Evaluation of Non-Negative Matrix Factorization of grey matter in age prediction

Deepthi P. Varikuti[1,2,3], Sarah Genon[1,2,3*], Aristeidis Sotiras[4], Holger Schwender[5], Felix Hoffstaedter [1,2,3], Christiane Jockwitz[6,7], Svenja Caspers[6,7], Susanne Moebus[8], Katrin Amunts[6,7], Christos Davatzikos[4], Simon B. Eickhoff [1,2,3*]

[1]Institute of Neuroscience and Medicine (INM-7, Brain & Behavior); Research Centre Jülich, Germany; [2]Institute of Systems Neuroscience, Medical Faculty; Heinrich-Heine University Düsseldorf, Düsseldorf, Germany; [3]Institute of Clinical Neuroscience and Medical Psychology; Heinrich-Heine University Düsseldorf, Düsseldorf, Germany; [4]Section for Biomedical Image Analysis, Center for Biomedical Image Computing and Analytics, University of Pennsylvania, Philadelphia, PA 19104, USA [5]Mathematical Institute, Heinrich Heine University Düsseldorf, Düsseldorf, Germany; [6]Institute of Neuroscience and Medicine (INM-1); Research Centre Jülich, Germany; [7]C. & O. Vogt Institute for Brain Research, Medical Faculty, Heinrich Heine University, Düsseldorf. Germany; [8]Institute of Medical Informatics, Biometry and Epidemiology, University of Duisburg-Essen, Essen, Germany.

Introduction

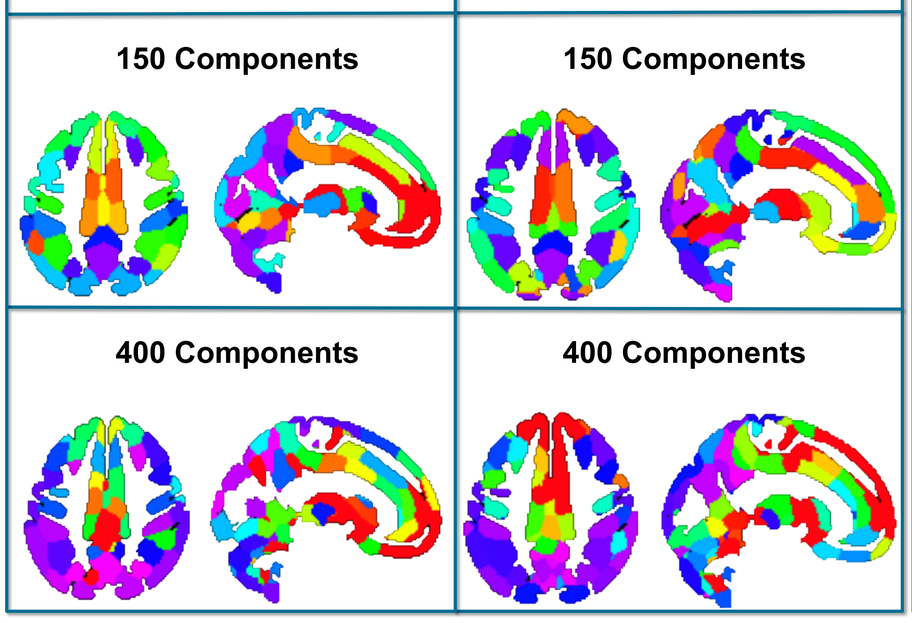
- **Applying machine-learning methods to voxel-based morphometry (VBM) data can predict the age by brain imaging [1].**
- **Data dimensionality reduction is a critical aspect of such brain-based prediction to counter the curse of dimensionality associated with voxel-wise analysis.**
- **While previous age-predictions have employed PCA based compression, non-negative matrix factorization (NNMF) has recently been suggested as a biologically more plausible factorization of high-dimensional VBM data [2].**
- **NNMF provides non-negative, orthonormal and sparse factorization, yielding more interpretable representations than PCA [3].**
- **In this study, we examine the application of NNMF to VBM data for brain age prediction with investigation of the following issues:**
  - **Does NNMF allow us to predict the biological age at a performance level comparable to previous reports [1]?**
  - **Are the factorizations transferrable between two different datasets, which differ in demographic characteristics and scanner protocols?**
  - **Is it possible to adjust the model bias inherent to many brain-based predictions?**
  - **Do the brain features contributing to the age prediction provide a meaningful spatial pattern?**

