Supplementary Materials

July 2, 2015

1 EEG-measures of consciousness

Table 1 makes explicit the abbreviations of the EEG-measures. Their computation closely follows Sitt et al. (2014) (supplement).

PE Θ	Permutation Entropy
K	Kolmogorov Complexity
wSMI Θ	Weighted Symmetrical Mutual Information
SMI Θ	Symmetrical Mutual Information
α	Alpha PSD
$\|\alpha\|$	Normalized Alpha PSD
β	Beta PSD
$\|\beta\|$	Normalized Beta PSD
δ	Delta PSD
$\|\delta\|$	Normalized Delta PSD
γ	Gamma PSD
$\|\gamma\|$	Normalized Gamma PSD
θ	Theta PSD
$\|\theta\|$	Normalized Theta PSD
MSF	Median Power Frequency
SE90	Spectral Edge 90
SE95	Spectral Edge 95
SE	Spectral Entropy
P1	Mid-latency Auditory Potential to the first sound
CNV	Contingent Negative Variation
P3A	Evoked Potential 280ms-340ms after the 5th sound
P3B	Evoked Potential 400ms-600ms after the 5th sound
MMN	Mismatch Negativity
ΔP3A	Contrasted P3A (Local Deviant vs Local Standard)
ΔP3B	Contrasted P3B (Global Deviant vs Global Standard)
ΔMMN	Contrasted MNN (Local Deviant vs Local Standard)
Decod Global	Classification based on the global effect
Decod Local	Classification based on the local effect

Table 1: Acronyms used for EEG-measures

2 Univariate area under the curve (AUC) analysis

The receiver-operator characteristic (ROC) plots the false positive rates (FPR) as a function of true positive rates (TPR). For example, in a comparison of the alpha power between CS and VS patients, one can observe the percentage of patients (TPR) who show a higher alpha power than an arbitrarily setup criterion C, and the corresponding percentage of VS patients (FPR). By testing all possible empirical criteria, a ROC curve can be drawn that allows to compute the AUC.

To illustrate the usage of univariate AUC scores in the present context, values higher than 0.5 mean MCS patients have on average higher alpha power than VS patients, and values lower than 0.5 imply that MCS patients have lower alpha power than VS patients. The AUC derived from the ROC curve provides a criterion-free estimate of the discrimination capacity of a given measure. Note that, based on on Sitt et al. (2014), the univariate AUC scores were computed from the full data and do not constitute cross-validated performance scores. Their statistical evaluation is instead based on traditional significance testing.

3 Supplementary analyses

3.1 Replication of univariate classification

To relate the current implementation of the computation of EEG-measures against the literature, a subset of the main analyses reported in Sitt et al. (2014) was repeated on the original data. The dataset comprised 76 vegetative state (VS) patient recordings, 69 minimally conscious state (MCS) and 24 councious state (CS) recordings. For this purpose, the cross-trial averages of the most discriminative EEG-measures were considered. Figure 1 depicts topographic comparisons and figure 2 depicts scalar summaries based on AUC scores. The results are highly consistent with the previous findings and argue in favor of a correct re-implementaiton of the EEG-measures.

3.2 Validation of multivariate classification against alternative software implementations

To further validate the reimplementation of then EEG-measures and its multivariate classification procedure, alternative algorithms were compared based on alternative software solutions. For this purpose the R-language for statistical computing was used (R Core Team, 2015) in concert with the caret package for predictive modeling (Kuhn, 2008). A set of linear models (Support Vector Machines, Elastic Net, partial least squares discriminant anlysis), rule-based models (Random Forests, C5.0) and non-linear models (Neural Networks) were tuned using a 5 times repeated 10 fold cross-validation with a fixed random seed and AUC scoring for the VS vs MCS comparison. For detailed descriptions of and references for these models see Kuhn & Johnson (2013). The default tuning parameters provided by the caret package were used. Models were then compared using resampling statistics as described in Eugster et al. (2008). The results ares summarized in 3. The overall performance is equivalent to the main results on the full electrode net. The resampled performance scores are very similar across modes, with the exception of the random forest classifier, which shows a slightly higher sensitivity than the other models. These findings suggest that the multivariate classification results reported here and in Sitt et al. (2014) are essentially independent of the particular classification algorithm used.

