

## Article

# Symptom-Dependent Changes in MEG-Derived Neuroelectric Brain Activity in Traumatic Brain Injury Patients with Chronic Symptoms

Don Krieger <sup>1,\*</sup>, Paul Shepard <sup>2</sup>, Ryan Soose <sup>3</sup>, Ava M. Puccio <sup>1</sup>, Sue Beers <sup>4</sup>, Walter Schneider <sup>5</sup>, Anthony P. Kontos <sup>6</sup>, Michael W. Collins <sup>6</sup> and David O. Okonkwo <sup>1</sup>

<sup>1</sup> Department of Neurological Surgery, University of Pittsburgh, Pittsburgh, PA 15232, USA; PuccAM@UPMC.EDU (A.M.P.); okonkwodo@upmc.edu (D.O.O.)

<sup>2</sup> Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, PA 15232, USA; shepard@pitt.edu

<sup>3</sup> Department of Otolaryngology, University of Pittsburgh, Pittsburgh, PA 15232, USA; soosrj@UPMC.EDU

<sup>4</sup> Department of Psychiatry, University of Pittsburgh, Pittsburgh, PA 15232, USA; BeersSR@upmc.edu

<sup>5</sup> Department of Psychology, University of Pittsburgh, Pittsburgh, PA 15232, USA; wws@pitt.edu

<sup>6</sup> Department of Sports Medicine, University of Pittsburgh, Pittsburgh, PA 15232, USA; akontos@pitt.edu (A.P.K.); collinsmw@upmc.edu (M.W.C.)

\* Correspondence: kriegerd@upmc.edu



**Citation:** Krieger, D.; Shepard, P.; Soose, R.; Puccio, A.M.; Beers, S.; Schneider, W.; Kontos, A.P.; Collins, M.W.; Okonkwo, D.O. Symptom-Dependent Changes in MEG-Derived Neuroelectric Brain Activity in Traumatic Brain Injury Patients with Chronic Symptoms. *Med. Sci.* **2021**, *9*, 20. <https://doi.org/10.3390/medsci9020020>

Academic Editors: Rodolfo Gabriel Gatto and Antoni Torres

Received: 29 December 2020

Accepted: 17 March 2021

Published: 25 March 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** Neuroelectric measures derived from human magnetoencephalographic (MEG) recordings hold promise as aides to diagnosis and treatment monitoring and targeting for chronic sequelae of traumatic brain injury (TBI). This study tests novel MEG-derived regional brain measures of tonic neuroelectric activation for long-term test-retest reliability and sensitivity to symptoms. Resting state MEG recordings were obtained from a normative cohort, Cambridge Centre for Ageing and Neuroscience (CamCAN), baseline:  $n = 619$ ; mean 16-month follow-up:  $n = 253$ ) and a chronic symptomatic TBI cohort, Targeted Evaluation, Action and Monitoring of Traumatic Brain Injury (TEAM-TBI), baseline:  $n = 64$ ; mean 6-month follow-up:  $n = 39$ ). For the CamCAN cohort, MEG-derived neuroelectric measures showed good long-term test-retest reliability for most of the 103 automatically identified stereotypic regions. The TEAM-TBI cohort was screened for depression, somatization, and anxiety with the Brief Symptom Inventory and for insomnia with the Insomnia Severity Index. Linear classifiers constructed from the 103 regional measures from each TEAM-TBI cohort member distinguished those with and without each symptom, with  $p < 0.01$  for each—i.e., the tonic regional neuroelectric measures of activation are sensitive to the presence/absence of these symptoms. The novel regional MEG-derived neuroelectric measures obtained and tested in this study demonstrate the necessary and sufficient properties to be clinically useful—i.e., good test-retest reliability, sensitivity to symptoms in each individual, and obtainable using automatic processing without human judgement or intervention.

**Keywords:** CamCAN; TEAM-TBI; post-concussion syndrome; normative atlas; test-retest reliability; insomnia; depression; anxiety; somatization

## 1. Introduction

Traumatic brain injury (TBI) is a major cause of death and disability. Localized neuroelectric correlates of persistent functional sequelae after TBI would provide significant clinical value for diagnosis, disease monitoring, and targeted therapy.

It has been the informed expectation for a century that the keys to understanding the human brain will be found in measuring and understanding the electrical activity of neurons. Today, clinical neurophysiologists routinely measure single neurons to aid implantation of therapeutic devices deep in the brain [1]. Epileptologists use arrays of

implanted “stereo EEG” electrodes and the population recordings obtained from them to diagnose and guide the treatment of intractable seizure disorders [2].

Population neuronal activity is presumed to be the basis for human behavior. Stereo EEG and comparable invasive methods produce voltage recordings with resolution of a few millimeters at best from up to a few hundred recording sites. Because the electric field interacts strongly with the conducting tissue in the brain, these measures are difficult to localize for currents at a distance from the electrodes. This problem is particularly pronounced when the recordings are made noninvasively from electrodes placed on the scalp.

Magnetoencephalography (MEG) provides an alternative noninvasive measurement approach with several advantages over scalp and even implanted EEG recordings. Electrical current flow within populations of neurons is a fundamental constituent of brain function. The magnetic fields produced by these electric currents within the brain are measured at the MEG sensor array with high fidelity. Unlike electric fields, magnetic fields do not interact with brain tissue [3–6]. Therefore, in principle, the currents which are the sources of the measured magnetic field are more readily localized and their time course quantified.

For almost all MEG studies, the neuroelectric sources of the recorded magnetic fields are presumed to be due to population postsynaptic currents within the cerebral cortex [4–6]. For this reason, most source level analyses are constrained to identify neuroelectric currents in the cortex only, although there are occasional reports in which neuroelectric dipoles are localized to the white matter—e.g., [7]. Additional remarks are found in the Discussion section.

In the current study, MEG recordings were processed using the referee consensus solver, a method which extracts profuse localized and validated neuroelectric current waveforms ( $p < 10^{-12}$  for each), not only from cortical and subcortical volumes within the brain, but also from white matter volumes [8,9]. We employed a large set of MEG recordings from healthy volunteers (CamCAN) to generate a normative database, including metrics of test–retest reliability across MEG studies repeated 16 months apart. Good test–retest reliability is a primary requirement for both scientific and diagnostic usefulness. We then analyzed MEG-derived regional neuroelectric measures in a cohort of chronic TBI subjects with persistent symptoms (Targeted Evaluation, Action and Monitoring of Traumatic Brain Injury (Team-TBI)) to identify patterns of brain activity which distinguish between those with and without specific TBI sequelae. Sensitivity to clinical symptoms is a second critical requirement for clinical usefulness.

The primary objective of this work was to obtain and validate clinically useful neuroelectric measures localized within the brain. The *mean* and *standard deviation* for each region from the CamCAN baseline MEG recordings were used to transform the corresponding regional measure for each recording from both cohorts to *z-scores*. The baseline and follow-up CamCAN *z-scores* were used to test for test–retest reliability. The baseline and follow-up Team-TBI *z-scores* were used to test for sensitivity to symptoms. Consistent with the translational objective of the effort, the data processing pipeline was deployed (a) to function without human judgement or intervention and (b) to fully process each new recording within 24 h [8].

The solver identifies high-fidelity waveforms for 80 msec instances of dipole electric current flow localized to about  $1 \text{ mm}^3$  [3,4]. The normalized count of these instances within each standardized region is the measure used in the present study—i.e., each measure is the total value over a brain region automatically identified from one subject’s scan. The fact that each standardized region is automatically identifiable for each scan using Freesurfer [10,11] enables generating standardized measures, i.e., norms, for each region across a large normative cohort. The values for those regions in any individual may then be compared with the norms to assess the normality of the individual’s regional measures.

Note that each measure’s localization is limited by the automated parcellation accuracy of Freesurfer and by the volume of each region. For the CamCAN and TEAM-TBI cohorts reported here, those volumes range from less than  $1.0 \text{ cm}^3$ , e.g., left or right nucleus accumbens, to  $30\text{--}70 \text{ cm}^3$ , e.g., left or right cerebellar cortex.

The results reported here enable rejection of the following null hypotheses. (a) Individuals with and without symptoms are indistinguishable. (b) The cohort membership of each individual (TEAM-TBI or CamCAN) cannot be determined. (c) Regional measures from an individual do not reliably repeat. (d) Within the individual, the neuroelectric activity within a cortical region is indistinguishable from the adjacent rim of white matter.

Joint rejection of hypotheses (a) and (c) supports the potential for these measures as clinically useful in the diagnosis and treatment of insomnia, depression, anxiety, and somatization, common sequelae of TBI. Rejection of hypothesis (a) and (b) suggests the possibility that these measures may be useful as biomarkers for TBI. The emphasis in the study design is towards symptoms rather than etiology. It is hoped that this shift in focus will produce insights which are useful in diagnosis and treatment of those symptoms, regardless of etiology. Repeated MEG recordings during drug or other treatment modalities may provide objective and useful information in assessing treatment efficacy and in making adjustments to the treatment.

## 2. Results

The primary objective of this work was to obtain and validate clinically useful neuroelectric measures localized within the brain. The *mean* and *standard deviation* for each region from the CamCAN baseline MEG recordings were used to transform the corresponding regional measure for each recording from both cohorts to *z-scores*. The baseline and follow-up CamCAN *z-scores* were used to test for test-retest reliability. The baseline and follow-up Team-TBI *z-scores* were used to test for sensitivity to symptoms.

Regional measures of neuroelectric activity for 17 subcortical, 68 cortical, and 18 deep white matter regions were extracted from the MEG recordings for each study participant. For the normative CamCAN cohort, the *mean* and *standard deviation* baseline values for each region are shown in Appendix A. These values were used to transform all regional measures to *z-scores*. The tables also include *correlations* and *differences* for baseline vs. follow-up measures. These were used to assess test-retest reliability. The presence/absence of relationships between these neuroelectric measures and measures of potential clinical relevance was tested.

This effort produced several types of results. Significant relationships were found between measures of tonic regional neuroelectric activity and (1) screening measures of insomnia and psychological distress and (2) subjects in the CamCAN control cohort vs. the TEAM-TBI chronically symptomatic group with history of concussion. These directly impact on the potential usefulness of the measures as diagnostics and as probes for scientific questions.

In addition, (3) both short-term (1 h) and long-term (16-month average) baseline vs. follow-up test-retest reliability results are reported. This too impacts directly on potential clinical utility. Finally, (4) the spatial resolution and statistical power of the neuroelectric measures is demonstrated by identifying significant differential activity between cortical and adjacent white matter regions.

Note that all regional neuroelectric activity measures were reduced to *z-scores*; the means and standard deviations of the baseline CamCAN recordings ( $n = 619$ ) were used for the *z-score* transformation.

### 2.1. Insomnia and Psychological Distress

MEG-derived neuroelectric measures were obtained from 63 TEAM-TBI subjects at baseline and from the 40 who returned for follow-up. Symptom surveys for insomnia and three symptoms of psychological distress were obtained from all but one of the baseline subjects. Standard inventories were for insomnia (Insomnia Severity Index, ISI, [12,13]) and somatization, depression and anxiety (Brief Symptom Inventory, BSI, [14,15]). Cut-offs of 15 (ISI) and 63 (BSI) were used to divide both the baseline and follow-up TEAM-TBI recordings into clinically negative or positive groups.

Regional measures of neuroelectric activity for 17 subcortical and 68 cortical regions were combined into classifiers using stepwise linear classification [16–18]. Classification accuracies with  $p$ -values are shown in Table 1 (Figure 1). The  $p$ -values were computed as follows. Considering line 1 of said table, 42 of 54 TEAM-TBI subjects who screened negative for insomnia were classified as negative and 12 as positive. The chance that this would happen by chance is equivalent to the odds of getting at least 42 heads when we flip a fair coin 54 times. For each symptom, both sides of the classification have significant  $p$ -values—i.e., the classifier does well in classifying both those who screen positive and those who screen negative. This provides confidence that the neuroelectric measures which comprise the classifier are related to the symptoms.

The regions whose measures were included are shown in Table 2 (Figure 1). For each symptom, a second classification function was constructed, for which the regions that were included in the first run were excluded. This second run produced significant classification accuracy for insomnia only, as indicated in the tables. This suggests (a) elevated confidence in the relationship between the regions whose neuroelectric measures were used for each classification and (b) that the regional measures included in the second classification function for insomnia are highly correlated with linear combinations of the first set. That is why they were not included in the first classification run.

Classification accuracy reached significance for all four symptoms for both clinically negative and positive groups. These results were weakest for anxiety. Classification accuracy was comparable for baseline and follow-up records; baseline vs. follow-up clinical rating changed for insomnia ( $n = 10$ ), somatization ( $n = 9$ ), depression ( $n = 10$ ), and anxiety ( $n = 15$ ), almost all for the better.

**Table 1.** Eight-five cortical and subcortical regional measures were trained as linear classifiers by groups determined by symptom survey clinical thresholds. Jackknifed classification accuracies are shown. The results for insomnia (2nd step) are highlighted in gray. These were obtained when the regions which were selected for the first classification analysis were excluded. Symptom survey scores and regional neuroelectric measures were used from both baseline ( $n = 62$ ) and follow-up ( $n = 40$ ) sessions for all subjects of the traumatic brain injury (TEAM-TBI) cohort. See the text for details.

	Classified Negative	Classified Positive	Percentage	$p$ -Value
insomnia				
clinically negative	42	12	77.8%	0.000007
clinically positive	11	37	77.1%	0.000031
insomnia—2nd step				
clinically negative	43	11	79.6%	0.0000017
clinically positive	13	35	72.9%	0.00036
somatization				
clinically negative	46	12	79.3%	0.000001
clinically positive	14	30	68.2%	0.0048
depression				
clinically negative	57	13	81.4%	0.00000001
clinically positive	9	23	71.9%	0.0035
anxiety				
clinically negative	37	18	67.3%	0.0032
clinically positive	12	35	74.5%	0.00017



**Table 2.** The regions whose neuroelectric activity values contributed to the symptom-Scheme 1. are shown. Activity in regions marked “+” was higher in those who screened positive; those marked “-” were lower. Regions marked “++” or “--” made the largest statistical contribution to the classifier. The results for insomnia (2nd step) are highlighted in gray. These were obtained when the regions which were selected for the first classification analysis were excluded.

	Insomnia	Somatization	Depression	Anxiety
Brain-Stem				+
L Amygdala		+		
L Hippocampus	-	-		
L Thalamus				-
R Hippocampus	+			
R Thalamus			++	
L caudal middle frontal		-		
L frontal pole	-			
L inferior parietal			-	
L insula	-			
L lateral occipital		-		
L lateral orbito-frontal		+	+	
L paracentral	-			
L pars opercularis			+	
L pars triangularis	+			
L precentral			-	
L rostral anterior cingulate				++
L rostral middle frontal		+		
L superior frontal		-		
L superior parietal	-			
L temporal pole	+			
R bankssts	-			
R caudal anterior cingulate	+	-	-	
R caudal middle frontal	-			
R cuneus			-	-
R inferior parietal		-		
R insula				++
R isthmus cingulate	-			
R lateral occipital		-	-	
R lateral orbito-frontal	-			
R medial orbito-frontal			+	
R middle temporal			-	
R pericalcarine	+			
R posterior central			+	+
R posterior cingulate			-	
R supramarginal	+	+		

The magnitude of the difference between the nonclinical and clinical groups was examined by testing the difference in the mean scores on each classifier between the groups. This approach provides a complementary view on the magnitude and significance of the differences between the groups achievable via measures of regional neuroelectric activity.

The results are shown in Table 3. As might be expected, this produced comparable results to those shown in Table 1, albeit with greater significance.

