementation of marching cubes' cases with topological guarantees. *J. Graph. Tools* 8, 1–15. <https://doi.org/10.1080/10867651.2003.10487582>.
- Li, K., Papademetris, X., Tucker, D.M., 2016. BrainK for structural image processing: creating electrical models of the human head. *Comput. Intell. Neurosci.* 2016. <https://doi.org/10.1155/2016/1349851>, 1349851.
- Liu, A., Joe, B., 1994. Relationship between tetrahedron shape measures. *BIT* 34, 268–287. <https://doi.org/10.1007/BF01955874>.
- Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M.A., Arbel, T., Avants, B.B., Ayache, N., Buendia, P., Collins, D.L., Cordier, N., Corso, J.J., Criminisi, A., Das, T., Delingette, H., Demiralp, Ç., Durst, C.R., Dojat, M., Doyle, S., Festa, J., Forbes, F., Geremia, E., Glocker, B., Golland, P., Guo, X., Hamamci, A., Iftekharruddin, K.M., Jena, R., John, N.M., Konukoglu, E., Lashkari, D., Mariz, J.A., Meier, R., Pereira, S., Precup, D., Price, S.J., Raviv, T.R., Reza, S.M.S., Ryan, M.,

- Sarikaya, D., Schwartz, L., Shin, H.C., Shotton, J., Silva, C.A., Sousa, N., Subbanna, N.K., Szekely, G., Taylor, T.J., Thomas, O.M., Tustison, N.J., Unal, G., Vasseur, F., Wintermark, M., Ye, D.H., Zhao, L., Zhao, B., Zikic, D., Prastawa, M., Reyes, M., Van Leemput, K., 2015. The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Trans. Med. Imaging* 34, 1993–2024. <https://doi.org/10.1109/TMI.2014.2377694>.
- Minjoli, S., Saturnino, G.B., Blicher, J.U., Stagg, C.J., Siebner, H.R., Antunes, A., Thielscher, A., 2017. The impact of large structural brain changes in chronic stroke patients on the electric field caused by transcranial brain stimulation. *NeuroImage Clin.* 15, 106–117. <https://doi.org/10.1016/j.nicl.2017.04.014>.
- Montes-Restrepo, V., Van Mierlo, P., Strobbé, G., Staelens, S., Vandenberghe, S., Hallez, H., 2014. Influence of skull modeling approaches on EEG source localization. *Brain Topogr.* 27, 95–111. <https://doi.org/10.1007/s10548-013-0313-y>.
- Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.M., 2011. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* <https://doi.org/10.1155/2011/156869>, 2011.
- Pechaud, M., Jenkinson, M., Smith, S., 2006. BET2 - MRI-Based Estimation of Brain, Skull and Scalp Surfaces FMRIB Technical Report TR06MP1. Report XXXIII, pp. 81–87.
- Perdue, K.L., Diamond, S.G., 2014. T1 magnetic resonance imaging head segmentation for diffuse optical tomography and electroencephalography. *J. Biomed. Opt.* 19 <https://doi.org/10.1117/1.JBO.19.2.026011>, 26011.
- Saturnino, G.B., Antunes, A., Thielscher, A., 2015. On the importance of electrode parameters for shaping electric field patterns generated by tDCS. *NeuroImage* 120, 25–35. <https://doi.org/10.1016/j.neuroimage.2015.06.067>.
- Shattuck, D.W., Leahy, R.M., 2002. Brainsuite: an automated cortical surface identification tool. *Med. Image Anal.* 6, 129–142. [https://doi.org/10.1016/S1361-8415\(02\)00054-3](https://doi.org/10.1016/S1361-8415(02)00054-3).
- Shattuck, D.W., Sandor-Leahy, S.R., Schaper, K.A., Rottenberg, D.A., Leahy, R.M., 2001. Magnetic resonance image tissue classification using a partial volume model. *NeuroImage* 13, 856–876. <https://doi.org/10.1006/nimg.2000.0730>.
- Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E.J., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobnjak, I., Flitney, D.E., Niazy, R.K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J.M., Matthews, P.M., 2004. Advances in functional and structural MR image analysis and implementation as FSL. In: *NeuroImage*. <https://doi.org/10.1016/j.neuroimage.2004.07.051>.
- Stenroos, M., Hunold, A., Hauelsen, J., 2014. Comparison of three-shell and simplified volume conductor models in magnetoencephalography. *NeuroImage* 94, 337–348. <https://doi.org/10.1016/j.neuroimage.2014.01.006>.
- Stenroos, M., Nummenmaa, A., 2016. Incorporating and compensating cerebrospinal fluid in surface-based forward models of magneto- and electroencephalography. *PLoS One* 11. <https://doi.org/10.1371/journal.pone.0159595> e0159595.
- Tadel, F., Baillet, S., Mosher, J.C., Pantazis, D., Leahy, R.M., 2011. Brainstorm: a user-friendly application for MEG/EEG analysis. *Comput. Intell. Neurosci.* 1–13. <https://doi.org/10.1155/2011/879716>, 2011.
- Thielscher, A., Antunes, A., Saturnino, G.B., 2015. Field modeling for transcranial magnetic stimulation: a useful tool to understand the physiological effects of TMS? *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS* 222–225. <https://doi.org/10.1109/EMBC.2015.7318340>.
- van der Walt, S., Schönberger, J.L., Nunez-Iglesias, J., Boulogne, F., Warner, J.D., Yager, N., Gouillart, E., Yu, T., 2014. scikit-image: image processing in Python. *PeerJ* 2. <https://doi.org/10.7717/peerj.453> e453.
- Van Leemput, K., Maes, F., Vandermeulen, D., Suetens, P., 1999. Automated model-based bias field correction of MR images of the brain. *IEEE Trans. Med. Imaging* 18, 885–896. <https://doi.org/10.1109/42.811268>.
- Vorwerk, J., Cho, J.H., Rampp, S., Hamer, H., Knösche, T.R., Wolters, C.H., 2014. A guideline for head volume conductor modeling in EEG and MEG. *NeuroImage* 100, 590–607. <https://doi.org/10.1016/j.neuroimage.2014.06.040>.
- Wells, W.M., Grimson, W.L., Kikinis, R., Jolesz, F. a., 1996. Adaptive segmentation of MRI data. *IEEE Trans. Med. Imaging* 15, 429–442. <https://doi.org/10.1109/42.511747>.
- Windhoff, M., Opitz, A., Thielscher, A., 2013. Electric field calculations in brain stimulation based on finite elements: an optimized processing pipeline for the generation and usage of accurate individual head models. *Hum. Brain Mapp.* 34, 923–935. <https://doi.org/10.1002/hbm.21479>.
- Wolters, C.H., Anwander, A., Tricoche, X., Weinstein, D., Koch, M.A., MacLeod, R.S., 2006. Influence of tissue conductivity anisotropy on EEG/MEG field and return current computation in a realistic head model: a simulation and visualization study using high-resolution finite element modeling. *NeuroImage* 30, 813–826. <https://doi.org/10.1016/j.neuroimage.2005.10.014>.
- Zhang, Y., Bajaj, C., Sohn, B.-S., 2005. 3D finite element meshing from imaging data. *Comput. Methods Appl. Mech. Eng.* 194, 5083–5106. <https://doi.org/10.1016/j.cma.2004.11.026>.
- Zhang, Y., Matuszewski, B.J., Shark, L.-K., Moore, C.J., 2008. Medical image segmentation using new hybrid level-set method. In: 2008 Fifth International Conference BioMedical Visualization: Information Visualization in Medical and Biomedical Informatics, pp. 71–76. <https://doi.org/10.1109/MediVis.2008.12>.
- Ziegler, E., Chellappa, S.L., Gaggioni, G., Ly, J.Q.M., Vandewalle, G., André, E., Geuzaine, C., Phillips, C., 2014. A finite-element reciprocity solution for EEG forward modeling with realistic individual head models. *NeuroImage* 103, 542–551. <https://doi.org/10.1016/j.neuroimage.2014.08.056>.