Predicting subject traits from brain spectral signatures: an application to brain ageing

Cecilia Jarne[a,b,c], Ben Griffin[d,c], Diego Vidaurre[c,d]

[a] *Departamento de Ciencia y Tecnologia de la Universidad Nacional de Quilmes, Bernal, Buenos Aires, Argentina*

[b] *CONICET, Buenos Aires, Argentina*

[c] *Center of Functionally Integrative Neuroscience, Department of Clinical Medicine, Aarhus University, Aarhus, Denmark*

[d] *Oxford Centre for Human Brain Activity, Department of Psychiatry, Oxford University, Oxford, UK*

Abstract

The prediction of subject traits using brain data is an important goal in neuroscience, with relevant applications in clinical research, as well as in the study of differential psychology and cognition. While previous research has primarily focused on neuroimaging data, our focus is on the prediction of subject traits from electroencephalography (EEG), a relatively inexpensive, widely available and non-invasive data modality. However, EEG data is complex and needs some form of feature extraction for subsequent prediction. This process is almost always done manually, risking biases and suboptimal decisions. Here, we propose a largely data-driven use of the EEG spectrogram, which reflects macro-scale neural oscillations in the brain. Specifically, the key idea of this paper is to reinterpret the spectrogram as a probability distribution and then leverage advanced machine learning techniques that can handle probability distributions with mathematical rigour and without the need for manual feature extraction. The resulting technique, Kernel Mean Embedding Regression (KMER), shows superior performance to alternative methods thanks to its capacity to handle nonlinearities in the relation between the EEG spectrogram and the trait of interest. We leveraged this method to predict biological age in a multinational EEG data set, HarMNqEEG, showing the method's capacity to generalise across experiments and acquisition setups.

Keywords:
Brain age, EEG, Machine Learning, Kernel methods, Maximum Mean Discrepancy

1. Introduction

Predicting behavioural and cognitive traits from brain data in a way that generalises to unseen subjects and is robust to acquisition idiosyncrasies is important because it can offer objective measures to otherwise elusive neurobiological constructs (Haynes and Rees (2006)). A widely studied example is brain age. While measuring actual age is a straightforward task, the concept of brain age provides a marker of mental health by quantifying how much a subject's brain appears to have aged with respect to the population average; that is, a predicted age that is lower than the individual's chronological age, for example, may indicate that the person has a brain that appears younger than expected for their actual age (Franke and Gaser (2019); Smith et al. (2019)).

Most of the work on the prediction of subject traits (such as age) from brain data has been done with resting-state fMRI data. Since we cannot straightforwardly predict from the raw data, an intermediate representation is typically used for prediction. In the case of fMRI, this is often a simple description of functional connectivity (Rosenberg et al. (2016)) or some model of brain dynamics (Liegeois et al. (2019); Vidaurre et al. (2021); Ahrends et al. (2023)). Here, we instead focus on EEG, a considerably less costly technique. As an intermediate representation, we consider the EEG frequency spectrum, reflecting neural oscillations that are well-known correlates of different behavioural and cognitive states (Buzsáki and Draguhn (2004)). But how to meaningfully extract features for prediction from an EEG spectrum is an open question. Most current efforts are based on manual feature extraction (Al Zoubi et al. (2018); Engemann et al. (2022)), or are based on models that need to be estimated from raw data and whose properties depend on the choice of configuration and hyperparameters (Vidaurre et al. (2013)).

In this paper, with no prior predefinition of frequency bands or any manual feature engineering besides basic preprocessing, we propose a method based on the idea of interpreting the EEG spectrogram as a probability distribution. This way, we can fully leverage all the powerful machinery of kernel learning on probability distributions. Specifically, we consider Kernel Mean Embeddings (KME), a technique used to construct a representation of the data in a high-dimensional feature space (Smola et al. (2007); Iyer et al. (2014); Borgwardt et al. (2006)). The KME technique maps joint, marginal and conditional probability distributions to vectors in a high (or even infinite) dimensional feature space that completely characterizes the distribution (Fukumizu et al. (2011)). Building upon KME, we used the Maximum Mean Discrepancy (MMD) (Smola et al. (2007)), a distance metric defined on the space of probability measures (here, EEG spectrograms) in combination with kernel ridge regression (Saunders et al. (1998)). We refer to our approach as Kernel Mean Embedding Regression (KMER).

We demonstrate the method for age prediction. Using a multisite, public resting-state EEG dataset with a very wide distribution of age from 1 to 97 years old (HarMNqEEG, Li et al. (2022)), we show that KMER leads to better prediction performances that can be projected on the EEG scalp for interpretation. KMER revealed that parietal sensors are the most accurate in predicting age, with slightly greater accuracy in women. We also showed that the predictions generalise well across experiments and acquisition sites, even considering the large differences in age distribution across sites (a well-known problem in machine learning referred to as prior shift). Overall, by demonstrating its predictive capacity and interpretability, we show how this method can help unveil insights about brain age or other neurobiological constructs.

2. Results

We developed Kernel Mean Embedding Regression (KMER), a novel method to predict subject traits from EEG spectrograms; see Methods and Figure 1. In short, KMER is based on the idea of interpreting EEG spectrograms as probability distributions so that techniques based on kernel learning of distributions are readily applicable. By forming a subject-by-subject kernel matrix of similarities between subjects (using a Gaussian kernel; see Methods), this technique allowed us to derive predictions for each EEG channel that can naturally accommodate non-linearities in a data-driven way without the need for manual feature extraction. We used this approach to predict age in the HarMNqEEG dataset (Li et al. (2022)), a multicountry resting-state EEG dataset of 1966 subjects encompassing people across the entire lifespan (here, we excluded the very youngest, using only subjects between 5 and 97 years old); see Figure S1 (SI) for some basic statistics about

the dataset in terms of age, gender, and geography.