**Table 3.** For each symptom, the difference between those who screened negative vs. positive was Table 1. Levene’s *F*-statistic was used to test for significant differences in the variability of the comparison groups. Although no significant differences were found in variability, the most conservative adjusted *df* values were used to compute the *p*-values corresponding to the *t*-statistics. As in Tables 1 and 2 above, results are highlighted in gray for insomnia (2nd step). These were obtained when the regions which were selected for the first classification analysis were excluded. See the text for details.

	<i>t</i>	<i>df</i> (Adjusted)	<i>p</i> -Value	Levene	<i>p</i> -Value
insomnia	9.20	87.8	10 <sup>−14</sup>	0.10	0.756
insomnia—2nd step	6.70	86.8	10 <sup>−9</sup>	0.39	0.532
somatization	8.63	73.6	10 <sup>−13</sup>	1.27	0.262
depression	10.15	56.6	10 <sup>−16</sup>	0.09	0.763
anxiety	2.14	90.6	0.025	0.00	0.980

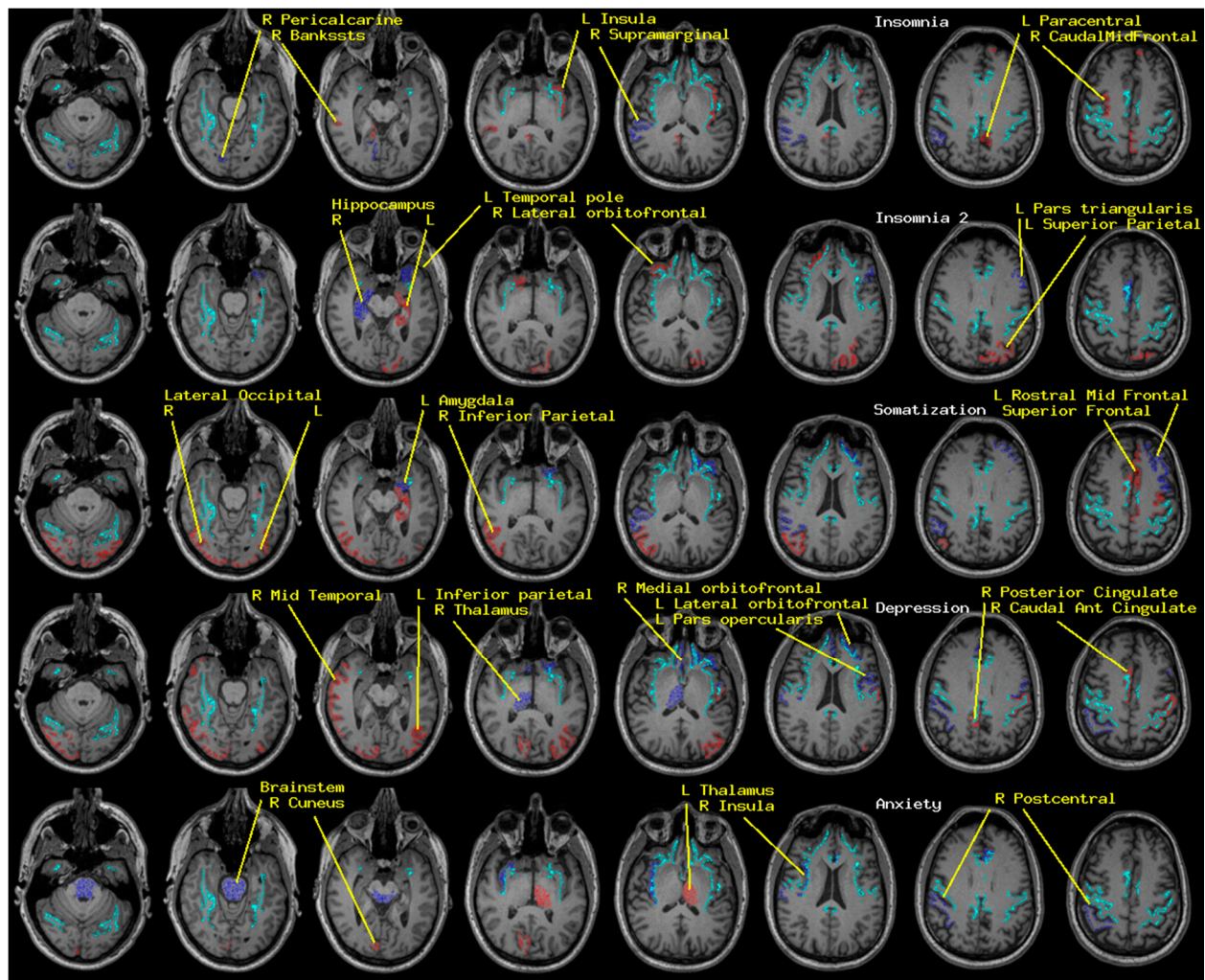
Finally, to provide information regarding the independence of the results for the different symptoms, coincidence rates for pairs of symptoms and correlations between classifier scores are shown in Table 4. These are the same classifiers reported in Table 1 and the same classifier scores reported in Table 3. In Table 4, the left panel shows that 60% or more of the TEAM-TBI subjects have at least two of the four symptoms. Note that the acceptance criteria for the TEAM-TBI study included “high symptom burden”. The correlations (right panel) are relatively high between pairs of BSI classifier scores and relatively low between the insomnia classifier score and each of the BSI classifiers.

**Table 4.** Left panel: symptom coincidence rate. The fraction of the TEAM-TBI cohort who screened positive for at least two symptoms is shown. For example, 61% of the TEAM-TBI cohort screened positive for both insomnia and depression. Right panel: correlations between classifier scores for the TEAM-TBI cohort. See the text for details.

Symptom Coincidence Rates			Classifier Score Correlations		
Somatization	Depression	Anxiety	insomnia	Depression	Anxiety
0.67	0.61	0.71	0.269	0.258	0.383
	0.66	0.70	0.599		0.666
		0.66			0.781

## 2.2. CamCAN vs. TEAM-TBI Cohort

MEG recordings were obtained from 619 CamCAN subjects at baseline, 253 at follow-up, 63 TEAM-TBI subjects at baseline and 40 at follow-up. High resolution T1 weighted MR imaging (MRI) was obtained from all members of both cohorts. Diffusion-weighted imaging (DWI) was obtained from 589 CamCAN subjects, and all of the TEAM-TBI subjects. DWI is required to obtain regional neuroelectric measures for deep white matter tracts.



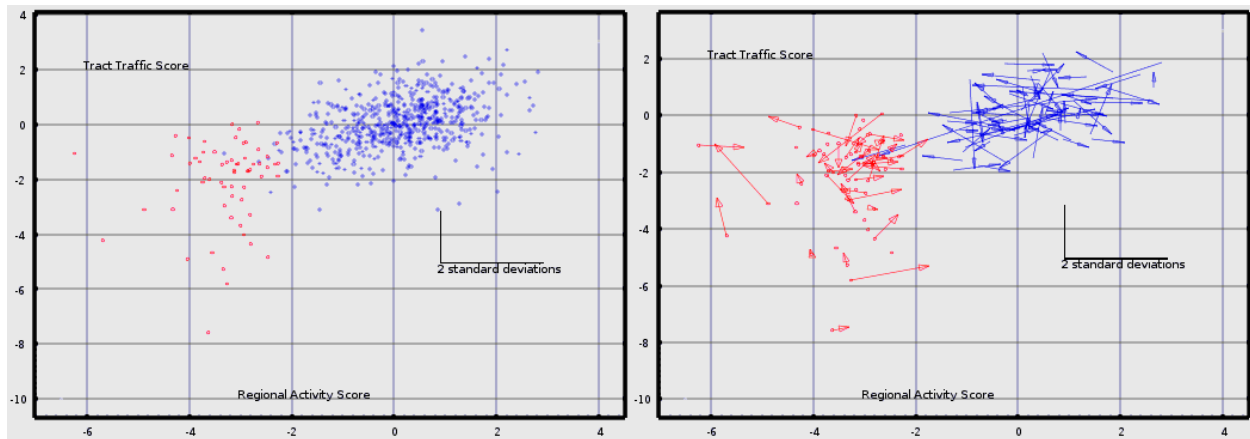
**Figure 1.** The regions whose neuroelectric activity values contributed to the symptom-specific classifiers reported in Tables 1 and 2 are shown. Each row shows the regions for the indicated symptom. From top to bottom, they are insomnia, insomnia (2nd step), somatization, depression, and anxiety. The MR imaging (MRI) slices range from left to right, inferior to superior in one-centimeter increments. The left side of each slice is the right brain. Activity in blue/red regions was higher/lower in those who screened positive. The cyan landmarks are the boundaries between gray and white matter in the precentral, cingulate, insula, and fusiform regions.

Regional measures of neuroelectric activity for 17 subcortical and 68 cortical regions were combined into classifiers using stepwise linear classification [16]. Classification accuracies with  $p$ -values are shown in Table 5. The cortical and subcortical regions which contributed to the classifier in order of their statistical contribution were R thalamus, R cerebellum, R middletemporal, R hippocampus, R lateraloccipital, L isthmuscingulate, L thalamus, L fusiform, brain-stem, and 11 others. When these 20 regions were excluded, classification accuracy remained highly significant albeit reduced. The contributing regions in order were L insula, L lingual, R superior temporal, L caudate, L parahippocampal, and seven others.

As for the cortical and subcortical regions, regional measures of neuroelectric activity for 18 deep white matter tracts were combined into classifiers. Classification accuracies with  $p$ -values are shown in Table 5. The regions which contributed to the classifier in order were left cortico-spinal tract (L\_cst), right inferior longitudinal fasciculus (R\_ilm), R\_cst, forceps minor (fminor), and six others.

Discriminant analysis reduces all of the measures for an individual to a single score for each classifier. Here, each individual has two scores, one for the cortical/subcortical

classifier and one for the deep white matter tract classifier. Figure 2 shows these scores in bivariate plots to illustrate the “spatial” separation which results in the classification accuracies shown in Table 5 (Figure 3).

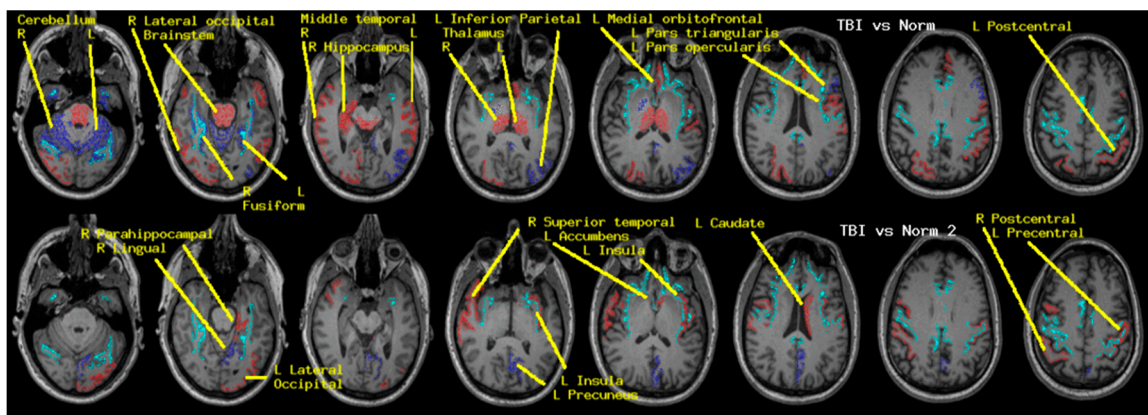


**Figure 2.** The classifier scores for each baseline CamCAN subject (blue,  $n = 589$ ) and each TEAM-TBI subject (red,  $n = 63$ ) are shown in the **left**. The score on the cortical/subcortical classifier is plotted on the x-axis; the score on the deep white matter classifier is plotted on the y-axis. These are the classifiers whose accuracies are shown in Table 5. In the **right** panel, all baseline TEAM-TBI scores are plotted,  $n = 63$ . Those who returned for follow-up ( $n = 40$ ) are plotted as arrows; the baseline is plotted at the base of the arrow; the follow-up is plotted at the arrow-head. 63 age and sex matched CamCAN subjects who returned for follow-up are plotted in blue. The standard deviation bars represent 2.0 *standard deviations* for the classifier scores.

**Table 5.** 85 cortical and subcortical regional measures were trained as linear classifiers by cohort using the baseline CamCAN and TEAM-TBI measures. Jackknifed classification accuracies are shown for classifiers using 85 cortical and subcortical regions (lines 1–10) and 18 deep white matter tracts (lines 11–15). The results highlighted in gray were obtained when the regions which were selected for the first classification analysis were excluded. See the text for details.

	CamCAN	TEAM-TBI	Percentage	<i>p</i> -Value
Cortical/Subcortical				
CamCAN baseline	581	38	93.9%	$10^{-126}$
CamCAN follow-up	227	25	90.1%	$10^{-42}$
TEAM-TBI baseline	0	63	100.0%	$10^{-18}$
TEAM-TBI follow-up	0	40	100.0%	$10^{-12}$
cortical/subcortical—2nd step				
CamCAN baseline	504	115	81.4%	$10^{-126}$
CamCAN follow-up	228	24	90.5%	$10^{-42}$
TEAM-TBI baseline	2	61	96.8%	$10^{-18}$
TEAM-TBI follow-up	3	37	92.5%	$10^{-12}$
deep white matter				
CamCAN baseline	518	71	87.9%	$10^{-85}$
CamCAN follow-up	208	31	87.0%	$10^{-33}$
TEAM-TBI baseline	14	48	77.4%	0.0000024
TEAM-TBI follow-up	9	30	76.9%	0.00015





**Figure 3.** The regions whose neuroelectric activity values contributed to the cohort-specific classifiers reported in Table 5 and Table 6 are shown. The MRI slices range left to right, inferior to superior, in one-centimeter increments. The left side of each slice is the right brain. Activity in blue/red regions was higher/lower in the TEAM-TBI cohort. The cyan landmarks are the boundaries between gray and white matter in the precentral, cingulate, insula, and fusiform regions.

The components of each classifier are the measures for each included region. The discriminant analysis algorithm identifies those regions for which there are differences between the cohorts and constructs a linear combination of those regions' measures which adds the differences to enable significant separation/classification.

Figure 4 illustrates the differences for each subcortical region. A dot is plotted for each subject of the CamCAN cohort (upper panel) and the TEAM-TBI cohort (lower panel). Since each measure is a z-score computed from the *mean* and *standard deviation* for that region from the CamCAN cohort, the plots for the CamCAN cohort all have a *mean* of 0.0, *standard deviation* of 1.0, and are nearly symmetrically distributed about the *mean*.

