Improving Individual-Specific Functional Parcellation Through Transfer Learning

Alexandre Le Bris, Louis Rouillard, Demian Wassermann

Improving Individual-Specific Functional Parcellation Through Transfer Learning

Alexandre Le Bris, Demian Wassermann, and Louis Rouillard

Université Paris-Saclay, Inria, CEA

Abstract. Resting-state functional magnetic resonance imaging enables the exploration of the functional brain organization and its representation via large-scale networks. Summary measures, such as functional connectivity profiles or networks' spatial topography, present inter-individual differences with applications in detecting cognitive disorders or predicting behavioral traits. However, the accuracy of these measures can be significantly disrupted due to the limited number of subjects in clinical data.
Transfer learning coupled with Bayesian modeling can be leveraged to overcome this issue. We simultaneously estimate the posteriors for the individual cortical topography and functional connectivity profiles by applying a variational inference technique over probabilistic graphical models. After extracting these features, inference performance is evaluated through the regression of behavioral scores. The Bayesian formalism allows information transfer by sharing parameters pre-trained on a large dataset, especially the posteriors inferred at a population level. This way, on the HCP dataset, we show that the knowledge captured on a large sub-population ($\sim$750 subjects and 4 scans per subject) helps improve the model trained on a much smaller sample (50 to 250 subjects with 1 scan per subject), even if this sample stems from another dataset, here CamCAN. By doing it with 50 subjects, we achieve comparable performance in behavioral prediction than a three times larger dataset.

Keywords: Neuroimaging · Functional Brain Networks · MRI · Bayesian Modeling · Transfer Learning

1 Introduction

Machine Learning-centric algorithms became the predominant techniques associating neuroimaging-based features with cognitive capabilities and diagnoses. However, the small sample sizes common in small-scale cognitive experiments and clinical settings hamper the application of such algorithms. To address this, we harness large-scale variational inference (VI) to implement a Bayesian-based transfer learning approach. So, by leveraging information learned in a large public dataset, we improve the capacity to predict behavioral measurements from neuroimaging data in small samples.

In predicting cognitive abilities from neuroimaging, recent works [17,2] propose employing both functional connectivity (FC) and topography as features. While topography pertains to spatial organization as the individualized parcellation of each subject's brain, FC relates to each subject's brain activity. FC is derived from functional magnetic resonance imaging (fMRI), which captures location-specific, time-varying brain

activity by measuring changes in oxygenation concentration, namely the Blood Oxygen Level Dependent (BOLD) signal. The FC between two brain regions is then determined as the temporal correlation of BOLD signals between them. This measurement reveals large-scale networks characterized by similar FC profiles within each [6]. In addition, these networks exhibit specific functions and architectonics, outlining a brain structure intertwining functional, structural, and connectivity organizations [31]. Many functional atlas have been produced, but mostly at a population level as the task requires averaging signals across a large cohort of subjects [27,5,23].

Recent endeavors have successfully mapped the cortex at the individual level, yielding compelling results in predicting human behavior from the inferred topography [17,18]. Individualized parcellation can also support the detection of cognitive disorders [6] or serve clinical purposes, e.g., in brain tumor care [20]. The latter use case involves clinical datasets involving a limited number of patients and a low signal-to-noise ratio.

Transfer learning, regarded as a strategy of leveraging knowledge learned on a source domain for a target task [28], can overcome the limits above. We present here a method harnessing the flexibility of Bayesian modeling [16,12] and VI [3] to provide individual probabilistic parcellations. This method is designed for computationally handling large parameter space induced by neuroimaging applications. Through predicting behavioral measures, we prove that both connectivity and spatial features inferred this way learn aspects of subject specificity. Then, we show the efficiency of transferring information from a pre-trained model to significantly enhance prediction performance on reduced datasets, whether sharing the same domain or coming from a different one.

2 Related Works

Compared to structural MRI, only a few studies focus on TL with fMRI data due to the added complexity introduced by the temporal dimension [1,28]. Among these, a typical application entails decoding brain states during behavior tasks and improving the classification accuracy on the target domain, which may differ from the source domain in terms of subjects, tasks, or acquisition sites [26,11,10]. Other favored applications include diagnosing Alzheimer's disease, major depressive disorder, or autism spectrum disorder, where domain adaptation strategies can mitigate the impact of data distribution shifts across different sites [30,9]. But, to the best of our knowledge, in fundamental research on cognition, no work has evaluated the benefits of TL for predicting cognitive scores from individualized parcellations.

TL is commonly performed by sharing parameters learned from the source domain data and fine-tuning part or all of them on the target domain, as seen on convolutional neural networks and kernel method [28]. Bayesian graphical models' methods fall into this category by identifying these parameters with prior knowledge. Suder et al. [24] distinguishes three primary strategies in Bayesian TL: parameters sharing via the likelihood, hierarchical modeling [14], and latent space sharing [32,19]. In this work, we will illustrate how these three strategies combined can transfer information from a large brain dataset to smaller studies.