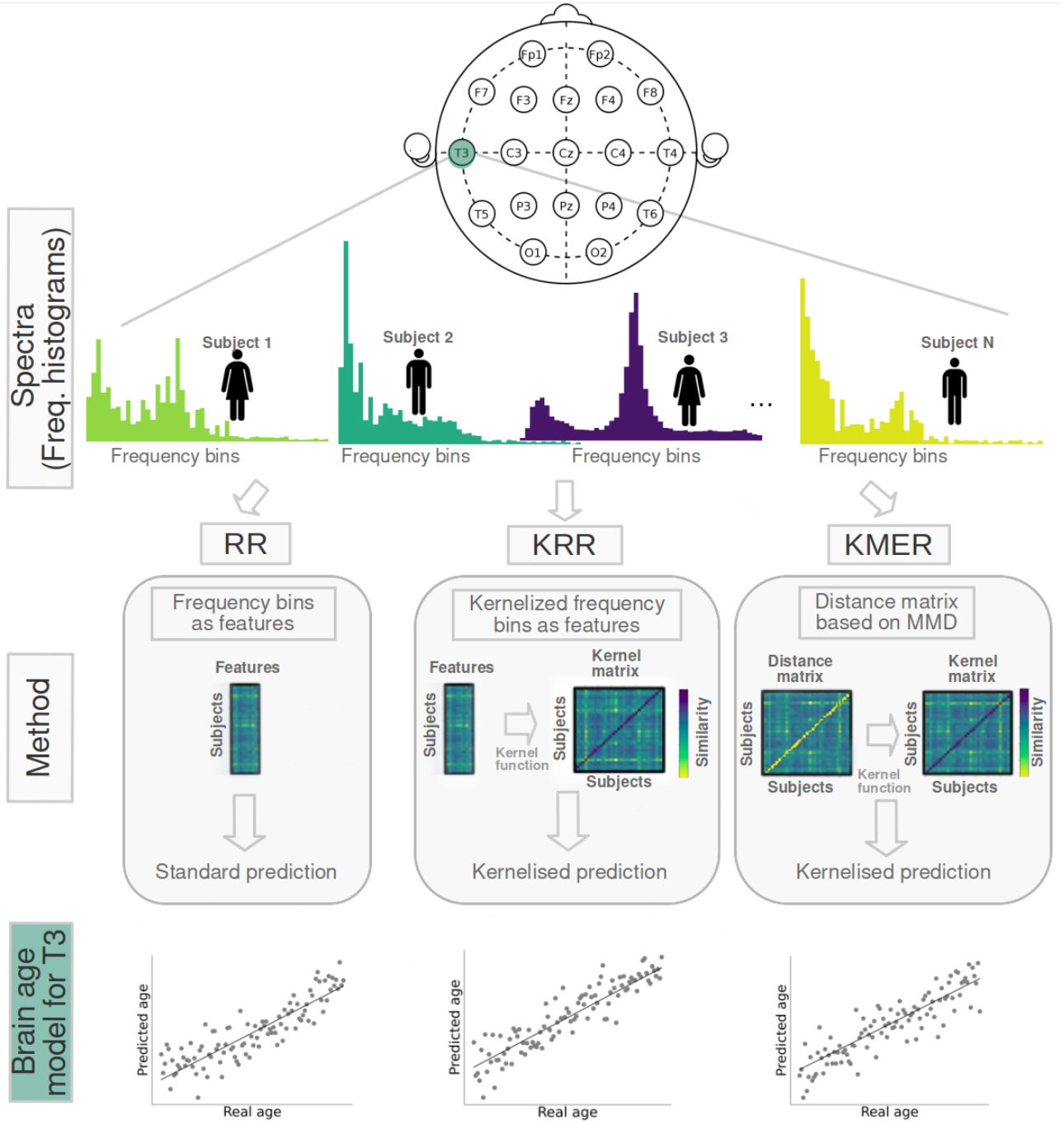

Figure 1: General workflow of the analysis. We used the HarMNqEEG dataset, which contains (adequately normalized) power densities per sensor and participant, as well as information on gender, age and scanning batch (acquisition site). We considered three prediction approaches, which we run separately per EEG sensor: (i) Ridge Regression (RR) using the power estimates at each frequency bin of the spectrogram as features; (ii) Kernel Ridge Regression (KRR), a kernelised version of RR (based on a non-linear Gaussian kernel) also based on the power estimates per bin; and (iii) our Kernel Mean Embedding Regression (KMER) approach, where we interpreted the power spectral estimates across bins as probability distributions so that we can leverage the mathematical machinery of kernel learning on probability measures for prediction.

2.1. Kernel Mean Embedding Regression outperforms existing alternatives

We compared the proposed method with two baseline alternatives. The first, Ridge Regression (RR), is the standard (linear) approach of using power as each frequency bin as a separate predictor, with a ridge (L_2) regularisation penalty. The second, Kernel Ridge Regression (KRR), is the kernelised version of RR, which allows the modelling of nonlinearities by the use of an appropriate kernel. For KRR, as well as for KMER, we used a Gaussian kernel and an L_2 regularisation penalty. The regularisation penalty was, in all cases, chosen using nested cross-validation. We assessed the performance of the three methods per EEG channel using two measures of accuracy: prediction explained variance (R^2) and mean absolute error (MAE). A comparative schematic of the three approaches is illustrated in Figure 1.

Figure 2 summarizes the results per EEG channel. Figure 2A illustrates an example of age vs. predicted age for a given channel (T3), where we can observe that the biggest errors concentrate at older age (this was the case for all predictions; not shown). Figure 2B and 2C show a comparison of the three methods across channels in terms of MAE and explained variance, respectively. KRR slightly outperforms RR, but both methods are dominated by KMER virtually in all channels, and sometimes by a large margin (e.g. sensors O1 and O2). This is likely due to the presence of nonlinearities in the relation between EEG spectrograms and age, which are more effectively captured by KMER than by KRR (and not captured at all by RR). As shown, MAE oscillates between 10 and 12 years approximately for all channels. There is a larger variation in terms of explained variance, ranging from $R^2 = 0.28$ for sensor Fp1 to $R^2 = 0.47$ for sensor C3. Finally, Figure 2D shows the explained variance in sensor space for KMER (MAE has a similar but inverted topography; not shown). We can observe that the parietal sensors are the most predictive of chronological age, whereas prefrontal and occipital sensors are the least predictive. Results for the linear and polynomial kernel are shown in Figures S2 and S3 (SI); these do not differ drastically from the Gaussian kernel, except that the linear kernel shows slightly less accuracy.