References

Eugster, Manuel J. A., Hothorn, Torsten, and Leisch, Friedrich. Exploratory and inferential analysis of benchmark experiments, 2008. URL `http://nbn-resolving.de/urn/resolver.pl?urn=nbn:de:bvb:19-epub-4134-6`.

Kuhn, Max. Building predictive models in r using the caret package. *Journal of Statistical Software*, 28(5):1–26, 11 2008. ISSN 1548-7660. URL `http://www.jstatsoft.org/v28/i05`.

Kuhn, Max and Johnson, Kjell. *Applied Predictive Modeling*. Springer, New York, Heidelberg, Dordrecht, London, 2013. URL `https://dl.dropboxusercontent.com/u/108263707/_book/KuhnJohnson2013apm.pdf`.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2015. URL `http://www.R-project.org/`.

Sitt, J D, King, J-R, El Karoui, I, Rohaut, B, Faugeras, F, Gramfort, A, Cohen, L, Sigman, M, Dehaene, S, and Naccache, L. Large scale screening of neural signatures of consciousness in patients in a vegetative or minimally conscious state. *Brain*, 137(8):2258–2270, 2014.

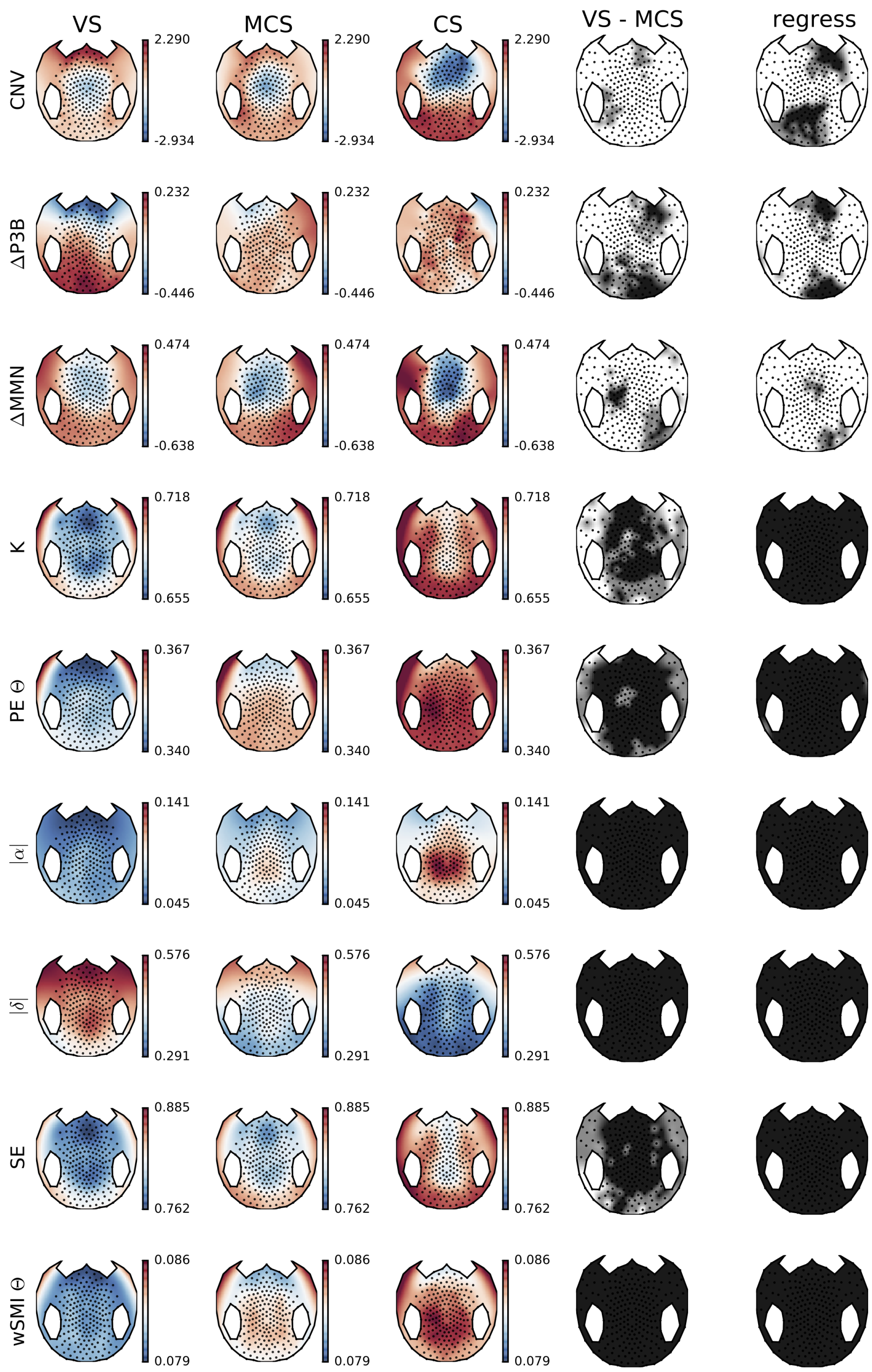

Figure 1: Scalps topographies of the most discriminatory measures from Sitt et al. (2014). The first three columns depict average topographies across subjects for VS, MCS and CS, respectively. The fourth column depicts sensor-wise significance levels based on a Mann-Whithet-U statistic for the VS-MCS contrast. The fifth row shows sensor-wise significance levels bases on a regression across diagnostic categories. Here a continuum of the following kind was was assumed: $VS < MCS < CS$. Significance levels are represented by black, grey and white, referring to $p < 0.01$, $p < 0.05$, and 'not significant', respectively.