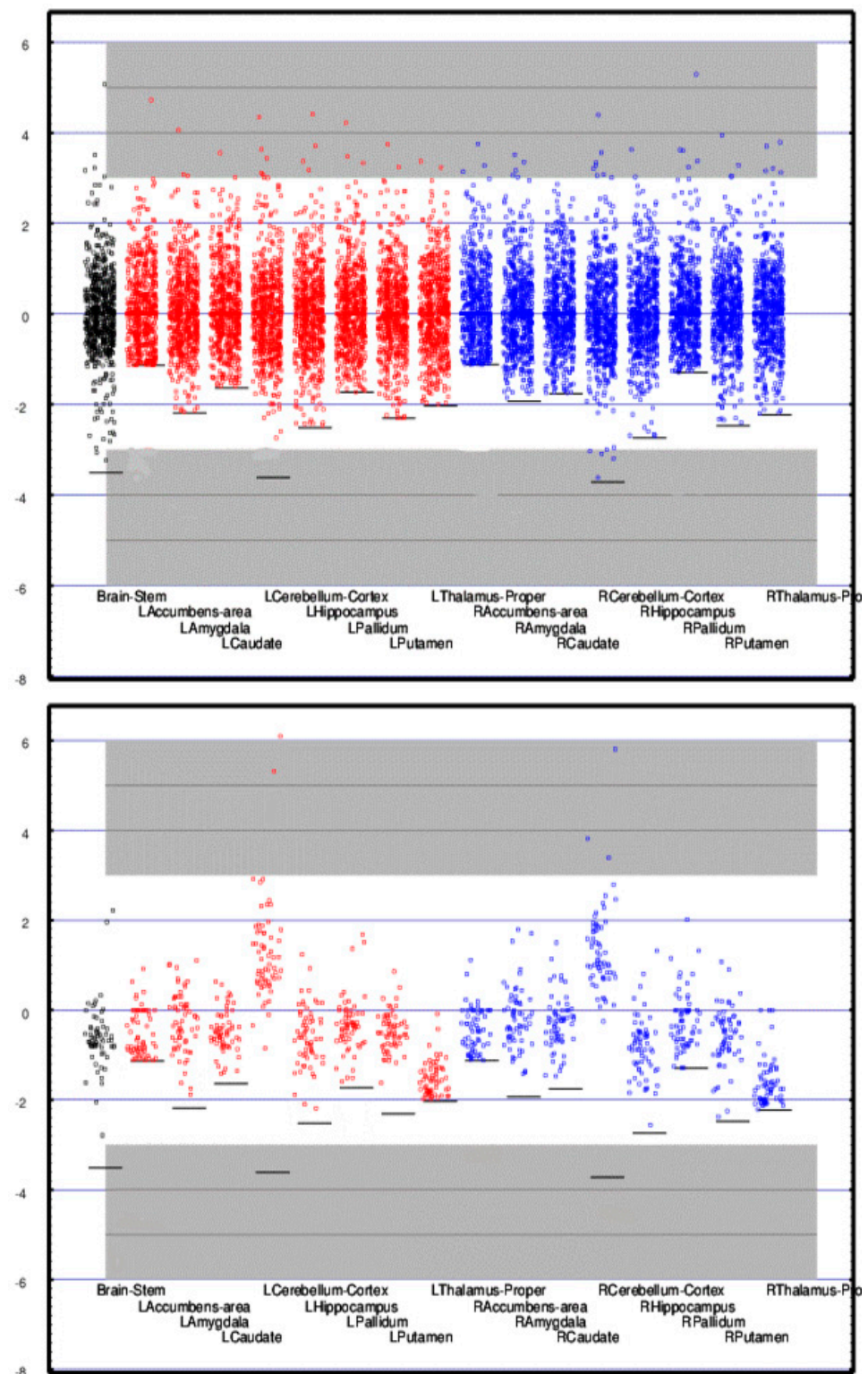
Lack of symmetry primarily results from the fact that each measure has a floor since each measure is a normalized count. The floor for each region is indicated in the figure with a horizontal bar. No individual's measure for a region can fall below that bar; this results in bunching of the values below the *mean*. This is most pronounced for the accumbens regions for which the floor is only slightly more than one *standard deviation* below the *mean*.

The regions whose contributions to the classifier were greatest are listed above and in Figure 4 (subcortical) and Figure 5 (right cortical). The *means* for each of those regions is markedly displaced from 0.0 as expected.

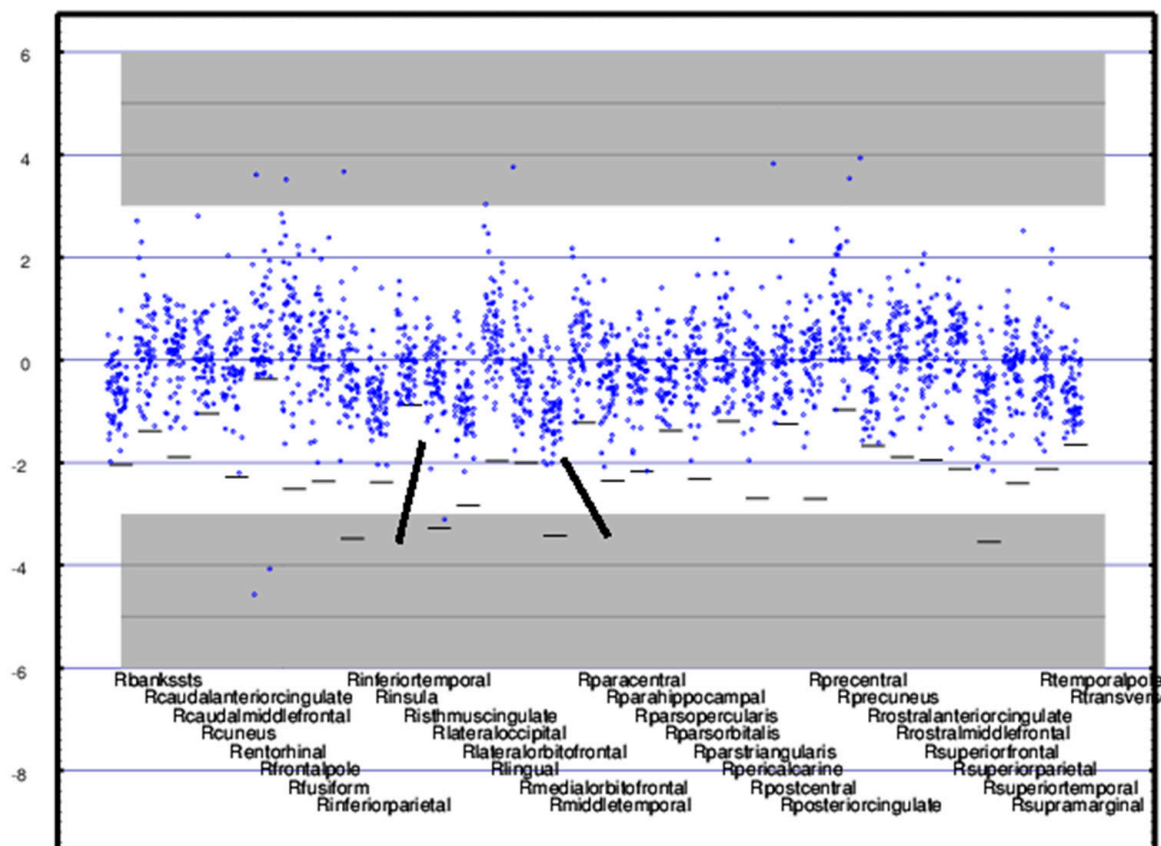
### 2.3. Test-Retest Reliability

Regional measures of neuroelectric activity for 103 cortical, subcortical, and deep white matter regions were extracted from the resting and task MEG recordings for each subject of both cohorts. The values for the CamCAN cohort were used to assess short-term and long-term test-retest reliability. For short-term reliability, baseline resting vs. task recordings obtained in the same sitting were used. For long-term, baseline vs. follow-up resting recordings were used (*mean* interval = 16 months). The long-term correlations and *mean* differences for each region are listed in Appendix A and are plotted on the y-axis in Figure 6 in blue and red, respectively; the short-term correlations and *mean* differences are plotted on the x-axis. Note that for *mean* difference, the values are z-scores.





**Figure 4.** Regional activity (*z-score*) for 17 subcortical regions is shown for each CamCAN (**upper** panel) and TEAM-TBI (**lower** panel) subject. Since the *means* and *standard deviations* for the CamCAN subjects were used to compute the *z-scores*, the *mean* for each of the CamCAN regions is zero and the *z-scores* for each region are distributed approximately normally about zero. The *mean* and Scheme 3. 0 or  $z > 3.0$ —i.e.,  $p < 0.0014$ . Individuals whose measures fall into those portions of the graph that are highlighted in gray are far from the norm. Note that assessing sub-*mean* normality is limited by the floor.



**Figure 5.** Regional activity (*z-score*) is shown for each TEAM-TBI subject for 34 right cortical regions. As in Figure 2, each horizontal bar represents the floor below which no measure can go. The right middle temporal and lateral occipital cortices, indicated with the wide black bars, contributed significantly to the classifier. Note that the *means* for some regions are greater than zero, some less. See the Figure 4 legend for details.

The correlations are centered about 0.8 for short-term reliability; for long-term reliability, they are centered around 0.45. The *mean* differences are consistently within 0.03 of 0.0 for the short-term reliability and range between  $\pm 0.55$  for the long-term reliability. Hence, test-retest reliability is high in the short term and moderate in the long term.

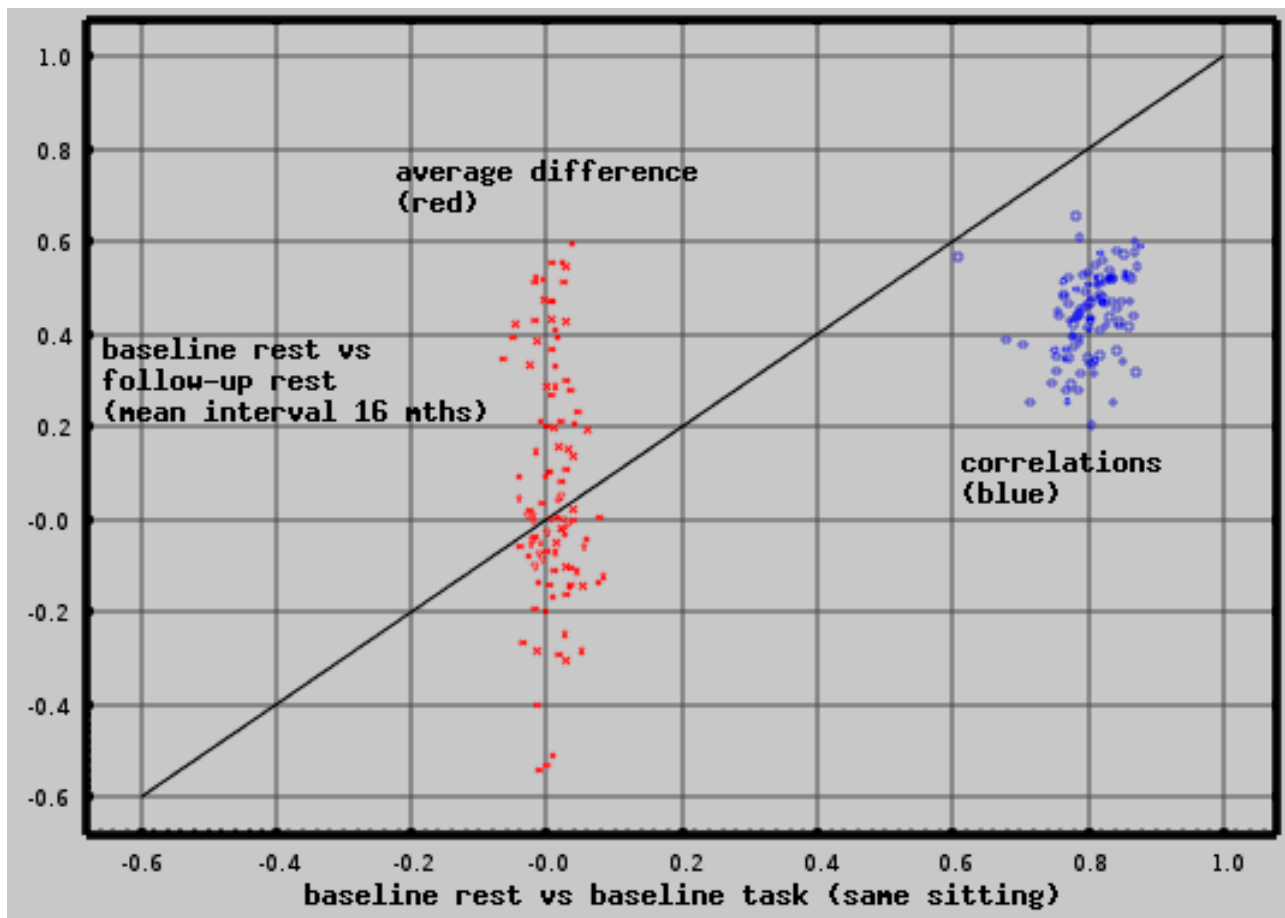
Note that the resting recordings were obtained with eyes closed, while the task recordings were obtained with eyes open. Yet, the comparisons of these, reported here as short-term correlations, are quite high, demonstrating that these measures, unlike fMRI measures, are relatively insensitive to eyes open vs. eyes closed.

#### 2.4. Differential Activity: Cortical vs. Adjacent White Matter Regions

For each of 68 cortical regions, Freesurfer identified an adjacent white matter region with a maximum thickness of 5 mm. For each such pair of regions, i.e., cortex and adjacent white matter rim, the difference in activity can be tested for significance by comparing the observed current counts within the regions to the expected counts given the volumes of the regions. This is not only a test of the spatial resolution of the referee consensus solver, but in addition may provide useful neurophysiological information. See the Discussion section for additional comments.

For most regions, there are thousands of counts so there is considerable statistical power to identify differences using the  $\chi^2$  statistic. For each of the 619 baseline CamCAN subjects, there are 68 cortex/white matter region pairs—i.e., 42,092 in total. To reduce false positives due to the large number of comparisons,  $p < 10^{-8}$  was used as the threshold for significance. In total, 14,187 (33.7%) of the pairs demonstrated greater cortical activity than white matter activity. This supports the claim that the solver's resolution is less than 5 mm.

Surprisingly, 18,129 (43%) of the pairs demonstrated greater white matter activity than cortical activity. Additional comments may be found in the Discussion section.



**Figure 6.** Short-term (same sitting) vs. long-term (*mean* 16-months) test-retest reliability is plotted for each of the 103 cortical, subcortical, and deep white matter regions. Regional measures (*z-scores*) of tonic neuroelectric activity were compared across all CamCAN subjects. Comparisons were between baseline resting and task recordings (short-term reliability, about one hour) and between baseline and follow-up resting recordings (long-term reliability, *mean* 16-month interval). Correlations (blue) and mean differences (red) are plotted—short term (x-axis) vs. long term (y-axis). Long-term test-retest reliability shows reduced correlations and increased differences. Note that mean differences  $>0.1$  or  $<-0.1$  are typically significant with  $p < 0.03$ .

### 3. Materials and Methods

Magnetoencephalographic (MEG) recordings were processed from each subject of two cohorts: (1) the normative CamCAN cohort,  $n = 619$  at baseline, ages 18–87 [19,20],  $n = 253$  at follow-up, and (2) the chronically symptomatic concussed TEAM-TBI cohort,  $n = 63$  at baseline, ages 21–60,  $n = 40$  at follow-up. The MEG recordings were coregistered with high-resolution T1-weighted and DWI weighted MRI scans obtained at baseline.

The primary objective of this work was to obtain and validate clinically useful neuroelectric measures localized within the brain. The *mean* and *standard deviation* for each region from the CamCAN baseline MEG recordings were used to transform the corresponding regional measure for each recording from both cohorts to *z-scores*. The baseline and follow-up CamCAN *z-scores* were used to test for test-retest reliability. The baseline and follow-up Team-TBI *z-scores* were used to test for sensitivity to symptoms. Consistent with the translational objective of the effort, the data processing pipeline was deployed (a) to function without human judgement or intervention and (b) to fully process each new recording within 24 h [8].

The raw MEG data from each subject were initially transformed to a collection of probabilistically validated neuroelectric currents. Each current is 80 msec in duration and is localized in time and space with a resolution of one millisecond (msec) and better than 5 mm (mm). This primary processing step yielded profuse high-resolution neurophysiological measures from within the brain of each subject. The total current count per subject per minute is typically in excess of 500,000.