In order to further compare the methods, we performed statistical testing across sensors (considering each sensor's accuracy as an observation). Specifically, we used the Wilcoxon signed-rank test (Wilcoxon (1945)) between each pair of methods. KMER was significantly superior across sensors than the alternative methods (RR, $p = 7.62 \times 10^{-06}$; KRR $p = 7.24 \times 10^{-05}$).

2.2. Sex differences

Females and males have previously shown to exhibit differences in their ageing trajectories (Hägg and Jylhävä (2021)). Here, we investigated differences in the model outcomes between sexes in terms of explained variance (R^2) using KMER, given the 884 females and 905 males in the dataset (137 were non-specified).

Figure 3 presents the results. Figure 3A and 3B show example scatter plots between age and predicted age, separately per sex. Figure 3B and 3C show R^2 across sensors in both barplot and sensor space format. Analogously, Figure 3D and 3E show MAE per sex.

As observed, females generally exhibited a higher prediction accuracy than males, potentially suggesting that their pace of biological change corresponds more closely to chronological age. Nevertheless, there are no striking differences between the two sexes; see Figure S4 (SI) for a male-vs-female accuracy scatter plot, where each dot represents a sensor (Wilcoxon signed-rank test with $pvalue = 0.0192$).

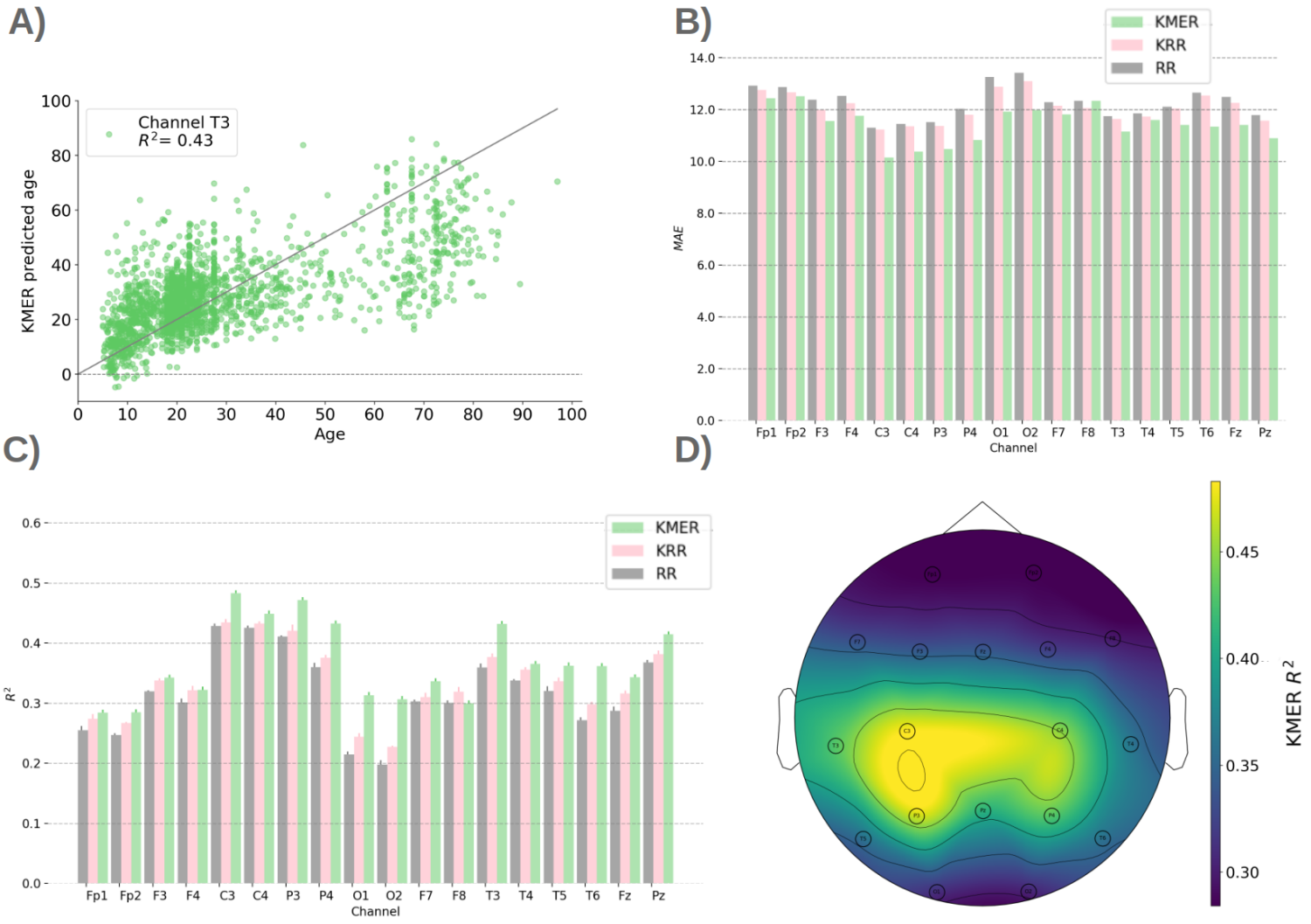

Figure 2: KMER shows lower errors and higher accuracies than standard alternatives, RR and KRR. **A)** Illustration of predicted vs. real age for channel T3 for KMER. **B)** Mean Absolute Error (MAE) per channel for KMER, RR and KRR. **C)** Explained variance R^2 per channel for the three methods. **D)** Explained variance projected on sensor space for KMER.