Methods

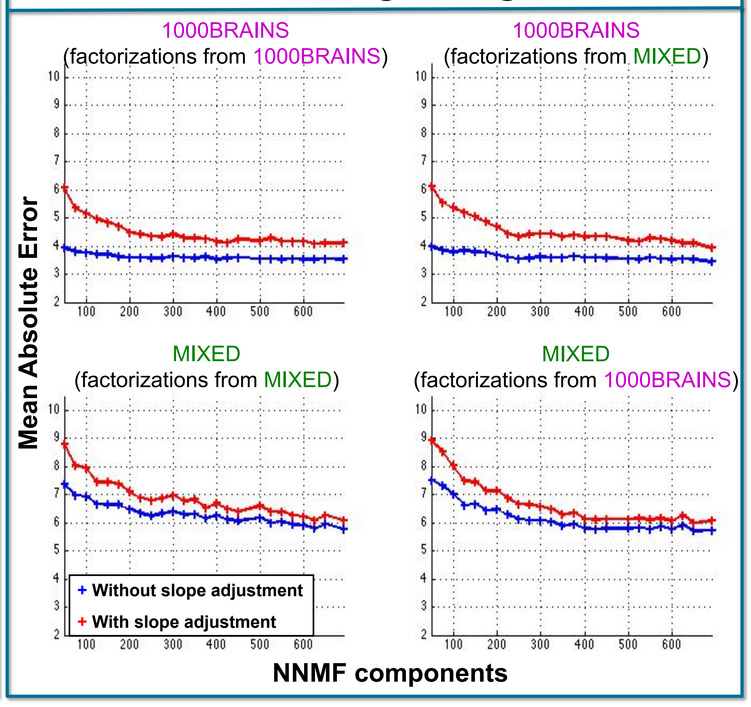
- **Samples: Two independent cohort's datasets:**
  - **uniform protocol in 693 healthy older subjects aged 55-76 (1000BRAINS [4])**
  - **heterogeneous multi-site data for 1084 healthy subjects aged 19-81 (MIXED)**
- **Grey matter volume (GMV) computation: VBM8 preprocessing [5] ; DARTEL normalization [6], only non-linear modulation, 8mm FWHM smoothing**
- **GMV Factorization: NNMF factorizes high dimensional voxel-wise data of the entire group into sparse components and subject-specific loading coefficients. And then factorizations for both groups were derived at different levels of granularity.**
- **Regression model: age prediction was performed by fitting LASSO regression models,**
  - **on the coefficient matrix obtained from the respective NNMF**
  - **on the coefficient matrix obtained by projecting a cohort's data on the respective other cohort's components**
- **Evaluation of model generalization: 10-fold cross-validation replicated 25 times.**
- **An additional linear adjustment was performed by fitting a regression-line between the real and predicted training set and then by using the parameters (such as slope and the intercept) to adjust the expected slope of the test set to 45 degrees.**

Results

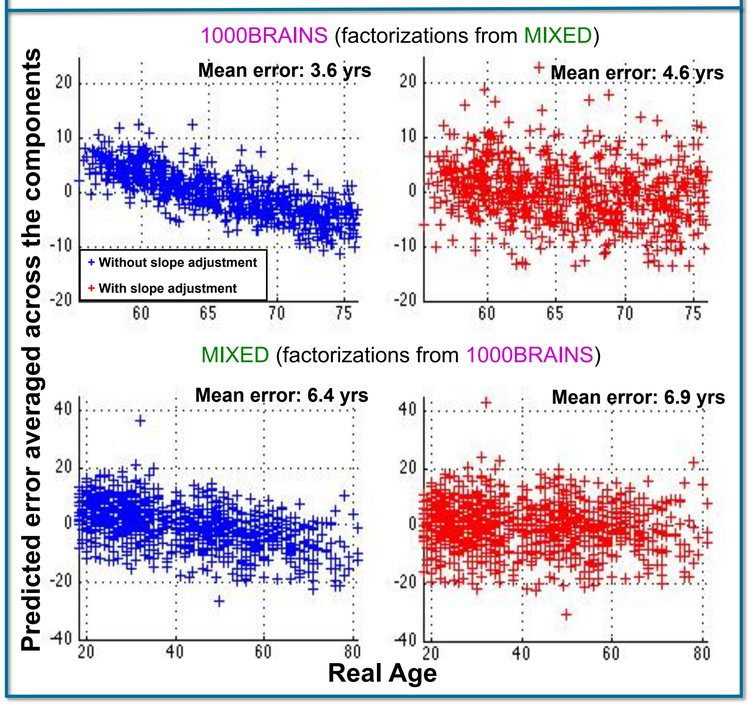
NNMF based clustering of the brain from the datasets