3 Method

We implement individualized parcellation as variational inference (VI) on a Bayesian generative model. In this model, the observed data are individual functional connectomes from each subject's scan (Fig. 1A). As a first benchmark, we show that our method captures individual specificity through both RSFC profiles and functional topography. To assess inference performance and validate our approach, we rely on behavioral measures prediction as a downstream task. Then, we test TL by pre-training a variational family on a large dataset of $\sim$ 750 subjects and 4 scans per subject (HCP subdataset). After transferring the parameters to the new model and freezing the population-level RVs, we perform inference on two smaller datasets: a subset of the HCP dataset with unseen subjects and a subset of the CamCAN dataset, both limited to 1 scan per subject. By comparing the performance without TL, we prove that our method, mixing VI and Bayesian modeling, is well-suited for efficient TL on computationally extensive applications such as brain parcellation.

3.1 Bayesian generative model

We place ourselves in the Bayesian framework, whose purpose is inferring the posterior $p(\Omega|X)$ of the parameters (or latent variables) Ω given the observations X. As $p(X|\Omega)\,p(\Omega) = p(X, \Omega)$, we can define a generative model $\mathcal{M}$ describing the joint distribution $p(X, \Omega)$ from the likelihood $p(X|\Omega)$ and the prior $p(\Omega)$. It can also be visualized as a probabilistic graphical model (PGM) [12], which expresses as a graph the conditional dependence structure between the latent random variables (RVs). This representation allows one to motivate, design, and share probabilistic models easily.

We adopt a generative model $\mathcal{M}$ inspired by the multi-session hierarchical Bayesian model (MS-HBM) proposed by Kong et al. [17]. $\mathcal{M}$ is composed of two sets of RVs (see Fig. 1A and Table 1). The first one is dedicated to the spatial location of networks on the cortex (topography): e.g., $\theta_l^{s,n}$ encodes the probability of the voxel n belonging to the network l for the subject s, and thus provides a probabilistic parcellation. The second set gathers the resting-state functional connectivity (RSFC) profiles for each network: e.g., $\mu_{l,d}^{s,t}$ represents the connectivity between the network l and the region d for the subject s during the scan t. Both sets are expressed at different levels (population, subject, scan) and are combined through a mixture at the session level to generate the observation X. We provide more detail in the supplementary material, Appendix A.

Level	Population	Subject (plate S)	Scan / Session (plate T)
Connectivity RVs (RSFC profiles)	$\mu_{l,d}^{g} \sim \mathcal{N}(0, 1)$	$\mu_{l,d}^{s} \sim \mathcal{N}(\mu_{l,d}^{g}, \epsilon)$	$\mu_{l,d}^{s,t} \sim \mathcal{N}(\mu_{l,d}^{s}, \sigma)$ $X_d^{s,n,t} \sim \mathcal{N}(\mu_{l,d}^{s,t}, \kappa \vert l = l^{s,n,t})$
Spatial RVs (parcellation)	$\theta_l^n \sim \mathcal{N}(0, 3)$ (logit)	$\theta_l^{s,n} \sim \mathcal{N}(\theta_l^n, \gamma)$ (logit)	$l^{s,t,n} \sim \text{Categorical}(\theta_l^{s,n})$ (label)

Table 1. RVs in the multi-session hierarchical Bayesian model $\mathcal{M}$ and their distribution