2.3. Performance variations across scanning sites

The HarMNqEEG dataset encompasses a wide range of age groups across various sites or geographical locations (14 different sites/experiments), each with a different number of individuals and specific age distributions; see Figure S1 (SI). Despite efforts made by the data curators to ensure homogeneity, it is always possible that differences in practice and instrumentation between the sites leak into the data. In the previous analyses, sites were never split across cross-validation folds; that is, when predicting the age of subjects in a given site (e.g. Colombia), none of the subjects of that site were part of the training data. We refer to this approach as leave-one-site-out cross-validation. Given that substantial differences in age distribution between sites may potentially coexist with other between-site differences due to e.g. instrumentation, leave-one-site-out cross-validation was performed to ensure that the reported prediction accuracies were purely reflective of age and not mixed with spurious factors. However, this conservative approach makes the problem more difficult, inducing what is known in the machine learning literature as a prior shift (Kouw (2018)). This means that the distribution of the dependent variable (here, age) changes be-

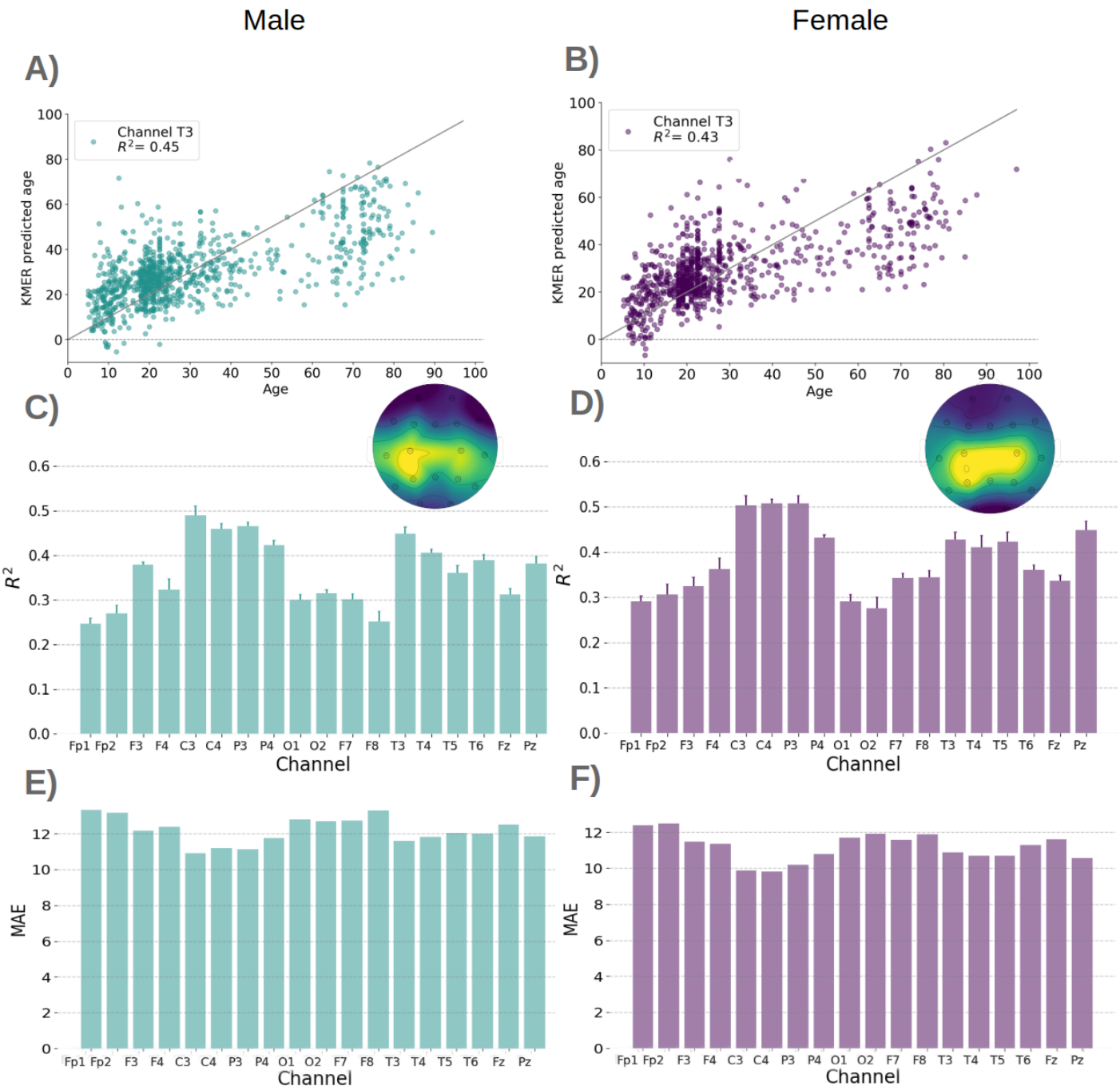

Figure 3: Sex differences in age estimation performed using KMER (left panels, male; right panels, female). **A)** and **B)** Predicted vs. real age per sex for channel T3; results are consistent for all channels (not shown). **C)** and **D)** Explained variance R^2 per sex and channel, in barplot format and projected on sensor space. **E)** and **F)** Analogously, MAE per sex and channel.

tween training and testing data. In this section, we investigated this issue by comparing KMER's performance for leave-one-site-out cross-validation vs. fold-blind cross-validation.