The current counts were normalized to produce measures of tonic activity for each of 171 standard regions of interest (ROIs): 17 subcortical regions, 68 cortical regions, 68 adjacent white matter regions, and 18 deep white matter tracts. Each regional measure is a count of all the neuroelectric currents localized within the region over the several-minute recording time. The statistical power when comparing counts to test for regional differences is high because the count for each region is high. A final adjustment was applied to each regional measure based on comparable measures obtained from empty-room recordings obtained on the same day.

### 3.1. CamCAN Dataset

The Cambridge Centre for Ageing and Neuroscience (CamCAN) Stage 2 cohort study is a large cross-sectional adult lifespan study (ages 18–87) of the neural underpinnings of successful cognitive ageing [19,20]. The study was conducted in compliance with the Helsinki Declaration, and was approved by the local ethics committee, Cambridgeshire 2 Research Ethics Committee (reference: 10/H0308/50) [20].

The work reported here utilized the majority subset ( $n = 619$ ) of the cohort for whom high-resolution (1 mm) anatomic T1-weighted MR imaging and MEG recordings were available. Of these, 253 follow-up resting recordings were obtained (Stage 3 longitudinal study) with a *mean* interval of 16 months between MEG studies. Diffusion-weighted imaging (DWI) was obtained for 589 of the baseline subjects, 240 of whom returned for follow-up.

MR imaging was obtained on all subjects at a single site using a 3T Siemens TIM Trio scanner with 32-channel head coil. T1 scans were obtained using the MPRAGE sequence. The field of view for these scans was  $256 \times 240 \times 192$  at 1 mm resolution. DWI scans were acquired ( $n = 589$ ) with a twice-refocused spin-echo sequence, with 30 diffusion gradient directions for each of the two b-values, 1000 and 2000  $s/mm^2$ , plus three images acquired with a b-value of 0. Other parameters are: TR = 9100 msec, TE = 104 msec, voxel size = 2.0 mm, FOV =  $96 \times 96$  mm, 66 axial slices [20].

MEG recordings were collected at a single site using a 306-channel VectorView MEG system (Elekta Neuromag, Helsinki). The data were sampled at 1 KHz with antialiasing low-pass filter at 330 Hz and high-pass filter at 0.03 Hz. Continuous head position measures were enabled throughout the recordings. All recordings were obtained with the subject sitting up.

At baseline, 560 s were recorded continuously with eyes closed resting [19]. In the same sitting, 560 s were recorded continuously during performance of a sensorimotor task ( $n = 619$ ). To test for short-term test-retest reliability, MEG-derived measures were compared between the baseline resting and sensorimotor task recordings. The implications of this test are detailed in the Discussion.

For the sensorimotor task, subjects detected visual and auditory stimuli and responded to detection of each with a button press with the right index finger. The stimuli were two circular checkerboards presented simultaneously to the left and right of a central fixation cross, 34 msec duration, and a binaural tone of 300 msec duration. The tone was at 300, 600, or 1200 Hz in equal numbers with the order randomized. In total, 121 trials were presented with simultaneous visual and auditory stimulation. Eight trials were randomly intermixed in which one stimulus was presented at a time—four visual and four auditory. This was carried out to discourage dependence on one stimulus modality only. The average intertrial interval was approximately 4.3 s



At follow-up, 320 s were recorded continuously with eyes closed resting,  $n = 253$  [20]. At both baseline and follow-up sessions, 60 s of empty-room recordings were obtained.

### 3.2. TEAM-TBI Dataset

The chronic TBI subject dataset was derived from the Targeted Evaluation, Action and Monitoring of Traumatic Brain Injury (TEAM-TBI) study, a personalized medicine research program for subjects with chronic TBI sequelae at the University of Pittsburgh (clinicaltrials.gov: NCT02657135). All TEAM-TBI subjects gave their informed consent for inclusion before they participated in the study. The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Institutional Review Board of The University of Pittsburgh (PRO13070121).

Inclusion criteria were ages 18–60 with a history of one or more TBI more than six months prior to the study, with severe persistent chronic TBI sequelae as assessed with the post-concussion symptom severity (PCSS) scale—i.e., high chronic symptom load [21]. In total, 61 of the 63 subjects with MEG recordings had sustained “mild” TBIs. TEAM-TBI subjects underwent a 4-day comprehensive clinical assessment, including advanced neuroimaging, followed by multidisciplinary adjudication of clinical syndromes. TEAM-TBI subjects then completed 6 months of supervised, targeted therapy. Subjects returned to Pittsburgh for a follow-up evaluation (*mean* interval = 6.4 months) to document impact of treatment on identified clinical disorders.

MR imaging was obtained for all subjects at a single site using a 3T Siemens TIM Trio scanner with 32-channel head coil. T1 scans were obtained using the MPRAGE sequence. The field of view for these scans was  $256 \times 256 \times 176$  at 1 mm resolution. DWI scans were acquired ( $n = 64$ ) with a twice-refocused spin-echo sequence, with 64 diffusion gradient directions at b-values of 1000 and 3000  $s/mm^2$ , and 128 directions at b-values of 5000 and 7000. Additional parameters for the four b-values were: TR = 3700, 3700, 4100, 4500 msec, TE = 92, 125, 147, 164 msec. voxel size = 2.4 mm, FOV = 230.4 mm, 63 axial slices.

MEG recordings were collected at a single site using a 306-channel VectorView MEG system, Elekta Neuromag, Helsinki. The data were sampled at 1 KHz with an antialiasing low-pass filter at 330 Hz and high-pass filter at 0.03 Hz. Continuous head position measures were enabled throughout the recordings. All recordings were obtained with the subject sitting up.

At baseline four 200 s resting recordings were obtained with eyes open and fixated with the room darkened ( $n = 63$ ). Four to eight recordings were obtained totaling 1500 s with the lights on during performance of a visual semantic decision task [20]. The protocol was the same at follow-up ( $n = 40$ ).

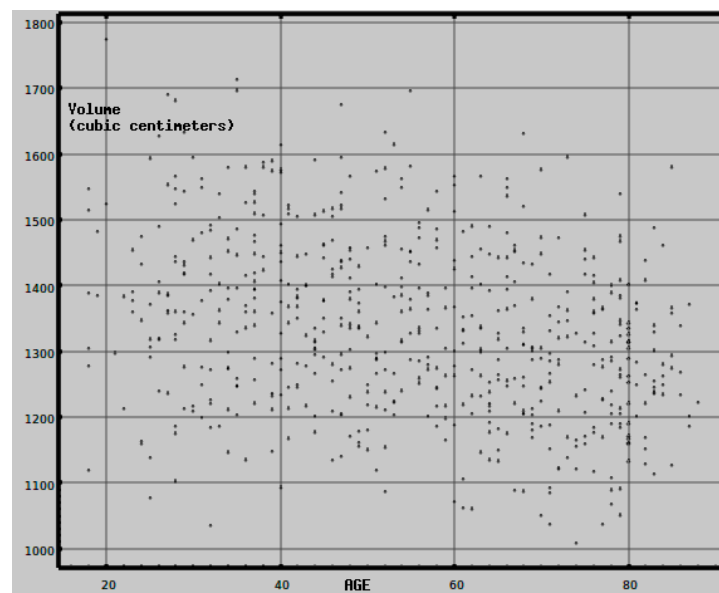
Baseline resting MEG recordings were used ( $n = 63$ ) for subjects whose high-resolution (1 mm) anatomic T1 and MEG recordings were available. Of these, 40 follow-up resting recordings were obtained—*mean* interval = 6.4 months. DWIs were obtained for 63 of the 64 baseline recordings and 39 of the 40 follow-up recordings. At both baseline and follow-up sessions, 300 s of empty-room recordings were obtained.

### 3.3. MRI Processing

Each high-resolution T1 scan was processed with Freesurfer, version 5.3, using its default Desikan–Killiany atlas parcellation [10,11]. Freesurfer is a segmentation package which automatically and reliably identifies brain regions. The 3-dimensional coordinates of the extent of the brain volume and 153 standardized regions of interest (ROIs) were identified—68 cortical regions, 68 adjacent white matter rims of tissue with thickness  $\leq 5.0$  mm, and 17 subcortical regions. Figure 7 shows the expected descending relationship between brain volume and age.

Each DWI scan was processed with Tracula, version 1.22 [22]. Tracula is a fully automated package which identifies the 3-dimensional coordinates of the volumes occupied by 18 deep white matter tracts.





**Figure 7.** The brain volume for each subject was obtained using Freesurfer (see text). Brain volume for this normative cohort shows the expected decrease with age [23] beginning at about age 60. Over the full age range, the Pearson’s correlation between brain volume and age is  $-0.295$ ,  $df = 617$ ,  $p < 10^{-18}$ .

### 3.4. MEG Processing

The MEG channels were each filtered using MNE tools with high and low pass at 10 and 250 Hz, 5 Hz roll-off [24]. A value of 250 Hz was used for the low pass to thoroughly remove the continuous head positioning signals present in the raw MEG at 293, 307, 314, and 321 Hz. Previous work has shown that the higher the low-pass frequency, the greater the yield of the solver [8,9]. Note that the 10 Hz high-pass filtering effectively demean each channel and removes much of the low-frequency content most commonly studied. A value of 10 Hz was used for the high pass because (a) the solver yield significantly increases with the low frequencies removed and (b) the solver was set to search one 80 msec data segment at a time. Data lengths greater than this reduce solver yields [8,9], presumably because current dipole orientation rarely remains stable for that long. This short data length provides very low sensitivity to frequencies below 12 Hz. However, the solver was stepped through the data in 40 msec increments; hence, a bolus of identified current dipoles was identified at 25 Hz. Analysis of the time course of those boluses can provide analysis of low frequencies; however, this is outside the scope of the present study.

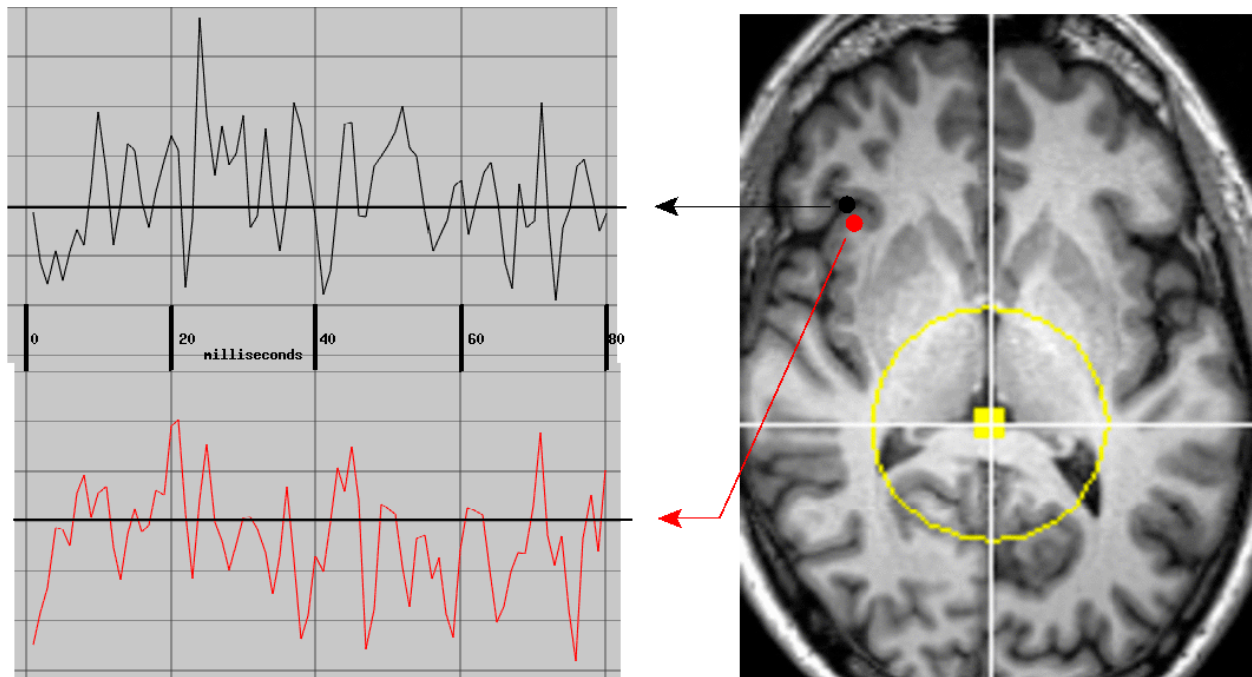
For each 1.24 s data segment, mains noise was removed from the CamCAN data at 50, 100, 150, 200, and 250 Hz using polynomial synchronous noise removal [25]. Mains noise was removed at 60, 120, 180, and 240 Hz from the TEAM-TBI data. No other preprocessing was applied and no data segments were excluded by manual artifact identification.

The subject’s head position within the MEG scanner was manually coregistered to the TI scan using Elekta’s MriLab visualization tool. The coordinates of the center point of a sphere most nearly approximating the brain were identified. These are the only operations in the data processing pipeline for which human judgement was applied. All other operations were fully automated.

Continuous head positioning measures were extracted using Elekta’s MaxFilter tool [26]. The coregistration of the MEG sensor array with the location of the subject’s head and brain was corrected once per second using the continuous head positioning information. This correction was applied to the forward solution used by the solver. The referee consensus solver is described in detail elsewhere. [8,9,27].

The forward solution is the mathematical relationship between a putative electric current within the brain and the resultant magnetic field measurements at the sensor array.

The solution we used models the brain as a uniformly conducting sphere [3]. Currents within 30 mm of the center of the sphere are nearly undetectable and the mathematical formulation for the forward solution behaves poorly for this volume; hence, it was excluded from the search. The intersection of this region with an MRI slice is shown in Figure 8. Note that the excluded volume typically includes the posterior thalamus, the posterior commissure, and much of the midbrain (not shown in the figure). The solver's search volume was delimited using the automated brain segmentation provided by Freesurfer with the 30 mm sphere at the center excluded.



**Figure 8.** Two typical simultaneously active neuroelectric currents were identified and validated by the referee consensus solver,  $p < 10^{-12}$  for each—i.e.,  $p < 10^{-4}$  for each when corrected for multiple comparisons. Each waveform has a duration of 80 msec sampled at 1000 Hz. The bandpass is 10–250 Hz. The currents are 5 mm apart with zero-lag cross-correlation of 0.157,  $df = 80$ ,  $p = 0.16$ . The yellow dot and circle delineate the region near the center of the head which is excluded from the search for neuroelectric currents. See text for details.