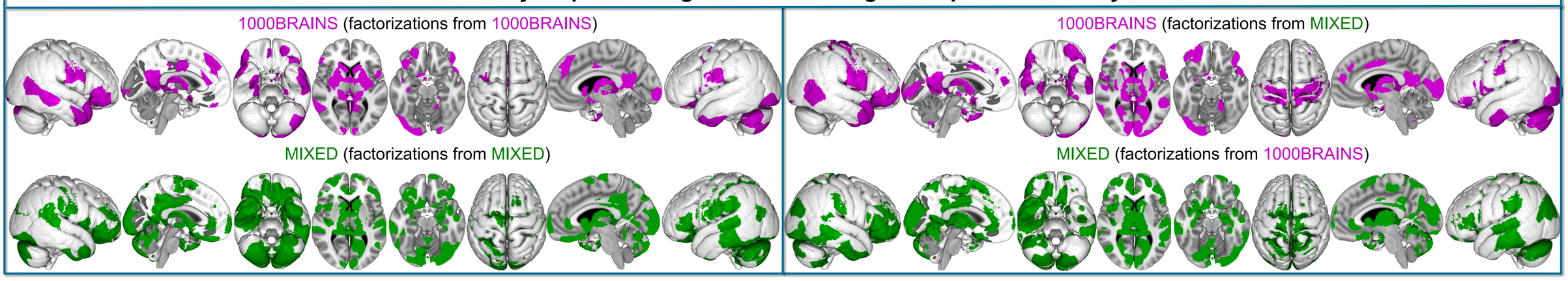
Predictions of age using NNMF



Prediction error averaged across the components



Summary map of the regions contributing in the prediction analysis



Conclusions

- **NNMF compression of VBM data over the lifespan allows predicting previously unseen subjects' age with a precision that is comparable to earlier reports using PCA for data compression [2], while offering the potential for neurobiological interpretation.**
- **Accuracy seems to be independent of whether the components were derived from the same dataset or from an independent dataset with different age distribution and scanner protocols.**
- **Accuracies tend to continuously increase with higher granularity, but reach a plateau around 400 components when using the factorization of the same data, whereas, for projection of the data, more components are required.**
- **Adjusting the inherent bias of sparse regression models yields unbiased out-of-sample predictions but comes at the expense of slightly higher mean errors.**
- **Most importantly, our framework combines prediction accuracy with interpretability quality regarding the patterns underlying the prediction**

References: [1] Franke et al., 2010, Neuroimage. 2010 Apr 15;50(3):883-92. [2] Sotiras et al., 2015, Neuroimage. 2015 Mar;108:1-16. [3] Lee and Sung 1999, Nature. 1999 Oct 21;401(6755):788-91. [4] Caspers et al., 2014, Front Aging Neuroscience. 2014 Jul 14;6:149. [5] http://www.neuro.uni-jena.de/vbm8. [6] Ashburner 2007, Neuroimage. 2007 Oct 15;38(1):95-113

Acknowledgments: This study was supported by the Deutsche Forschungsgemeinschaft (DFG, EI 816/4-1, LA 3071/3-1), the National Institute of Mental Health (R01-MH074457), the Helmholtz Portfolio Theme "Supercomputing and Modelling for the Human Brain" and the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 7202070 (HBP SGA1).

E-mail: d.varikuti@fz-juelich.de