Figure 4A shows predicted age vs. chronological age with colours indicating stratification by site, with two different perspectives: 3D to better appreciate each site individually and 2D to compare the sites side-by-side. Here, we can observe substantial differences between sites in both accuracy

and age distribution. Figure 4B shows distributions of delta across sites, defined as the difference between predicted and chronological age (Franke and Gaser (2019); Smith et al. (2019)). When compared with Figure S1, where we show the age distributions explicitly, we can observe that the distribution of errors is contingent on the age range within each group, e.g. underestimating the age of older individuals (see Switzerland as an example, which has the oldest population). The observed pattern remains consistent across different channels (not shown). These results highlight the importance of age distributions in prediction.

In order to characterize the impact of the cross-validation scheme, Figure 4C shows the difference between the two cross-validation approaches in terms of both R^2 and MAE. As observed, leave-one-site-out cross-validation results in a substantially lower accuracy. While this is most likely due to the different distributions of age across sites, we cannot completely discard other between-site effects. We further compared the cross-site to a within-site prediction, where we ran a cross-validated prediction within one given site without using data from the other sites. While this can only be done effectively for sites with a sufficient number of subjects, we observed substantial increases in accuracy in this type of prediction. Figure S5 of Supplementary Information shows the case of New York, where, the explained variance went up to 0.63 and MAE became as small as 5.1.

Overall, these results demonstrate the challenges of predicting across sites when the age distributions differ, but also that KMER can still produce reasonable accuracies even in the face of this problem.

3. Methods

The general workflow of our analysis is presented in Figure 1. In summary, we first considered the frequency spectrum of each individual across subjects for every subject. Predictions of age were made per channel in a cross-validation fashion, where the acquisition sites (e.g. Russia) were used as folds. We used three approaches to predict age from the EEG spectrograms: Ridge Regression, Kernel Ridge Regression, and our Kernel Mean Embedding Regression. For scoring prediction performance, we used explained variance (computed as the coefficient of determination; R^2) and the mean absolute error (MAE), which offer complementary information (Engemann et al. (2022)). We also computed the so-called delta, defined as the signed difference between predicted and actual age, which is often used to quantify brain age. All code used in this paper is open and publicly available at: https://github.com/katejarne/Kernel_Max_mean_discrepancy_EEG_Age.

3.1. The HarMNqEEG dataset

We used EEG normalized spectra from the HarMNqEEG dataset (Li et al. (2022)), which originated from a legacy dataset associated with the Cuban Human Brain Mapping initiative (Valdes-Sosa et al. (2021)). The data used in our study were collected from 9 countries, 12 EEG systems, and 14 experiments. Overall, there are 1966 subjects, of which 40 subjects without a recorded age were excluded. Also, we discarded babies and toddlers and considered only subjects between 5 and 97 years old.

Recordings were taken from the 19 channels of the 10/20 International Electrodes Positioning System: Fp1, Fp2, F3, F4, C3, C4, P3, P4, O1, O2, F7, F8, T3/T7, T4/T8, T5/P7, T6/P8, Fz, Cz, Pz, where the Cz is the reference electrode. Data was formatted as cross-spectral matrices sampled from 1.17 to 19.14 Hz, with a 0.39 Hz resolution. The scalp EEG cross-spectrum was calculated using Bartlett's method (Møller (1986)) by averaging the periodograms of more than 20