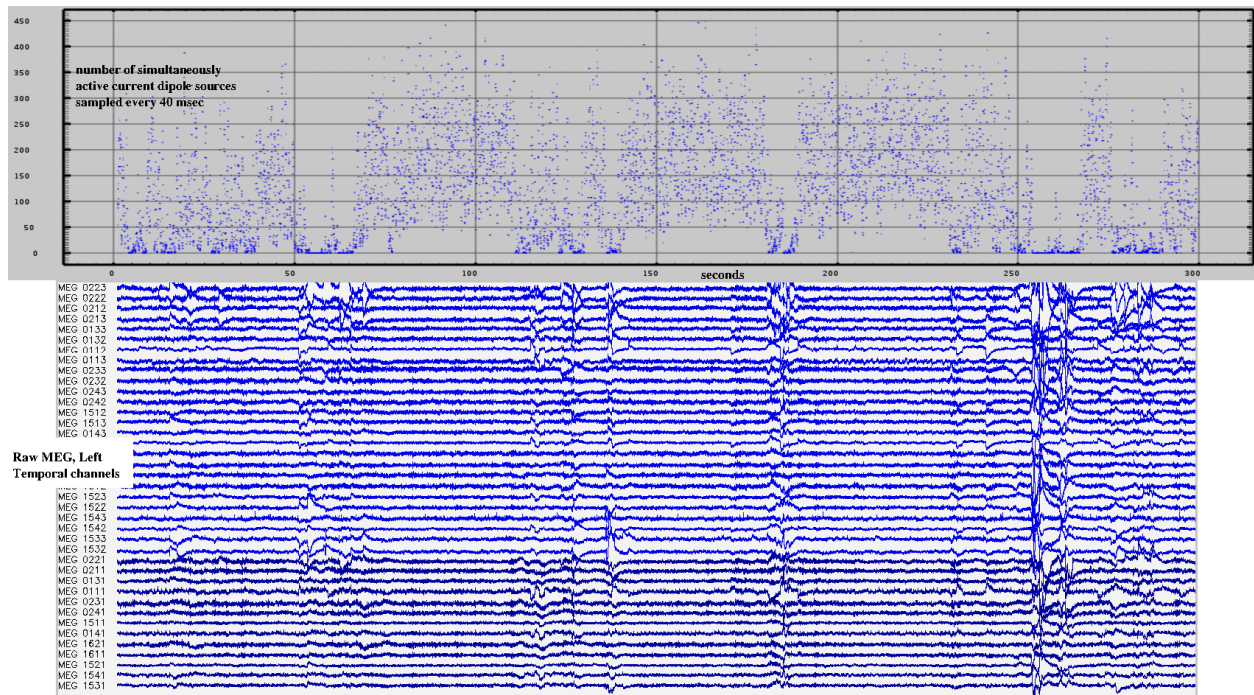
The solver was deployed on The Open Science Grid (OSG), an international distributed supercomputing partnership for data-intensive research [28,29]. The work described here utilized more than 70,000,000 processor-hours on the OSG. The solver is detailed in [8,9] and in Appendices B and C.

When applied to continuous MEG recordings, the solver typically identifies and validates more than 400 neuroelectric currents within the brain per 40 msec step through the data stream— $p < 10^{-12}$  for each and  $p < 10^{-4}$  for each when conservatively corrected for multiple comparisons (Bonferroni). This is more than 600,000 currents per minute of recorded MEG data identified with millimeter and millisecond resolutions. Note that data segments contaminated by movement or other artifacts were not manually identified for removal. Instead, artifact rejection relied upon the referee consensus solver's inherent failure to validate neuroelectric currents when presented with noisy data, as shown in Figure 9 [27].

The validated currents within each of the 171 automatically identified brain regions were counted over the duration of the recording. Each count was normalized to current density,  $\rho_{\text{roi}}$ :

$$\rho_{\text{roi}} = (\text{count}_{\text{region}} / \text{count}_{\text{total}}) \div (\text{vol}_{\text{region}} / \text{vol}_{\text{total}}) \quad (1)$$

The purpose of this normalization is to enable comparisons of a region within or between individuals, during different states, at different times, or comparisons of one region with another. The normalization is defined so that  $\rho_{\text{region}} = 1.0$  for all regions if the neuroelectric currents are uniformly distributed throughout the brain. In the isotropic case, (a) the regional count fraction is always equal to the regional volume fraction and (b) no difference is found for any comparison. Dividing the counts for a region by the total count normalizes  $\rho$  for variations due to both data quality and record length. The normalization for data quality is important since the yield of the solver changes from moment to moment as data quality waxes and wanes, e.g., Figure 9 [8,27].



**Figure 9.** The referee consensus solver automatically fails when the recordings are noisy. 300 s of raw magnetoencephalographic (MEG) (**lower**) and neuroelectric currents (**upper**) are shown. The number of validated ( $p < 10^{-12}$ ) currents identified drops markedly when the MEG is noisy.

#### 4. Normative Measures

Normative values for regional measures of static neuroelectric activity were established. To accomplish this, the  $mean_{\rho}$  and  $standard\ deviation_{\rho}$  for each regional current density ( $\rho$ ) were obtained from the CamCAN recordings.  $\rho$  for any region for any individual may then be compared with the norm for that region by converting it to a  $z$ -score with a corresponding  $p$ -value under the assumption that the current densities for the normative population is normally distributed.

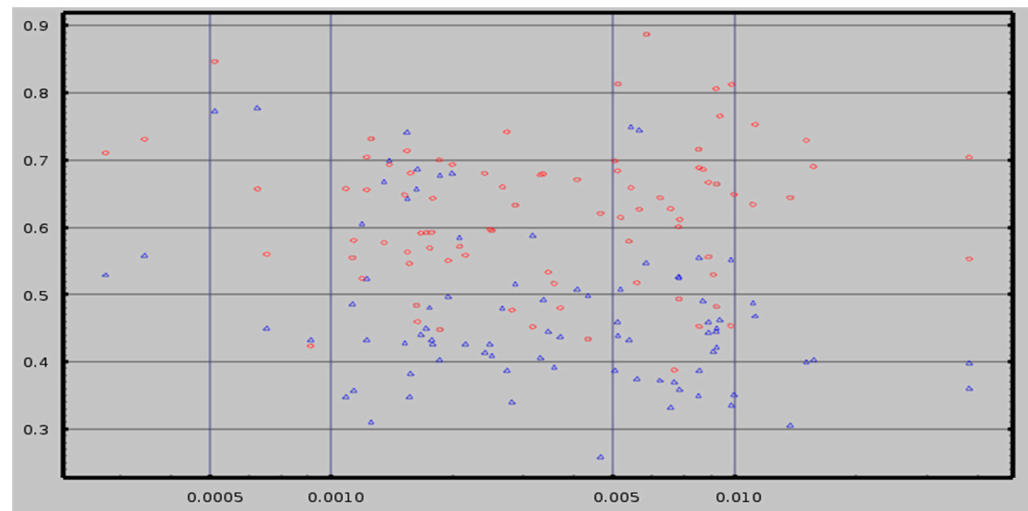
$$z\text{-score}_{\rho} = (\rho - mean_{\rho}) \div (sd_{\rho}) \quad (2)$$

The tables of CamCAN *means* and *standard deviations* are presented in Appendix A. They constitute an atlas which may be used to transform the current densities from any individual to  $z$ -scores and then assess the normality of deep white matter tonic neuroelectric traffic and cortical/subcortical tonic neuroelectric activity.

Note that transformation of the density measures to  $z$ -scores nominally equalizes the variances of the norms for all of the regions. This ensures that for a composite measure, e.g., a linear classifier composed of the 18 tract  $z$ -scores, the impact of each of the 18 densities is approximately equal.

### Empty Room Correction

An ideal method for extracting neuroelectric measures from MEG recordings would consistently yield a value of zero from empty room recordings. The referee consensus solver falls short of this ideal—i.e., there is a significant “dark count”. In addition, the neuroelectric measures extracted from empty room recordings consistently demonstrate significant correlations to measures extracted from human resting recordings obtained on the same day (Figure 10). Each correlation is computed across all of the subjects of one of the cohorts. They are plotted on the y-axis in the figure in blue for the CamCAN cohort ( $n = 619$ ) and in red for the TEAM-TBI cohort ( $n = 63$ ).



**Figure 10.** The correlations between measures for empty room and human resting for 85 cortical and subcortical regions for the CamCAN (blue,  $n = 619$ ) and the TEAM-TBI (red,  $n = 63$ ) cohorts are plotted on the y-axis. The number of events found within each region as a fraction of the total number of events found is plotted on the x-axis; the x-axis is logarithmic. The correlations are greater for TEAM-TBI than CamCAN as expected given the large difference in the  $n$  values. See text for details.

Each  $\rho_{\text{region}}$  was adjusted to compensate for the contribution of the empty-room “dark count”—i.e., the presumed contribution of falsely validated currents. We define the following:

$\rho_{\text{region-empty}}$  is the result of the  $\rho_{\text{region}}$  calculation applied to the empty-room data.

$\text{corr}_{\text{region}}$  is the correlation across the CamCAN subjects between  $\rho_{\text{region}}$  and  $\rho_{\text{region-empty}}$ —i.e., between the regional activity measured with the subject present and absent in the scanner. Then

$$\rho_{\text{region-corrected}} = \rho_{\text{region}} - (\text{corr}_{\text{region}} \times \rho_{\text{region-empty}}) \quad (3)$$

This correction is an approximate one which uses the correlations as estimates of detecting the competitive advantages that true vs. false neuroelectric currents have. A likely more accurate approach would use a classifier to decide on inclusion or exclusion of one current at a time. For each subject by region, there are typically thousands of identified currents; hence, there is considerable statistical power to accurately train such a classifier. Pursuing this approach is beyond the scope of the work reported here.

### 5. Classification

Regional measures of neuroelectric activity for 17 subcortical and 68 cortical regions were combined into classifiers using stepwise linear classification [16–18]. This is an automated computer algorithm which performs discriminant analysis between two groups by computing a linear classification function in a stepwise manner. The groupings for classification were determined by symptom survey scores to test for sensitivity of the

measures to symptoms and by cohort membership to test for differences between the cohorts. These are detailed in the Results section.

## 6. Discussion

MEG-derived regional brain measures of tonic neuroelectric activation were tested for long-term test-retest reliability in a large normative cohort, CamCAN, and for sensitivity to symptoms in a chronic TBI cohort, TEAM-TBI. The studied symptoms were insomnia, depression, anxiety, and somatization. Good test-retest reliability was found as well as sensitivity to all four symptoms. Hence, the measures reported here may prove of significant clinical utility in the diagnosis and treatment of these symptoms. In addition, the measures enable classification of each individual into her/his cohort—i.e., normative vs. chronic TBI. Hence, the measures may prove useful as biomarkers for TBI.

The analysis and all results were obtained “by region”. Since we are seeking measures with good test-retest reliability and which can be compared between subjects. The volumetric units we used are regions, i.e., volumes, which can be reliably identified by automated algorithms because they are common to the anatomically normal human brain. As more detailed atlases with finer structures are developed, the measures reported here will be tested for those volumes. For the present, the volumes to which the measures reported here apply are the regions identifiable with *Freesurfer 5.3* [10,11] and *tracula 1.22.2.12* [22].

Each regional value which demonstrates long-term test-retest reliability is a measure of regional neuroelectric tonus—i.e., the static level of regional neuroelectric activation. Elevated or reduced regional tonus within an individual may prove emblematic of tonic alterations in network function. The ability to assess many such regional measures simultaneously may provide substantive useful information which is complementary to the measures which have specificity to TBI—e.g., bloodborne markers [30,31] and MEG-derived slow waves [32–34]. Patterns of altered regional tonus may prove useful in monitoring response to treatment. Analysis of the patterns may enable identification of regions to target for treatment. In particular, the several-centimeter localization of the measures is comparable to the localization precision of transcranial magnetic stimulation (TMS) [35–38]. The deviations seen in a particular individual may prove sufficient to identify individualized target regions for TMS [39,40] rather than the standardized left and/or right prefrontal cortex currently in use [41,42].

### 6.1. Potential Clinical Utility

This study was undertaken to utilize and assess MEG-derived measures for the diagnosis and monitoring of treatment for chronic symptoms of TBI. We report results which demonstrate (a) sensitivity to the presence/absence of insomnia, somatization, depression, and anxiety (Tables 1 and 3, Figure 1) and (b) sensitivity to history of concussion and/or chronic symptoms (Table 5, Figures 2–5). We cannot directly tie these MEG results to TBI. However, for clinical purposes, the etiology may not matter so long as we can use the measures to more effectively diagnose and treat.

The symptomatic identification accuracies shown in Tables 1 and 3 are reliably significant, but the percentages are not yet high enough for this classification method to be useful clinically. It is likely that classification accuracy can be increased by (a) refining the measures of neuroelectric activity and by (b) using nonlinear or machine-learning classification methods,

The primary results of the study combine the information contained in many regional neuroelectric measures into patterns of brain activity which are related to chronic symptoms in chronic TBI. We also report cohort-wide differences in regional activity (Figures 4 and 5). These are findings which suggest ways to study the mechanisms which underlie presentation and recovery from symptoms. Productive scientific use of these findings may be complemented by a working theoretical conjecture. To this end, we propose a phantom pain conjecture: all symptoms of psychological distress result from hyperactivity in brain regions responsible for attention and response to pain. In support of this conjecture, many



regions which show fMRI-derived differential activation in response to painful stimuli [43] (Table 1) show differential activation in MEG-derived measures in the TEAM-TBI cohort when compared with the CamCAN cohort (Table 6)—e.g., thalamus, precentral, postcentral, and precuneus cortexes, and portions of the cingulate, orbito-frontal, and insular cortexes.

**Table 6.** The regions whose neuroelectric activity values contributed to the cohort-specific classifiers reported in Table 5 are shown. Activity in regions marked “+” was higher in the Team-TBI cohort; those marked “-” were lower. Regions marked “++” or “-” made the largest statistical contribution to the classifier. The results for the 2nd step are highlighted in gray. These were obtained when the regions which were selected for the first classification analysis were excluded.

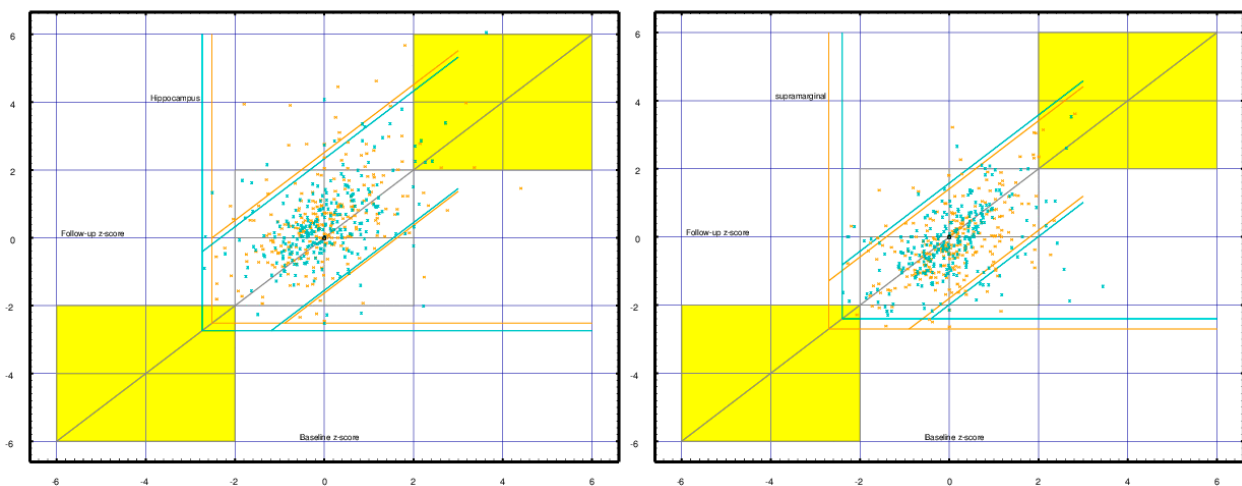
Brain-Stem	++	
	Left	Right
Accumbens	-	
Cerebellum	+	++
Caudate	-	
Hippocampus		-
Pallidum		+
Thalamus	-	-
fusiform	++	+
inferior parietal	+	
insula	-	
isthmus cingulate	++	
lateral occipital	-	-
lingual	++	
medial orbito-frontal	-	
middle temporal	-	-
para-hippocampal	-	
pars opercularis	-	-
pars triangularis	+	
peri-calcarine		-
postcentral	-	-
precuneus	+	
precentral	-	
superior parietal		-
superior temporal		-
transverse temporal		-

## 6.2. Test-Retest Reliability

We report short-term (1 h,  $n = 619$ ) and long-term (*mean* 16-months,  $n = 253$ ) test-retest reliability for the CamCAN normative cohort for each of the 103 brain regions. The shorter average follow-up for the TEAM-TBI cohort of 6.4 months would be expected to have better test-retest reliability in a normative population. This strengthens confidence in the primary findings of the study that MEG-derived neuroelectric measures change in those TEAM-TBI cohort members whose symptom scores changed.