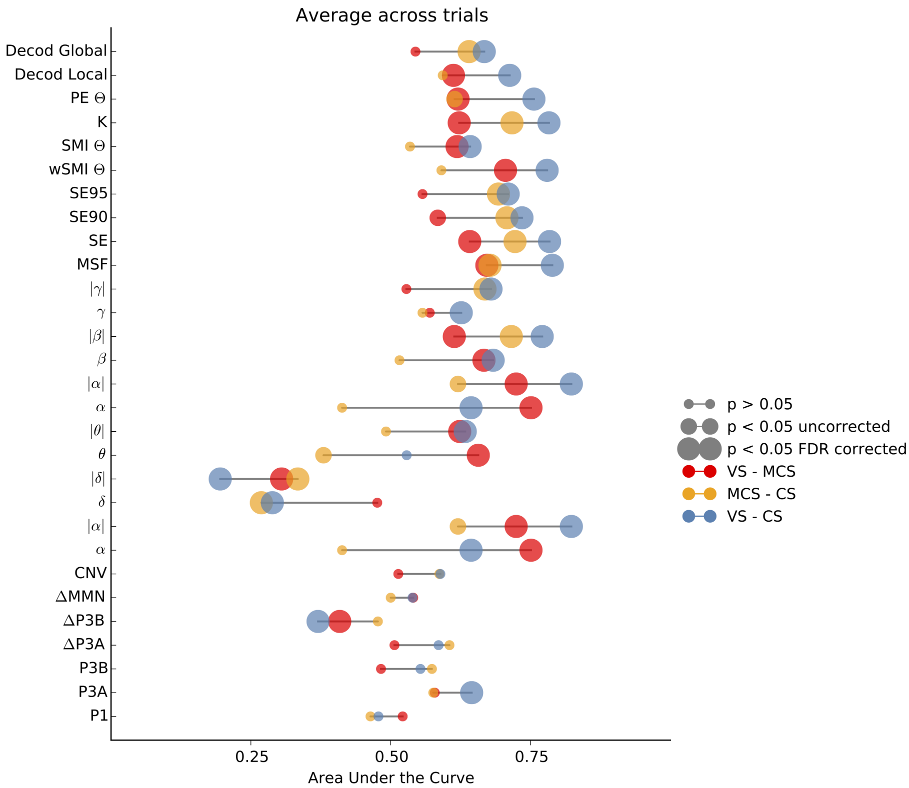

Figure 2: Discrimination performance for a subset of EEG-measures reported in Sitt et al. (2014). Each line provides a summary report of its respective measure. The meaning of each acronym can be found in table 1. The location of each dot corresponds to the AUC for a pair-wise comparison between two states of consciousness. Chance level corresponds to a AUC score of 0.5. A AUC > 0.5 suggests that the corresponding measure is correlated with the state of consciousness (from VS to MCS and CS). A AUC score < 0.5 suggests that the corresponding measure is negatively correlated with the state of consciousness. Color and size of the plotting markers indicate the type and the