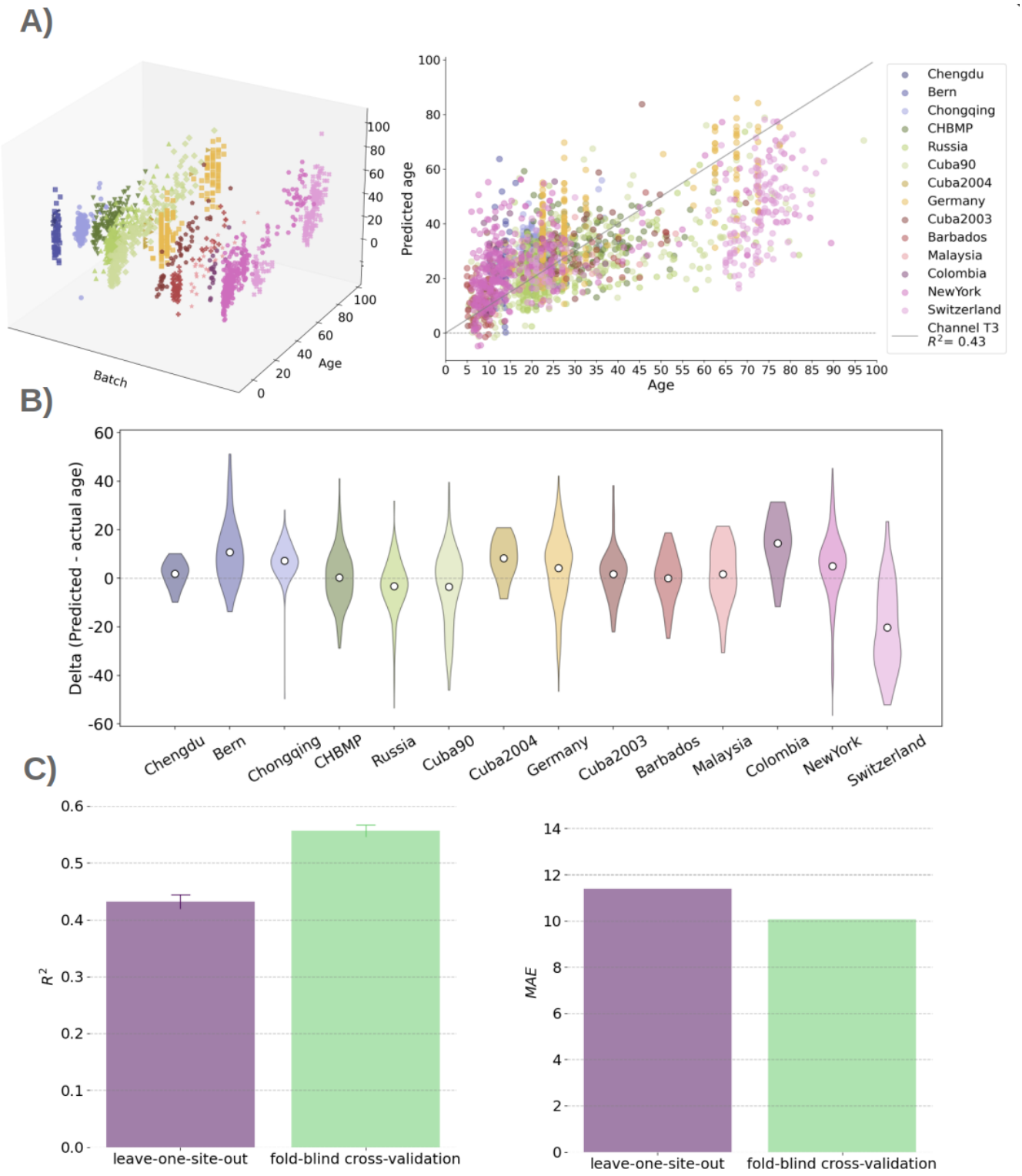

Figure 4: Exploration of the effect of site on the KMER predictions. **A)** Predicted vs. real age in 3D (left) and 2D (right). **B)** Distribution of delta (defined as predicted minus real age) per site **C)** Comparison of leave-one-site-out cross-validation (where the sites were taken into account for structuring the folds) versus fold-blind cross-validation (where they were not); a consistently higher accuracy is observed in terms of both R^2 and MAE for fold-blind cross-validation.

consecutive and non-overlapping segments. Because the spectra from different sites have different maximum cutoff frequencies, we used the lowest maximum cutoff frequency across sites for full compatibility. Thus, the histogram of each channel has 49 bins corresponding to a maximum cutoff frequency of 19.13 Hz.

3.2. Maximum Mean Discrepancy

We next describe the mathematical foundation of our method. Kernel embeddings fully characterize the distributions by mapping joint, marginal, and conditional probability distributions to vectors in a high (or infinite) dimensional feature space (Fukumizu et al. (2011); Smola et al. (2007)). This implies that any two distributions P and Q with differences in any moment are mapped to separate points in a reproducing Hilbert space (Fukumizu et al. (2011)). KME can be applied to classification, regression, and clustering problems as it enables a manageable data representation while maintaining the pertinent information of the original data distribution. To quantify differences in the distributions we considered Maximum Mean Discrepancy, or MMD, a distance metric established on the space of probability measures.

Following Gretton et al. (2012), the MMD can be expressed in terms of the kernel mean embedding (KME) of the distributions P and Q, defined as:

$$\mu_P = \mathbb{E}_{x\sim P}[\phi(x)] \qquad \text{and} \qquad \mu_Q = \mathbb{E}_{z\sim Q}[\phi(z)] \tag{1}$$

where ϕ is the mapping function to a Hilbert feature space $\mathcal{H}$, and z is a data point. The MMD can then be expressed as:

$$MMD(\mathcal{F}, P, Q) = \|\mu_P - \mu_Q\|_{\mathcal{H}}\,. \tag{2}$$

That is, the MMD is the norm in the feature space $\mathcal{H}$ of the difference between the KME of the distributions P and Q. Expanding the MMD equation with the KME of the distributions P and Q in terms of μ_P and μ_Q, we have:

$$MMD(\mathcal{F}, P, Q) = \|\mu_P - \mu_Q\|_{\mathcal{H}} = \|\mathbb{E}_{x\sim P}[\phi(x)] - \mathbb{E}_{z\sim Q}[\phi(z)]\|_{\mathcal{H}} \tag{3}$$

Given some selection of a characteristic kernel, we obtain:

$$MMD(\mathcal{F}, P, Q) = \frac{1}{n^2}\sum_{i,j=1}^{n} k(x_i, x_j) + \frac{1}{m^2}\sum_{i,j=1}^{m} k(z_i, z_j) - \frac{2}{nm}\sum_{i,j=1}^{n,m} k(x_i, z_j), \tag{4}$$

where $X = \{x_1, \ldots, x_n\}$ and $Z = \{z_1, \ldots, z_m\}$ are samples from two distributions P and Q respectively; $k(x, z) = \langle\phi(x), \phi(z)\rangle_{\mathcal{H}}$ is the kernel associated with the Hilbert space $\mathcal{H}$; and $\langle\cdot,\cdot\rangle_{\mathcal{H}}$ denotes the inner product in $\mathcal{H}$. In the last equality, we have used the definition of the kernel k, where $k_X(i, j) = k(x_i, x_j)$ and $k_Z(i, j) = k(z_i, z_j)$, respectively.

The empirical estimation of the MMD in terms of the kernel function, called $\widehat{\text{MMD}}_k$, can be expressed as:

$$\widehat{\mathrm{MMD}}_k(\mathcal{F}, X, Z) = \frac{1}{n(n-1)} \sum_{i,j=1}^{n} k(x_i, x_j) + \frac{1}{m(m-1)} \sum_{i,j=1}^{m} k(z_i, z_j) - \frac{2}{nm} \sum_{i,j=1}^{n,m} k(x_i, z_j) \tag{5}$$

where $k(x, z)$ is the kernel function used to compute the MMD, and $\mathcal{F}$ is the space of functions for which the MMD distance is measured. Here, therefore, we considered MMD as our distance metric, which is defined on the space of probability measures.

We used frequency distributions (histograms) of each channel as the probability distributions. The central idea is to interpret EEG spectrograms as probability distributions and then use this information for prediction. Specifically, we applied the empirical estimator $\widehat{\mathrm{MMD}}_k$ defined in Equation 5 and considered different kernel functions to estimate the distances between each subject in the dataset. This produced a $(N \times N)$ distance matrix per kernel, where N is the number of subjects.