We used Pearson’s correlation and *mean* difference in test-retest values—Tables A1–A3 and Figure 6. The difference measure may be used to correct a follow-up measure to compare with baseline.

Short-term test-retest reliability ranged around a *mean* correlation of 0.8 with *mean* average difference well under the  $z$ -score = 0.03. Long-term test-retest reliability ranged around a *mean* correlation of 0.45, with the *mean* average difference as high as  $|z\text{-score}| = 0.55$ . This is visualized for the hippocampus and supramarginal cortex in Figure 11. These regions were selected because their baseline vs. 16-month follow-up correlations are typical and their difference values are significantly different from zero and visibly so in the figure.



**Figure 11.** Baseline (x-axis) vs. 16-month follow-up (y-axis) z-score regional current densities are plotted for each CamCAN subject ( $n = 253$ )—right brain in green, left brain in orange. The **left** panel shows the hippocampus. Correlations are 0.456 (L) and 0.476 (R); differences are 0.524 (L) and 0.408 (R). The **right** panel shows the supramarginal cortex. Correlations are 0.524 (L) and 0.482 (R); differences are  $-0.264$  (L) and  $-0.194$  (R). The significance of the differences is readily seen in the displacement of the upward (hippocampus) and downward (supramarginal) distributions. The horizontal and vertical green/orange lines indicate the right/left floor values for the measures. The diagonal lines are 2.0 z-scores from the mean difference. Yellow highlights the areas for which both baseline and follow-up z-scores are  $\geq 2.0$ .

A survey of recent test–retest reliability reports shows reliability ranging widely. As would be expected, test–retest reliability is more consistent for time-locked measures in evoked response paradigms [44–51]. For free-running paradigms, most measures are for resting connectivity confined to frequency bands [52–58]. The test–retest intervals in most studies are days to weeks. Recasens et al. [47] report results from a 7-week interval and both Piitulainen et al. [57] and Dunkley et al. [48] report results for intervals greater than one year. Over all the reports, the maximum number of subjects was 40 [46]. Cohorts with clinical diagnoses were reported by Candelaria-Cook et al. [55] (psychosis) and Dunkley et al. [48] (PTSD).

Both short-term and long-term reliability values we report compare favorably with all others. For the work reported here (a) the  $n$  values are much larger, (b) the long-term interval is 16 months, (c) the measures are free-running rather than synchronized to an event, and (d) the measures are from raw rather than averaged data.

### 6.3. Cortical vs. Adjacent White Matter Regions

We report considerable detectable neuroelectric activity from the white matter with positive differentials in favor of adjacent white matter for 56% of those pairs for which the differential is significant with the threshold  $p < 10^{-8}$ . Both previously reported measurements and neurophysiological understanding speak to the plausible validity of these findings. MEG-derived responses from thalamocortical fibers have been reported [59,60]. The source magnetic fields were presumed due to synchronous volleys of action potentials (APs). Each AP produces a travelling current quadrupole. The approximate amplitude was estimated at  $100 \text{ Amp}^{-15} \text{ m}$  in an unmyelinated axon [7], with separation of 1 mm between the two dipoles forming the quadrupole, assuming a propagation velocity of 1 m/second. It is presumed that the velocity is greater in the myelinated fibers which comprise the white matter. Hence, the velocity and dipole separation would be greater. This would decrease the distance-dependence of the magnetic field strength and so enhance the detectability of this activity. In addition, the magnetic field, due to an action potential in a single axon, was directly measured at about  $150 \times 10^{-12} \text{ Tesla}$  [60].

A trivial explanation of the profuse findings we report is that cortical activity is localized in nearby white matter due either to poor resolution or to head movements. The robust finding of differential activity between adjacent cortical and white matter ROIs argues

against this. As does the design of the method which relies on gradients between points within the brain that are 1 mm apart [8,9] (Appendix B), coupled with the use of one correction per second to the forward solution using continuous head positioning information.

Under the assumption that the white matter is in fact the source of profuse measurable neuroelectric activity, the measured magnetic field components can only be due to synchronous volleys of APs. These would produce transient longitudinal intra-axonal currents which are nearly synchronous in many parallel running axons due to near simultaneous passage of propagating APs.

The detected magnetic field waveforms, e.g., Figure 8, are envelopes which follow the high-frequency waveforms of several AP volleys in sequence. The envelope of a single highly synchronized AP volley would require well under 10 msec to rise and fall. Hence, this type of activity would be dominated by high-frequency contents. This is consistent with the observation that the yield of the solver improves when the low-pass cut-off with which the signals are preprocessed is increased from 150 to 330 Hz [8] (Figure 3). It is also consistent with the observation from a typical task recording that the frequency content of the current waveforms includes profuse resonant activity with frequency content above 70 Hz [61,62].

Additional work outside the scope of the present study is needed to understand the mechanisms which underlie the detection of profuse activity localized to the white matter.

#### 6.4. *CamCAN vs. TEAM-TBI Differences*

The robust differences seen between the two cohorts must be interpreted with caution. We cannot rule out the possibility that these differences are due to differences in the MEG scanners, the scanner noise environments, or for the white matter results to the differences in DWI scan parameters. Given the robustness of the differences, this question can be answered by running a cohort of 40 neurologically normal individuals in the scanner used for the TEAM-TBI cohort. The classifiers developed for the present study can be applied to the measures from such a control cohort. If they are different from the CamCAN cohort, then the differences between the CamCAN and TEAM-TBI cohorts must be presumed to be due to differences in the scanners.

A second potential confounding factor of these results is that the CamCAN resting recordings were obtained with eyes closed, whereas the Team-TBI recordings were obtained with eyes open. To test this, consider that the short test-retest reliability results were obtained by comparing baseline CamCAN resting (eyes closed) with CamCAN sensorimotor task (same sitting, eyes open). The test-retest reliability is very good, i.e., the differences between resting and task are very small. In addition, linear classifiers fail to distinguish between baseline CamCAN resting and task recordings. Hence, this difference in recording conditions, i.e., eyes closed vs. eyes open, does not account for the differences between the cohorts.

It is noteworthy that differences found between TEAM-TBI cohort members with and without specific symptoms is not affected by these questions. The same applies to the test-retest reliability results and to the comparisons between cortical and adjacent white matter volumes. Only the cause of the differences between the cohorts is in question.

Under the assumption that the differences found between cohorts are due to differences from the norm in the neuroelectric brain activity of those with TBI, correspondences between the normal vs. TBI classification results we report and those reported by others may be useful. The high classification accuracy found between CamCAN and TEAM-TBI cohorts, i.e., greater than 90%, provides confidence in the validity of the regions whose measures contribute most to the classifier (Table 6). Of the frontal regions, only the left medial-orbito frontal was a contributor, so significant correspondence was found in our findings to the bilateral orbito-frontal localization of high-amplitude slow waves reported by Lewine et al., Huang et al. and others [32–34].

A striking correspondence was also found with a recent study of MEG-derived functional connectivity in TBI [63]. The results of this study highlight the importance of changes

in thalamic function in TBI. Table 6 shows that both left and right thalamic measures are major contributors to the classifier and that tonic thalamic activation in the TEAM-TBI cohort members is bilaterally reduced. At present, we can only speculate what the neuro-physiologic mechanisms that tie altered regional activation, functional connectivity, and symptoms together are. The growth in our ability to reliably measure such alterations and to target specific regions with drug and TMS therapies may enable us to understand these mechanisms and to more effectively treat them.

**Author Contributions:** Conceptualization, D.K., A.M.P., S.B., W.S., A.P.K., D.O.O.; methodology, ALL.; software, D.K., P.S.; formal analysis, D.K., P.S., S.B., A.P.K., D.O.O.; investigation, D.K., P.S., R.S., A.M.P., S.B., W.S., A.P.K., M.W.C., D.O.O.; resources, D.K., S.B., W.S., A.P.K., M.W.C., D.O.O.; data curation, D.K., A.M.P., S.B., W.S., A.P.K., D.O.O.; writing—original draft preparation, D.K., P.S., D.O.O.; writing—review and editing, D.K., P.S., R.S., A.M.P., S.B., W.S., A.P.K., M.W.C., D.O.O.; supervision, A.M.P., S.B., W.S., A.P.K., M.W.C., D.O.O.; project administration, D.K., A.M.P., A.P.K., D.O.O. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the US Department of Defense (USAMRAA W81XWH-14-2-0002), by the Extreme Science and Engineering Development Environment (XSEDE IBN130001), and Texas Advanced Computing Center (TACC TG-IBN130001).

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Institutional Review Board of The University of Pittsburgh (PRO13070121).

**Informed Consent Statement:** Participants' approvals for participation were obtained using written informed consent approved by the University of Pittsburgh IRB (PRO13070121).

**Data Availability Statement:** The CamCAN datasets are available on request from the Cambridge Centre for Ageing and Neuroscience (CamCAN) at the University of Cambridge, UK, <https://www.cam-can.org/>, (accessed on 22 March 2021). The TEAM-TBI datasets are available on request from the Federal Interagency Traumatic Brain Injury Research (FITBIR) Informatics System at the National Institutes of Health, USA, <https://fitbir.nih.gov/> (accessed on 22 March 2021). The collection of currents identified by the referee consensus solver ( $\approx 5$  TBytes) are available on request from the corresponding author—kriegerd@upmc.edu.

**Acknowledgments:** We gratefully acknowledge the invaluable contributions made to this effort by the Cambridge (UK) Centre for Ageing and Neuroscience, the Extreme Science and Engineering Development Environment (XSEDE), The Texas Advanced Computing Center (TACC), the Open Science Grid (OSG), the San Diego Supercomputing Center, the Pittsburgh Supercomputing Center, the XSEDE Neuroscience Gateway, Darren Price, Ethan Knights, Mats Rynge, Rob Gardner, Frank Wurthwein, Derek Simmel, Mahidhar Tatineni, and Ali Marie Shields. Data used in the preparation of this work were obtained from the CamCAN repository [10,11]. The OSG [17,18] is supported by the National Science Foundation, 1148698, and the US Department of Energy's Office of Science.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A. Atlas of Normative Regional Neuroelectric Measures

Regional measures of neuroelectric activity for 103 cortical, subcortical, and deep white matter regions were extracted from both the baseline ( $n = 619$ ) and 16-month follow-up ( $n = 253$ ) CamCAN recordings. The *means* and *standard deviations* for the baseline current density measures are listed in the left-most four columns of Table A1 (18 deep white matter tracts), Table A2 (17 subcortical regions), and Table A3 (68 cortical regions).

These density measures were transformed into *z-scores*; the *correlations* and *mean differences* between the baseline and follow-up *z-scores* are listed in the right-most four columns of the tables.

**Table A1.** Deep white matter tract traffic norms. Left columns—Atlas. The column labelled *mean* contains the normalized current density for the tract. *s.d.* == *standard deviation*. These were obtained from the CamCAN resting MEG recordings. A value greater/less than 1.0 indicates that normal tonic neuroelectric activity level for the tract is greater/less than that for the total volume of the brain. The activity measured within a tract for an individual is converted to a z-score using the corresponding *mean* and *standard deviation* listed in the table. Note that 589 of 619 CamCAN subjects had DWI scans; 240 returned for follow-up studies. Right columns—good test-retest reliability. The *mean* baseline to follow-up interval was 16.0 months. For each tract, the cohort-wide follow-up vs. baseline correlation (*Corr*) and difference (*Diff*) are shown. *Diff* is the *mean z-score* difference: follow-up–baseline. The correlations are all high and the differences are all low. Moderate significance (i.e.,  $0.03 > p > 0.002$ ;  $-3.0 < t < 3.0$ ) is indicated with *italics*; high significance (i.e.,  $p < 0.002$ ;  $t \geq 3.0$ ) is indicated in bold.

Deep White Matter Tracts CamCAN Baseline (n = 589)				Baseline vs. Follow-up (n = 240) Correlation and Mean Difference			
<i>mean</i>		<i>s.d.</i>		<i>corr</i>	<i>diff</i>		
0.319		0.168		0.255	-0.055		
0.561		0.234		0.345	-0.102		
Left		Right		Left		Right	
<i>mean</i>	<i>s.d.</i>	<i>mean</i>	<i>s.d.</i>	<i>corr</i>	<i>diff</i>	<i>corr</i>	<i>diff</i>
0.924	0.275	0.828	0.278	0.427	-0.001	0.348	0.050
0.914	0.304	1.039	0.325	0.468	0.044	0.507	-0.085
0.477	0.245	0.511	0.259	0.366	-0.059	0.435	-0.021
1.004	0.227	0.878	0.217	0.471	0.041	0.369	-0.052
0.958	0.254	0.796	0.253	0.589	-0.012	0.532	-0.030
0.826	0.260	0.856	0.254	0.515	-0.003	0.576	0.008
0.980	0.197	1.020	0.236	0.497	0.006	0.408	-0.073
1.087	0.288	0.947	0.288	0.523	-0.000	0.524	-0.046

**Table A2.** Subcortical regional activity norms. These norms and test-retest reliability measures were obtained as described in Table A1.

Subcortical Regions CamCAN Baseline (n = 619)				Baseline vs. Follow-up (n = 253) Correlation and Mean Difference			
<i>mean</i>		<i>s.d.</i>		<i>corr</i>	<i>diff</i>		
0.713		0.203		0.434	-0.510		
Left		Right		Left		Right	
<i>mean</i>	<i>s.d.</i>	<i>mean</i>	<i>s.d.</i>	<i>corr</i>	<i>diff</i>	<i>corr</i>	<i>diff</i>
0.746	0.658	0.635	0.566	0.388	0.157	0.349	0.136
1.057	0.483	0.921	0.476	0.523	<b>0.474</b>	0.560	<b>0.513</b>
0.569	0.347	0.580	0.330	0.488	0.022	0.609	-0.050
0.666	0.184	0.697	0.187	0.441	<b>0.207</b>	0.546	0.004
0.876	0.348	0.985	0.360	0.456	<b>0.524</b>	0.476	<b>0.408</b>
0.957	0.552	0.715	0.552	0.488	<b>0.287</b>	0.523	<b>0.333</b>
0.944	0.409	1.062	0.429	0.509	<b>0.194</b>	0.656	<b>0.287</b>
0.622	0.306	0.730	0.328	0.520	0.146	0.549	-0.079

Left panel—Atlas. Table (right)—good test-retest reliability. As for Tables A1 and A3, *Diff* is the *mean z-score* difference: follow-up–baseline; *mean* interval is 16 months. Moderate significance (i.e.,  $0.03 > p > 0.002$ ;  $-3.0 < t < 3.0$ ) is indicated with *italics*; high significance (i.e.,  $p < 0.002$ ;  $t \geq 3.0$ ) is indicated in bold.



**Table A3.** Cortical regional activity norms. These norms and test-retest reliability measures were obtained as described in Table A1.