significance level of the comparison (cf. legend). Note that the contingent negative variation and MMN, which are negative EEG components, where not sign-flipped (in contrast to Sitt et al. (2014)).

.

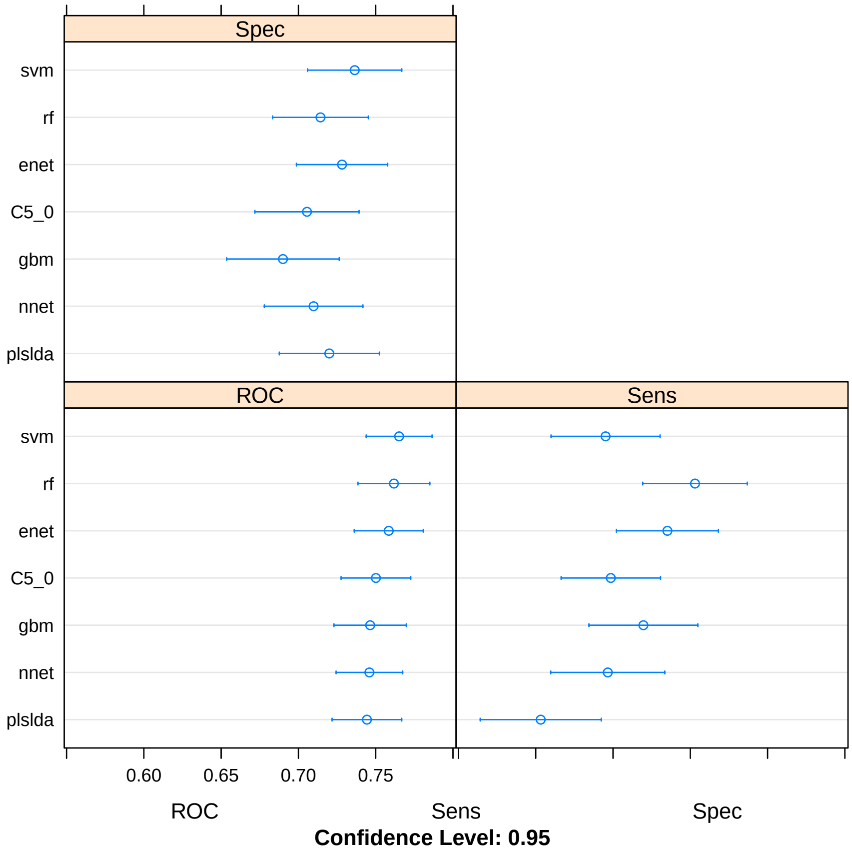

Figure 3: AUC (lower left), sensitivity (lower right) and specificity (upper left) for alternative statistical models. In each subplot, each line shows the 95% confidence intervals on resampled performance scores for a given statistical model. From top to bottom, Support Vector Machines (SVM), Random Forests (RF), Elastic Net (ENET), the C5.0 rule-based model, gradient boosted tree (GBM), , Neural Networks (NNET), and partial least squares discriminant analysis (PLSDA).