3.3. Predictive models

Three methods were considered. The first, Ridge Regression (RR), used the power estimations across bins as features in a standard regression model with an L_2 penalty. The other two, Kernel Ridge Regression (KRR) and Kernel Mean Embedding Regression (KMER), are both based on the construction of a kernel matrix. While KRR is based on the power across bins as features for the construction of the kernel matrix, KMER interprets the spectrogram as a probability distribution, using MMD as a metric of dissimilarity to construct the kernels.

3.3.1. Ridge regression

RR is a form of linear regression that includes an L_2 penalty term to the loss function. This term is controlled by a regularization hyperparameter called λ, which is tuned through cross-validation. RR, often used when there are many correlated predictors in the model, produces estimates as per:

$$\hat{\boldsymbol{y}}_{ridge} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X} + \lambda \boldsymbol{I})^{-1}\boldsymbol{X}^T\boldsymbol{y}, \tag{6}$$

where $\hat{\boldsymbol{y}}_{ridge}$ is the predicted response variable, $\boldsymbol{X}$ is the matrix of predictor variables, $\boldsymbol{y}$ is the vector of response variables, λ is the regularization hyparameter and $\boldsymbol{I}$ is the identity matrix.

Here, we performed RR to estimate the age of individuals based on their EEG frequency histograms, using all frequency bins as features.

3.3.2. Kernel-based regression

We considered a kernelised version of RR that can readily incorporate non-linearities by taking advantage of the so-called kernel trick (Vidaurre et al. (2021)). Both, KRR and KMER use this approach, although they construct the kernel matrix differently. In essence, a kernel prediction is performed as

$$\hat{f}(\boldsymbol{x}) = \sum_{i=1}^{N} \alpha_i k(\boldsymbol{x}, \boldsymbol{x}_i), \tag{7}$$

where $\hat{f}(\boldsymbol{x})$ is the predicted response variable, N is the number of samples, $\boldsymbol{x}$ is the data vector for a testing subject, $\boldsymbol{x}_i$ is the i-th training sample, $k(\cdot,\cdot)$ is the kernel function and α_i are the coefficients to be estimated.

In the case of KMER, $k(\cdot,\cdot)$ is based on the MMD; while KRR is based on ordinary kernels computed on the spectrograms directly.

4. Discussion

In this paper, we have introduced Kernel Mean Embedding Regression (KMER), a method for the prediction of individual traits from EEG spectral information based on the idea of interpreting channel spectrograms as probability distributions. By doing this, we could leverage mathematical principles from kernel learning. Although not pursued in this paper, the same principles can also be applied to other applications, such as unsupervised clustering of subjects from EEG spectral information. Even though the mathematical foundations of the method are not necessarily trivial, KMER is simple to implement and use, and it is not computationally costly in comparison to more complex (e.g. deep learning) approaches that are estimated on the raw data.

In order to demonstrate the improved performance of our model over alternative approaches, we selected the prediction of age as the target variables from the HarMNqEEG dataset. HarMNqEEG is a recently published international initiative that made EEG spectrograms available across different cohorts and countries. Compared to other works that predicted age within a single cohort, predicting across sites (where cross-validation folds were constructed such that sites were not split) presents a more challenging problem because the distribution of age across sites varies substantially and also because of potential protocol or infrastructure differences between the sites. Despite this difficulty, the presented results are good in terms of performance compared to previous EEG studies (Al Zoubi et al. (2018); Engemann et al. (2022)). Also, when predicting exclusively within New York (as an example of within-site prediction, given that New York has a large number of subjects and the broadest age range of all sites), the accuracy increased to the level of the highest accuracies reported in the literature using more complex models, while being much less computationall expensive. It is worth noting that this was the case despite the fact that, in contrast to these studies, we only used the spectrograms and had no access to the raw data.

Although not exhaustively explored here, the method can be further optimised by testing different kernel functions and hyperparameters. Here, we have considered three kernel functions: polynomial, Gaussian, and linear. We observed that the linear kernel underperformed with respect to the other two kernels, supporting our claim that a non-linear relationship exists between the EEG spectrogram and age.

As a final remark, here we demonstrated the benefit of KMER over alternative approaches by predicting age using EEG, but KMER can be applied to other modalities (e.g., MEG and fMRI) to predict other individual traits such as cognitive ability or clinical variables, for which accurate predictions are crucial for the clinical translation of findings.

Acknowledgments

D. Vidaurre is supported by a Novo Nordisk Foundation Emerging Investigator Fellowship (NNF19OC-0054895), an ERC Starting Grant (ERC-StG-2019-850404), and a DFF Project 1 from the Independent Research Fund of Denmark (2034-00054B). This research was funded in part by the Wellcome

Trust (215573/Z/19/Z). For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission. We acknowledge support from PICT 2020-01413. We also thank Sonsoles Alonso for her help.

Conflict of interest statement

The authors declare no competing conflicts of interest.

References

Ahrends, C., Woolrich, M. and Vidaurre, D. (2023), 'Predicting individual traits from models of brain dynamics accurately and reliably using the fisher kernel', *bioRxiv* .
URL: *https://www.biorxiv.org/content/early/2023/07/21/2023.03.02.530638*

Al Zoubi, O., Ki Wong, C., Kuplicki, R. T., Yeh, H.-w., Mayeli, A., Refai, H., Paulus, M. and Bodurka, J. (2018), 'Predicting age from brain eeg signals—a machine learning approach', *Frontiers in Aging Neuroscience* **10**.
URL: *https://www.frontiersin.org/articles/10.3389/fnagi.2018.00184*

Borgwardt, K. M., Gretton, A., Rasch, M. J., Kriegel, H.-P., Schölkopf, B. and Smola, A. J. (2006), 'Integrating structured biological data by Kernel Maximum Mean Discrepancy', *Bioinformatics* **22**(14), e49–e57.
URL: *https://doi.org/10.1093/bioinformatics/btl242*

Buzsáki, G. and Draguhn, A. (2004), 'Neuronal oscillations in cortical networks', *Science* **304**(5679), 1926–1929.
URL: *https://www.science.org/doi/abs/10.1126/science.1099745*

Engemann, D. A., Mellot, A., Höchenberger, R., Banville, H., Sabbagh, D., Gemein, L., Ball, T. and Gramfort, A. (2022), 'A reusable benchmark of brain-age prediction from m/eeg resting-state signals', *NeuroImage* **262**, 119521.
URL: *https://www.sciencedirect.com/science/article/pii/S105381192200636X*

Franke, K. and Gaser, C. (2019), 'Ten years of brainage as a neuroimaging biomarker of brain aging: What insights have we gained?', *Frontiers in Neurology* **10**(JUL).