Cortical Regions CamCAN Baseline (n = 619)					Baseline vs. Follow-up (n = 253) Correlation and Mean Difference			
Left		Right			Left		Right	
<i>mean</i>	<i>s.d.</i>	<i>mean</i>	<i>s.d.</i>		<i>corr</i>	<i>diff</i>	<i>corr</i>	<i>diff</i>
0.961	0.415	0.903	0.442	bankssts	0.538	0.203	0.523	0.091
0.565	0.426	0.547	0.156	caudalanteriorcingulate	0.252	−0.142	0.377	−0.200
0.729	0.408	0.848	0.448	caudalmiddlefrontal	0.434	<b>0.210</b>	0.459	<b>0.231</b>
0.276	0.276	0.287	0.275	cuneus	0.253	−0.134	0.355	−0.042
1.015	0.405	0.954	0.419	entorhinal	0.393	<b>0.422</b>	0.532	<b>0.345</b>
0.124	0.310	0.117	0.313	frontalpole	0.567	− <b>0.169</b>	0.418	−0.137
0.578	0.229	0.611	0.243	fusiform	0.419	<b>0.395</b>	0.420	<b>0.331</b>
0.661	0.268	0.553	0.234	inferiorparietal	0.426	−0.068	0.530	− <b>0.284</b>
1.041	0.277	0.988	0.284	inferiortemporal	0.520	<b>0.556</b>	0.521	<b>0.471</b>
0.697	0.267	0.652	0.274	insula	0.581	<b>0.213</b>	0.508	<b>0.428</b>
0.227	0.286	0.263	0.298	isthmuscingulate	0.318	−0.002	0.337	−0.112
0.889	0.263	0.855	0.261	lateraloccipital	0.577	− <b>0.432</b>	0.472	<b>0.369</b>
0.952	0.300	0.907	0.320	lateralorbitofrontal	0.478	<b>0.517</b>	0.471	<b>0.386</b>
0.470	0.260	0.497	0.253	lingual	0.473	−0.020	0.520	− <b>0.284</b>
0.517	0.263	0.537	0.269	medialorbitofrontal	0.337	−0.140	0.384	− <b>0.531</b>
0.716	0.256	0.862	0.252	middletemporal	0.4675	<b>0.546</b>	0.601	<b>0.471</b>
0.335	0.290	0.334	0.274	paracentral	0.317	−0.146	0.422	−0.123
0.886	0.378	0.909	0.386	parahippocampal	0.410	<b>0.394</b>	0.372	<b>0.301</b>
0.845	0.355	0.843	0.389	parsopercularis	0.431	−0.161	0.354	−0.112
0.506	0.362	0.462	0.335	parsorbitalis	0.525	0.152	0.342	0.104
0.803	0.362	0.882	0.381	parstriangularis	0.443	−0.033	0.415	−0.038
0.352	0.297	0.359	0.302	pericalcarine	0.490	0.109	0.292	0.036
0.784	0.276	0.712	0.264	postcentral	0.466	−0.060	0.399	0.019
0.382	0.356	0.403	0.324	posteriorcingulate	0.316	−0.102	0.319	− <b>0.290</b>
0.793	0.281	0.777	0.287	precentral	0.459	0.092	0.480	<b>0.198</b>
0.150	0.154	0.160	0.165	precuneus	0.205	− <b>0.305</b>	0.365	− <b>0.248</b>
0.730	0.438	0.701	0.419	rostralanteriorcingulate	0.278	− <b>0.401</b>	0.280	− <b>0.540</b>
0.560	0.276	0.559	0.296	rostralmiddlefrontal	0.450	0.081	0.394	0.102
0.530	0.258	0.512	0.262	superiorfrontal	0.442	0.005	0.457	−0.143
0.618	0.273	0.612	0.289	superiorparietal	0.442	−0.105	0.447	−0.071
0.892	0.232	0.902	0.255	superiortemporal	0.468	<b>0.596</b>	0.513	<b>0.430</b>
0.768	0.284	0.731	0.304	supramarginal	0.524	− <b>0.264</b>	0.482	− <b>0.194</b>
0.642	0.387	0.856	0.403	temporalpole	0.351	<b>0.512</b>	0.438	<b>0.555</b>
0.808	0.450	0.832	0.503	transversetemporal	0.521	<b>0.267</b>	0.293	<b>0.280</b>

Left Panel—Atlas. Table (right)—good test-retest reliability. As for Tables A1 and A3, *Diff* is the mean z-score difference: follow-up–baseline; mean interval is 16 months. Moderate significance (i.e.,  $0.03 > p > 0.002$ ;  $-3.0 < t < 3.0$ ) is indicated with *italics*; high significance (i.e.,  $p < 0.002$ ;  $t \geq 3.0$ ) is indicated in bold.

## Appendix B. The Referee Consensus Method

The 2-fold task of the solver is to (1) provide a robust measure of confidence that a dipole current is detected at location X and (2) estimate the time course of the current amplitude.

The solver was applied to one 80 msec data segment ( $M_{t=1, \dots, 80}$ ) at a time. A decision was made for one location at a time—for example: “Is there a dipole current present at location X?”. To answer this question, spatial filters were constructed from the “viewpoints” of each of the 90 distant “referee” locations distributed widely through the volume of the brain—e.g., R.

Filter  $P_{R|X'}$  was constructed with gain 1.0 at R and gain 0.0 at X' 1 mm from X.  $P_{R|X'}$  was applied to the 80 data vectors,  $M_{t=1, \dots, 80}$ , to produce the 80-point univariate time series,  $V_{R|X'}$ . A second filter was constructed,  $P_{R|X}$ , with gain 1.0 at R and gain 0.0 at X.  $P_{R|X}$  was also applied to  $M_{t=1, \dots, 80}$  to produce the 80-point univariate time series,  $V_{R|X}$ . Note that there was a small contribution to  $V_{R|X'}$  from activity at X but none from X'. Contrariwise, there was a small contribution to  $V_{R|X}$  from activity at X', but none from X. The difference filter was constructed,  $P_{R|X'-R|X}$ . This has gain 0.0 at R and nearly equal and opposite gains at X and X'.  $P_{R|X'-R|X}$  applied to  $M_{t=1, \dots, 80}$  produced  $V_{R|X'-R|X}$ , the difference— $V_{R|X'} - V_{R|X}$ . Note that there is no contribution to this from R. Note too that each of these three filters was constructed with gain 0.0 at each of 89 other “referee” locations coarsely covering the brain, so  $V_{R|X'-R|X}$  includes only small contributions from other neuroelectric currents. This ensures that the primary contributors to  $V_{R|X'} - V_{R|X}$  are currents close to X and/or X'.

The “opinion” from the viewpoint of referee R regards the presence of a current at X, which is obtained by evaluating this inequality:

$$(V_{R|X'-R|X} \bullet V_{R|X'})^2 > (V_{R|X'-R|X} \bullet V_{R|X})^2 \quad (A1)$$

If the inequality is true, then there is a current at X from the viewpoint of R since  $V_{R|X'}$  (left side) has no contribution from X',  $V_{R|X}$  (right side) has none from X, and  $V_{R|X'-R|X}$  has nearly equal contributions from both.

This procedure is repeated for two vector components for each of the 90 referee locations to produce 180 yes/no “opinions”. In total, 114 or more must be “yes” ( $p \ll 0.01$ ) to produce an acceptable “consensus” for this differential. The same procedure was repeated for each of the other five differentials since there were two differentials along each of the three spatial axes. Only if all six exceed the threshold, i.e., 57 or more of 90 for each of the 6, was a current accepted. Therefore,  $0.01^6 = 10^{-12}$  is the threshold for accepting a current.

Once a location was validated, an eigenvector analysis was used to identify the 80-point time course of the current at that location as the waveform which captures the most variance in the complete set of  $V_{R|X'-R|X}$  values. The estimated signal/noise enhancement of 10 provided by this operation is detailed in the introduction. Note that the validation ensures that there is a current present at X and not at any of the six Xs. Hence,  $V_{R|X'-R|X}$  was used because the primary contributor to all of the  $V_{R|X'-R|X}$  values was due to the current at X. Because the current at X must dominate any current present at any of the Xs, the ability of the method to identify two currents near each other was limited to twice the distance—X-X', i.e., 2 mm.

## Appendix C. : The Referee Consensus Method—Advantages

Akin to beamformers, the filters used in the referee consensus method were generated in sets, seven at a time. However, unlike standard filters which were optimized to yield source space measures at the target/test location, X, all seven referee consensus filters were optimized to yield source space measures at *referee* locations remote from the test location X.

Note that the estimated neuroelectric currents are two-dimensional vector quantities [4]. For the purpose of intelligibility, they were treated as scalars in the following explanation without loss of generalization.

For a particular referee location, R, all of the filters have gain 1.0 at R. There was one filter that was constrained to have zero gain at X—filter  $P_{RIX}$ . This filter, designated as “R not X”, was optimized to measure the signal at the referee location but with no contribution from activity at X. There are six other filters, each constrained to have zero gain at one of the six points 1.0 mm away from X along the x-, y-, or z-axis. Hence, these filters were also optimized to measure the signal at the referee location but with no contribution from one of the locations 1 mm away from X along one of the coordinate axes.

These seven filters were used to generate six difference filters—e.g.,  $P_{RIX'-RIX}$ , where  $X'$  is the location +1 mm away from X along the x-axis. Note that the difference filters were constrained to have zero gain at the referee location and near equal but opposite gains at X and the location 1 mm away. Hence, they were optimized to measure the difference between the signals at X and at a location 1 mm away. The magnitude of the gain of these filters at these two “differential” locations was typically 0.05–0.08 [64], (Figure 6). Note that the differencing delivers signal/noise enhancement of about 2.0 in measuring the differential between activity at X and activity 1 mm from X.

Conceptually, this approach to source space measurement is upside down. The filters for the location at which measurements are made do not have gain 1.0 at the test location, X, as is the standard, but rather have zero gain either at X or very near it. The power of this approach comes from the use of families of these filters to develop a consensus *decision* on the question: is there or is there not a neuroelectric current at the test location, X? For example, the filters  $P_{RIX'}$ ,  $P_{RIX}$ , and  $P_{RIX'-RIX}$  were each applied to an 80 msec MEG data segment to yield three 80 msec data traces,  $V_{RIX'}$ ,  $V_{RIX}$ , and  $V_{RIX'-RIX}$ . If  $V_{RIX'-RIX} \bullet V_{RIX'} > V_{RIX'-RIX} \bullet V_{RIX}$ , then there is a current at X from the “point of view” of referee R.

Note the number of MEG measures used to assess this inequality between two numbers—i.e., the two dot products. Each element of each of the Vs is the dot product of the corresponding filter with the 306 MEG measures for a single time point. Were both the filter weights and the MEG measures uncorrelated and normally distributed, the signal/noise enhancement due to the filtering operations would be approximately  $\sqrt{306} \approx 17$ . To account for the certain significant departure from both assumptions, we divided the 306 degrees of freedom by 10 as a nominal correction to this estimate. Hence, we estimated the signal/noise enhancement due to the filters as  $\sqrt{30} \approx 5$ .

For the decision dot product, i.e., for evaluation of  $V_{RIX'-RIX} \bullet V_{RIX'} > V_{RIX'-RIX} \bullet V_{RIX}$ , each V was composed of 80 such filtered MEG measures. Hence, the numbers  $V_{RIX'-RIX} \bullet V_{RIX'}$  and  $V_{RIX'-RIX} \bullet V_{RIX}$  are each obtained using  $306 \times 80 = 24,480$  degrees of freedom. By the same reasoning as above, with the same factor of 10 for nominal correction, the estimated signal/noise enhancement due to the filters and the use of the 80-point times series is  $\sqrt{2448} \approx 50$ . Note that this estimate ignores the factor of 2.0 signal/noise enhancement inherent in the difference filter. Note too that this high-signal/-noise enhancement applies only to the numbers used in evaluating the inequality, not to the individual current amplitude measures—i.e., the elements of the Vs. This is why the emphasis in this paper is on the direct results of those decisions—i.e., the counts of validated neuroelectric currents.

There are 1080 binary decisions computed for a test location, X. One for each of the two orthogonal orientations for each 90 *referee* location and one for each of the six Xs. The collection of  $2 \times 90 \times 6 = 1080$  *decisions* provides a cost function that is used to assess the *consensus*: is there or is there not a neuroelectric current at X? This procedure is robust enough to use  $10^{-12}$  as the threshold *p*-value to accept X as a true source of a detectable magnetic field. If and only if this probabilistic threshold is met were the 1080 outputs of the difference filters, the  $V_{RIX'-RIX}$  values, combined by an eigenvector calculation, which is akin to averaging [Appendix B] to generate an estimate of the 80-point time course for the current.

Computation for each of these decisions generated its own optimal difference filter,  $P_{RIX'-RIX}$ , and corresponding optimal 80-point current time series estimate,  $V_{RIX'-RIX}$ . As detailed above, the signal/noise enhancement estimate for the elements of the Vs, i.e., the

current magnitude measures, are modest,  $\approx 5$ , and likely comparable to those for most filtering methods. However, the estimate obtained by this average-like operation over 1080 such difference filters again enhances the signal/noise by an additional factor of perhaps  $\sqrt{1080/10} \approx 10$ , producing total signal/noise enhancement for the current time series waveform estimate of about 50.