Fukumizu, K., Song, L. and Gretton, A. (2011), Kernel bayes'rule, *in* J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira and K. Weinberger, eds, 'Advances in Neural Information Processing Systems', Vol. 24, Curran Associates, Inc.

Gretton, A., Borgwardt, K. M., Rasch, M. J., Schölkopf, B. and Smola, A. (2012), 'A kernel two-sample test', *J. Mach. Learn. Res.* **13**(1), 723–773.
URL: *http://dl.acm.org/citation.cfm?id=2503308.2188410*

Haynes, J.-D. and Rees, G. (2006), 'Decoding mental states from brain activity in humans', *Nature Reviews Neuroscience* **7**(7), 523–534.
URL: *https://doi.org/10.1038/nrn1931*

Hägg, S. and Jylhävä, J. (2021), 'Sex differences in biological aging with a focus on human studies', *eLife* **10**, e63425.
URL: *https://doi.org/10.7554/eLife.63425*

Iyer, A. S., Jagarlapudi, S. and Sarawagi, S. (2014), Maximum mean discrepancy for class ratio estimation: Convergence bounds and kernel selection, *in* 'International Conference on Machine Learning'.

Kouw, W. M. (2018), 'An introduction to domain adaptation and transfer learning', *CoRR* **abs/1812.11806**.
URL: *http://arxiv.org/abs/1812.11806*

Li, M., Wang, Y., Lopez-Naranjo, C., Hu, S., Reyes, R. C. G., Paz-Linares, D., Areces-Gonzalez, A., Hamid, A. I. A., Evans, A. C., Savostyanov, A. N., Calzada-Reyes, A., Villringer, A., Tobon-Quintero, C. A., Garcia-Agustin, D., Yao, D., Dong, L., Aubert-Vazquez, E., Reza, F., Razzaq, F. A., Omar, H., Abdullah, J. M., Galler, J. R., Ochoa-Gomez, J. F., Prichep, L. S., Galan-Garcia, L., Morales-Chacon, L., Valdes-Sosa, M. J., Tröndle, M., Zulkifly, M. F. M., Abdul Rahman, M. R. B., Milakhina, N. S., Langer, N., Rudych, P., Koenig, T., Virues-Alba, T. A., Lei, X., Bringas-Vega, M. L., Bosch-Bayard, J. F. and Valdes-Sosa, P. A. (2022), 'Harmonized-multinational qeeg norms (harmnqeeg)', *NeuroImage* **256**, 119190.
URL: *https://www.sciencedirect.com/science/article/pii/S1053811922003147*

Liegeois, R., Li, J., Kong, R., Orban, C., Van De Ville, D., Ge, T., Sabuncu, M. R. and Yeo, B. T. T. (2019), 'Resting brain dynamics at different timescales capture distinct aspects of human behavior', *Nature Communications* **10**(1), 2317.
URL: *https://doi.org/10.1038/s41467-019-10317-7*

Møller, J. (1986), 'Bartlett adjustments for structured covariances', *Scand. J. Statist.* **13**, 1 – 15. Cited by: 17.

Rosenberg, M. D., Finn, E. S., Scheinost, D., Papademetris, X., Shen, X., Constable, R. T. and Chun, M. M. (2016), 'A neuromarker of sustained attention from whole-brain functional connectivity', *Nature Neuroscience* **19**(1), 165–171.
URL: *https://doi.org/10.1038/nn.4179*

Saunders, C., Gammerman, A. and Vovk, V. (1998), Ridge regression learning algorithm in dual variables, *in* 'Proceedings of the Fifteenth International Conference on Machine Learning', Morgan Kaufmann, pp. 515–521. Edited by J.Shavlik.

Smith, S. M., Vidaurre, D., Alfaro-Almagro, F., Nichols, T. E. and Miller, K. L. (2019), 'Estimation of brain age delta from brain imaging', *NeuroImage* **200**, 528–539.
URL: *https://www.sciencedirect.com/science/article/pii/S1053811919305026*

Smola, A., Gretton, A., Song, L. and Schölkopf, B. (2007), A hilbert space embedding for distributions, *in* M. Hutter, R. A. Servedio and E. Takimoto, eds, 'Algorithmic Learning Theory', Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 13–31.

Valdes-Sosa, P. A., Galan-Garcia, Lidice AU own, S., Valdes-Urrutia, L., Evans, A. C. and Valdes-Sosa, M. J. (2021), 'The cuban human brain mapping project, a young and middle age population-based eeg, mri, and cognition dataset', *Scientific Data* **8**(1), 45.
URL: *https://doi.org/10.1038/s41597-021-00829-7*

Vidaurre, D., Bielza, C. and Larrañaga, P. (2013), 'Classification of neural signals from sparse autoregressive features', *Neurocomputing* **111**, 21–26.
URL: *https://www.sciencedirect.com/science/article/pii/S0925231213000271*

Vidaurre, D., Llera, A., Smith, S. and Woolrich, M. (2021), 'Behavioural relevance of spontaneous, transient brain network interactions in fmri', *NeuroImage* **229**, 117713.
URL: *https://www.sciencedirect.com/science/article/pii/S1053811920311988*

Wilcoxon, F. (1945), 'Individual comparisons by ranking methods', *Biometrics Bulletin* **1**(6), 80–83.
URL: *http://www.jstor.org/stable/3001968*