## References

1. Telkes, I.; Jimenez-Shahed, J.; Viswanathan, A.; Abosch, A.; Ince, N.F. Prediction of STN-DBS Electrode Implantation Track in Parkinson's Disease by Using Local Field Potentials. *Front. Neurosci.* **2016**, *10*, 198. [\[CrossRef\]](#)
2. Moroni, F.; Nobili, L.; Curcio, G.; De Carli, F.; Fratello, F.; Marzano, C.; De Gennaro, L.; Ferrillo, F.; Cossu, M.; Francione, S.; et al. Sleep in the Human Hippocampus: A Stereo-EEG Study. *PLoS ONE* **2007**, *2*, e867. [\[CrossRef\]](#)
3. Sarvas, J. Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem. *Phys. Med. Biol.* **1987**, *32*, 11–22. [\[CrossRef\]](#)
4. Cheyne, D.; Weinberg, H. Neuromagnetic fields accompanying unilateral finger movements: Pre-movement and movement-evoked fields. *Exp. Brain Res.* **1989**, *78*, 604–612. [\[CrossRef\]](#)
5. Harrop, R.; Weinberg, H.; Brickett, P.; Dykstra, C.; Robertson, A.; Cheyne, D.; Baff, M.; Crisp, D. The biomagnetic inverse problem: Some theoretical and practical considerations. *Phys. Med. Biol.* **1987**, *32*, 1545–1557. [\[CrossRef\]](#)
6. Hämäläinen, M.S.; Ilmoniemi, R.J. Interpreting magnetic fields of the brain: Minimum norm estimates. *Med Biol. Eng. Comput.* **1994**, *32*, 35–42. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Papadelis, C.; Leonardelli, E.; Staudt, M.; Braun, C. Can magnetoencephalography track the afferent information flow along white matter thalamo-cortical fibers? *NeuroImage* **2012**, *60*, 1092–1105. [\[CrossRef\]](#)
8. Krieger, D.; McNeil, M.; Zhang, J.; Puccio, A.; Schneider, W.; Li, X.; Okonkwo, D.O. Very high resolution neuroelectric brain imaging by referee consensus processing. *Int. J. Adv. Comput. Sci.* **2014**, *1*, 15–25.
9. Krieger, D.; Shepard, P.; Okonkwo, D.O. Normative atlases of neuroelectric brain activity and connectivity from a large human cohort. *arXiv* **2018**, arXiv:1805.01552.
10. Fischl, B.; Sereno, M.I.; Dale, A.M. Cortical Surface-Based Analysis. *NeuroImage* **1999**, *9*, 195–207. [\[CrossRef\]](#) [\[PubMed\]](#)
11. Reuter, M.; Schmansky, N.J.; Rosas, H.D.; Fischl, B. Within-subject template estimation for unbiased longitudinal image analysis. *NeuroImage* **2012**, *61*, 1402–1418. [\[CrossRef\]](#)
12. Morin, C.M.; Belleville, G.; Bélanger, L.; Ivers, H. The Insomnia Severity Index: Psychometric Indicators to Detect Insomnia Cases and Evaluate Treatment Response. *Sleep* **2011**, *34*, 601–608. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Gagnon, C.; Bélanger, L.; Ivers, H.; Morin, C.M. Validation of the Insomnia Severity Index in Primary Care. *J. Am. Board Fam. Med.* **2013**, *26*, 701–710. [\[CrossRef\]](#)
14. Lancaster, M.A.; McCrea, M.A.; Nelson, L.D. Psychometric properties and normative data for the Brief Symptom Inventory-18 (BSI-18) in high school and collegiate athletes. *Clin. Neuropsychol.* **2016**, *30*, 321–333. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Raghavan, S.S.; Rosenfeld, B.; Rasmussen, A. Measurement Invariance of the Brief Symptom Inventory in Survivors of Torture and Trauma. *J. Interpers. Violence* **2017**, *32*, 1708–1729. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Dixon, W.J. *BMDP Statistical Software Manual*; Chapter on BMDP7M; University of California Press: Berkeley, CA, USA, 1990.
17. Afifi, A.; Clark, V. *Computer Aided Multivariate Analysis*; Wadsworth Publishing: London, UK, 1984.
18. Tabachnick, B.G.; Fidell, L.S. *Using Multivariate Statistics*; Pearson/Allyn & Boe: Boston, MA, USA, 2007.
19. Taylor, J.R.; Williams, N.; Cusack, R.; Auer, T.; Shafto, M.A.; Dixon, M.; Tyler, L.K.; Henson, R.N. The Cambridge Centre for Ageing and Neuroscience (CamCAN) data repository: Structural and functional MRI, MEG, and cognitive data from a cross-sectional adult lifespan sample. *NeuroImage* **2015**, *144*, 262–269. [\[CrossRef\]](#)
20. Shafto, M.A.; Can, C.; Tyler, L.K.; Dixon, M.; Taylor, J.R.; Rowe, J.B.; Cusack, R.; Calder, A.J.; Marslen-Wilson, W.D.; Duncan, J.; et al. The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) study protocol: A cross-sectional, lifespan, multidisciplinary examination of healthy cognitive ageing. *BMC Neurol.* **2014**, *14*, 1–25. [\[CrossRef\]](#)
21. Collins, M.W.; Kontos, A.P.; Okonkwo, D.O.; Almquist, J.; Bailes, J.; Barisa, M.; Bazarian, J.; Bloom, O.J.; Brody, D.; Cantu, R.; et al. Concussion is treatable: Statements of agreement from the targeted Evaluation and Active Management (TEAM) approaches to treating concussion meeting held in Pittsburgh, October 15–16, 2015. *Neurosurgery* **2016**, *79*, 912–929. [\[CrossRef\]](#)
22. Yendiki, A.; Panneck, P.; Srinivasan, P.; Stevens, A.; Zöllei, L.; Augustinack, J.; Wang, R.; Salat, D.; Ehrlich, S.; Behrens, T.; et al. Automated probabilistic reconstruction of white-matter pathways in health and disease using an atlas of the underlying anatomy. *Front. Neuroinform.* **2011**, *5*, 23. [\[CrossRef\]](#)
23. Scahill, R.I.; Frost, C.; Jenkins, R.; Whitwell, J.L.; Rossor, M.N.; Fox, N.C. A Longitudinal Study of Brain Volume Changes in Normal Aging Using Serial Registered Magnetic Resonance Imaging. *Arch. Neurol.* **2003**, *60*, 989–994. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Hämäläinen, M.; Hari, R.; Ilmoniemi, R.J.; Knuutila, J.; Lounasmaa, O.V. Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain. *Rev. Mod. Phys.* **1993**, *65*, 413–497. [\[CrossRef\]](#)
25. Krieger, D.; Onodipe, S.; Charles, P.J.; Scabassi, R.J. Real time signal processing in the clinical setting. *Ann. Biomed. Eng.* **1998**, *26*, 462–472. [\[CrossRef\]](#)
26. *MaxFilter User's Guide*; Software Version 2.1; Elekta Neuromag Oy: Helsinki, Finland, 2008.



27. Krieger, D.; Shepard, P.; Zusman, B.; Jana, A.; Okonkwo, D. Shared high value research resources: The CamCAN human lifespan neuroimaging dataset processed on the open science grid. In Proceedings of the 2017 IEEE International Conference on Bioinformatics and Biomedicine, Kansas City, MO, USA, 13–16 November 2017; IEEE: Piscatawy, NJ, USA, 2017; pp. 1815–1822.
28. Pordes, R.; Petravick, D.; Kramer, B.; Olson, D.; Livny, M.; Roy, A.; Avery, P.; Blackburn, K.; Wenaus, T.; Würthwein, F.; et al. The open science grid. *J. Phys.* **2008**, *78*. [[CrossRef](#)]
29. Sfiligoi, I.; Bradley, D.C.; Holzman, B.; Mhashilkar, P.; Padhi, S.; Wurthwein, F. The Pilot Way to Grid Resources Using glideinWMS. In Proceedings of the 2009 WRI World Congress on Computer Science and Information Engineering, Los Angeles, CA, USA, 31 March–2 April 2009; Volume 2, pp. 428–432. [[CrossRef](#)]
30. Gan, Z.S.; Stein, S.C.; Swanson, R.; Guan, S.; Garcia, L.; Mehta, D.; Smith, D.H. Blood biomarkers for traumatic brain injury: A quantitative assessment of diagnostic and prognostic accuracy. *Front. Neurol.* **2019**. [[CrossRef](#)] [[PubMed](#)]
31. Peltz, C.B.; Kenney, K.; Gill, J.; Diaz-Arrastia, R.; Gardner, R.C.; Yaffe, K. Blood biomarkers of traumatic brain injury and cognitive impairment in older veterans. *Neurology* **2020**, *95*, e1126–e1133. [[CrossRef](#)] [[PubMed](#)]
32. Lewine, J.D.; Davis, J.T.; Sloan, J.H.; Kodituwakku, P.W.; Orrison, W.W., Jr. Neuromagnetic assessment of pathophysiologic brain activity induced by minor head trauma. *Am. J. Neuroradiol.* **1999**, *20*, 857–866. [[PubMed](#)]
33. Huang, M.; Theilmann, R.J.; Robb, A.; Angeles, A.; Nichols, S.; Drake, A.; Dandrea, J.; Levy, M.; Holland, M.; Song, T.; et al. Integrated imaging approach with MEG and DTI to Detect Mild Traumatic Brain Injury in Military and Civilian Patients. *J. Neurotrauma* **2009**, 1213–1226. [[CrossRef](#)]
34. Huang, M.; Risling, M.; Baker, D.G. The role of biomarkers and MEG-based imaging markers in the diagnosis of post-traumatic stress disorder and blast-induced mild traumatic brain injury. *Psychoneuroendocrinology* **2016**, *63*, 398–409. [[CrossRef](#)]
35. Sack, A.T.; Kadosh, R.C.; Schuhmann, T.; Moerel, M.; Walsh, V.; Goebel, R. Optimizing Functional Accuracy of TMS in Cognitive Studies: A Comparison of Methods. *J. Cogn. Neurosci.* **2009**, *21*, 207–221. [[CrossRef](#)]
36. Alexander, B.; Laycock, R.; Crewther, D.P.; Crewther, S.G. An fMRI-Neuronavigated chronometric TMS investigation of V5 and intraparietal cortex in motion driven attention. *Front. Hum. Neurosci.* **2018**, *11*. [[CrossRef](#)]
37. Inuggi, A.; Filippi, M.; Chieffo, R.; Agosta, F.; Rocca, M.A.; González-Rosa, J.J.; Cursi, M.; Comi, G.; Leocani, L. Motor area localization using fMRI-constrained cortical current density reconstruction of movement-related cortical potentials, a comparison with fMRI and TMS mapping. *Brain Res.* **2010**, *1308*, 68–78. [[CrossRef](#)]
38. Trapp, N.T.; Uitermarkt, B.; Johnson, M.K.; Kosciak, T.R.; Garrett, L.; Heinzerling, A.; Zanaty, M.; Holland, M.T.; Howard, M.; Boes, A.D. A new device to improve target localization for transcranial magnetic stimulation therapy. *Brain Stimul.* **2019**, *12*, 1600–1602. [[CrossRef](#)]
39. Luber, B.M.; Davis, S.; Bernhardt, E.; Neacsu, A.; Kwapil, L.; Lisanby, S.H.; Strauman, T.J. Using neuroimaging to individualize TMS treatment for depression: Toward a new paradigm for imaging-guided intervention. *NeuroImage* **2017**, *148*, 1–7. [[CrossRef](#)]
40. Van Dun, K.; Bodranghien, F.; Manto, M.; Mariën, P. Targeting the Cerebellum by noninvasive neurostimulation: A review. *Cerebellum* **2017**, *16*, 695–741. [[CrossRef](#)] [[PubMed](#)]
41. Sonmez, A.I.; Camsari, D.D.; Nandakumar, A.L.; Voort, J.L.V.; Kung, S.; Lewis, C.P.; Croarkin, P.E. Accelerated TMS for Depression: A systematic review and meta-analysis. *Psychiatry Res.* **2019**, *273*, 770–781. [[CrossRef](#)]
42. Alexopoulos, G.S. Mechanisms and treatment of late-life depression. *Transl. Psychiatry* **2019**, *9*, 1–16. [[CrossRef](#)] [[PubMed](#)]
43. Mayhew, S.D.; Hylands-White, N.; Porcaro, C.; Derbyshire, S.W.; Bagshaw, A.P. Intrinsic variability in the human response to pain is assembled from multiple, dynamic brain processes. *NeuroImage* **2013**, *75*, 68–78. [[CrossRef](#)]
44. Tan, H.-R.; Gross, J.; Uhlhaas, P. MEG—measured auditory steady-state oscillations show high test–retest reliability: A sensor and source-space analysis. *NeuroImage* **2015**, *122*, 417–426. [[CrossRef](#)] [[PubMed](#)]
45. Legget, K.T.; Hild, A.K.; Steinmetz, S.E.; Simon, S.T.; Rojas, D.C. MEG and EEG demonstrate similar test–retest reliability of the 40 Hz auditory steady-state response. *Int. J. Psychophysiol.* **2017**, *114*, 16–23. [[CrossRef](#)]
46. Onishi, H.; Otsuru, N.; Kojima, S.; Miyaguchi, S.; Saito, K.; Inukai, Y.; Yamashiro, K.; Sato, D.; Tamaki, H.; Shirozu, H.; et al. Variability and reliability of paired-pulse depression and cortical oscillation induced by median nerve stimulation. *Brain Topogr.* **2018**, *31*, 780–794. [[CrossRef](#)]
47. Recasens, M.; Uhlhaas, M. Test-retest reliability of the magnetic mismatch negativity response to sound duration and omission deviants. *NeuroImage* **2017**, *157*, 184–195. [[CrossRef](#)]
48. Dunkley, B.T.; Wong, S.M.; Jetly, R.; Wong, J.K.; Taylor, M.J. Post-traumatic stress disorder and chronic hyperconnectivity in emotional processing. *NeuroImage Clin.* **2018**, *20*, 197–204. [[CrossRef](#)]
49. Takeuchi, N.; Sugiyama, S.; Inui, K.; Kanemoto, K.; Nishihara, M. Long-latency suppression of auditory and somatosensory change-related cortical responses. *PLoS ONE* **2018**, *13*, e0199614. [[CrossRef](#)]
50. Ahonen, L.; Huotilainen, M.; Brattico, E. Within- and between-session replicability of cognitive brain processes: An MEG study with an N-back task. *Physiol. Behav.* **2016**, *158*, 43–53. [[CrossRef](#)]
51. Solomon, J.; Boe, S.G.; Bardouille, T. Reliability for non-invasive somatosensory cortex localization: Implications for pre-surgical mapping. *Clin. Neurol. Neurosurg.* **2015**, *139*, 224–229. [[CrossRef](#)]
52. Colclough, G.L.; Woolrich, M.W.; Tewarie, P.K.; Brookes, M.J.; Quinn, A.J.; Smith, S.M. How reliable are MEG resting-state connectivity metrics? *NeuroImage* **2016**, *138*, 284–293. [[CrossRef](#)]
53. Dimitriadis, S.I.; Routley, B.; Linden, D.E.; Singh, K.D. Reliability of Static and Dynamic Network Metrics in the Resting-State: A MEG-Beamformed Connectivity Analysis. *Front. Neurosci.* **2018**, *12*, 506. [[CrossRef](#)] [[PubMed](#)]



54. Martín-Buro, M.C.; Garcés, P.; Maestú, F. Test-retest reliability of resting-state magnetoencephalography power in sensor and source space. *Hum. Brain Mapp.* **2016**, *37*, 179–190. [[CrossRef](#)]
55. Candelaria-Cook, F.T.; Schendel, M.E.; Ojeda, C.J.; Bustillo, J.R.; Stephen, J.M. Reduced parietal alpha power and psychotic symptoms: Test-retest reliability of resting-state magnetoencephalography in schizophrenia and health controls. *Schizophr. Res.* **2019**, *215*, 229–240. [[CrossRef](#)] [[PubMed](#)]
56. Marquetand, J.; Vannoni, S.; Carboni, M.; Hegner, Y.L.; Stier, C.; Braun, C.; Focke, N.K. Reliability of Magnetoencephalography and High-Density Electroencephalography Resting-State Functional Connectivity Metrics. *Brain Connect.* **2019**, *9*, 539–553. [[CrossRef](#)] [[PubMed](#)]
57. Piitulainen, H.; Illman, M.; Laaksonen, K.; Jousmäki, V.; Forss, N. Reproducibility of corticokinematic coherence. *NeuroImage* **2018**, *179*, 596–603. [[CrossRef](#)]
58. Garcés, P.; Martín-Buro, M.C.; Maestú, F. Quantifying the test-retest reliability of magnetoencephalography resting-state functional connectivity. *Brain Connect.* **2016**, *6*, 448–460. [[CrossRef](#)]
59. Kimura, T.; Ozaki, I.; Hashimoto, I. Impulse Propagation along Thalamocortical Fibers Can Be Detected Magnetically outside the Human Brain. *J. Neurosci.* **2008**, *28*, 12535–12538. [[CrossRef](#)]
60. Roth, B.; Wikswo, J. The magnetic field of a single axon. A comparison of theory and experiment. *Biophys. J.* **1985**, *48*, 93–109. [[CrossRef](#)]
61. Krieger, D.; McNeil, M.; Zhang, J.; Schneider, W.; Li, X.; Okonkwo, D.O. Referee consensus. In Proceedings of the Conference on Extreme Science and Engineering Discovery Environment: Gateway to Discovery, XSEDE 2013, San Diego, CA, USA, July 2007; ACM: New York, NY, USA, 2007. ISBN 978-1-4503-2170-9. [[CrossRef](#)]
62. Krieger, D.; Becker, J.; Fabrizio, M.; McNeil, M.; Schneider, W.; Habeych, M.; Okonkwo, D. LP35: Task relevant high frequency neuroelectric oscillations in humans. *Clin. Neurophysiol.* **2014**, *125* (Suppl. S1). [[CrossRef](#)]
63. Antonakakis, M.; Dimitriadis, S.I.; Zervakis, M.; Papanicolaou, A.C.; Zouridakis, G. Aberrant Whole-Brain Transitions and Dynamics of Spontaneous Network Microstates in Mild Traumatic Brain Injury. *Front. Comput. Neurosci.* **2020**, *13*, 1–19. [[CrossRef](#)] [[PubMed](#)]
64. Krieger, D.; McNeil, M.; Zhang, J.; Li, X. Referee consensus: A platform technology for nonlinear optimization. In *Proceedings of the Conference on Extreme Science and Engineering Discovery Environment*; Association for Computing Machinery: New York, NY, USA, 2013; Article No. 12; ISBN 9781450321709.