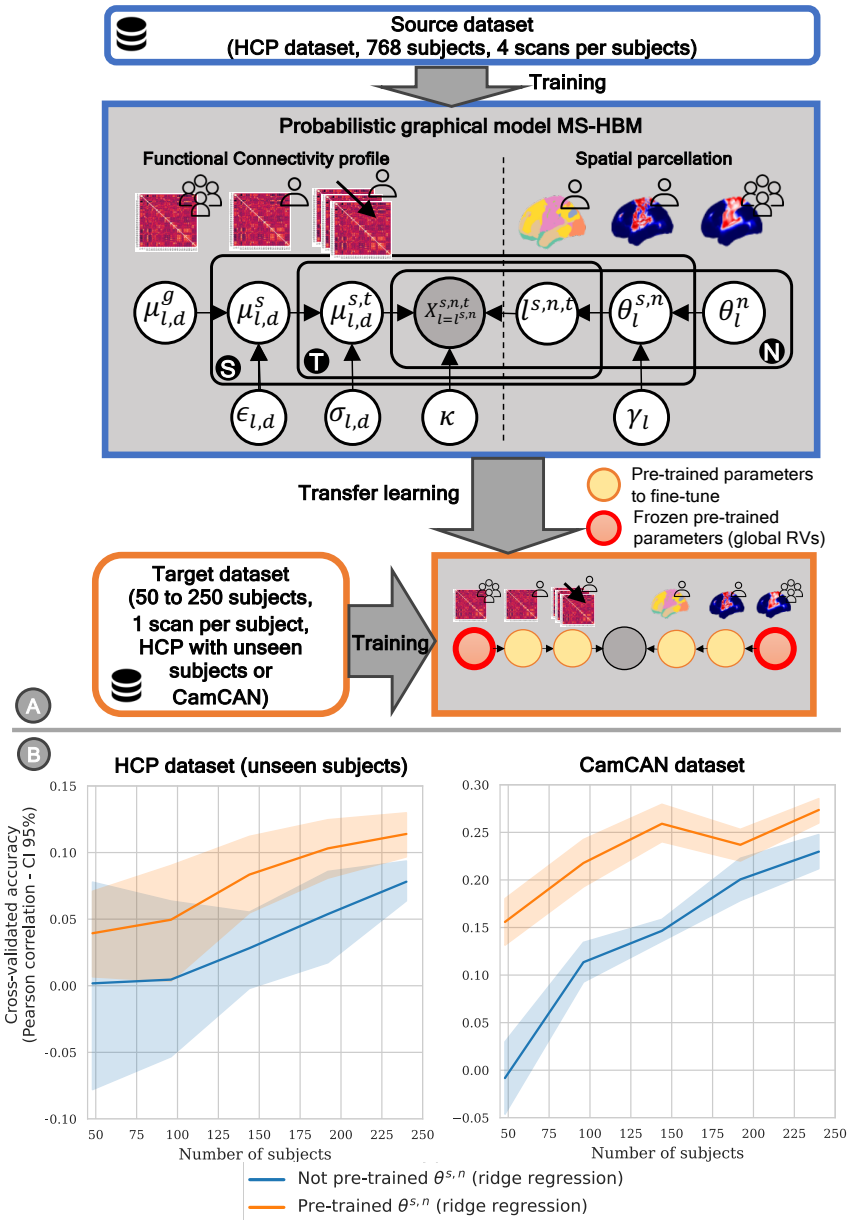

Fig. 1. Panel A - Graphical abstract of our approach for testing TL. We divided the HCP dataset into a large dataset, with all the scans for pre-training the model, and a smaller one, reduced to 1 scan per subject in order to simulate quantitatively a clinical dataset. We also selected the CamCAN project to validate our approach on an unrelated dataset with unseen conditions (e.g., new protocol, shorter acquisition time, larger lifespan range of the cohort). This way, we also test the robustness of our method to domain shift. We set the number of networks l to 17.**Panel B** - Learning curves for both datasets when predicting behavioral scores from individualized probabilistic parcellations $\theta_l^{s,n}$, with and without transfer learning. The accuracy was averaged over 13 behavioral scores for both cases (see supplementary material, Appendix D). The error bands represent the variability when modifying the composition of the source and target datasets in terms of subjects (we performed ten iterations for each sample size). The TL acts as a positive shift for small cohorts, capturing some subject-specific information, even with a dataset of 50 patients. Then, when increasing the population, the gap narrows, at least for the CamCAN dataset.

3.2 Variational inference and plate amortization

We use a VI-based approach to estimate the RVs posterior $p(\Omega|X)$ As a major advantage, this technique avoids heavy analytical derivation, as required by the Expectation-Maximization algorithm [17,23] VI approximates the true posterior with a parametric family of distributions $q_\phi(\Omega)$, also called the variational family. Maximizing the Evidence Lower BOund (ELBO) yields an estimate for the parameters ϕ:

$$\phi^\star = \arg\max_\phi \mathrm{ELBO}(q_\phi) \tag{1}$$

$$\text{where} \quad \mathrm{ELBO}(q_\phi) = \mathbb{E}_{\Omega\sim q}[\log p(\Omega, X)] - \mathbb{E}_{\Omega\sim q}[\log q_\phi(\Omega)] \tag{2}$$

However, this technique becomes computationally prohibitive when the volume of latent parameters and observations massively increases, as in neuroimaging applications. To handle this issue, we simplify the variational distribution $q_\phi(\Omega)$ by assuming factorization over the latent variables ω^i and by leveraging the conditional independence of ground RVs $(\omega^{i,k})_k$ within one or several plates $\mathcal{P} \in \text{Plates}(\omega^i)$:

$$\log q_\phi(\Omega) = \sum_{i=1}^{I} \sum_{n=0}^{N_i} \log q^{i,k}_{\phi^{i,k}}(\omega^{i,k}), \tag{3}$$

where I is the number of RV templates ω^i and N_i the total number of repeated ground RV $\omega^{i,k}$: $N_i = \prod_{\mathcal{P}\in\text{Plates}(\omega^i)} |\mathcal{P}|$ (see supplementary material, Appendix B).

By recalling that most of the distributions in the generative model $\mathcal{M}$ were designed as Gaussian ones, whose conjugate distributions belong to the same family, we select diagonal Gaussian distribution with mean $\alpha^{i,k}$ and standard deviation $\beta^{i,k}$ parameters for the surrogate posteriors $q^{i,k}_{\phi^{i,k}}$ (Mean Field VI [3]):

$$\omega^{i,k} \underset{q^{i,k}}{\sim} \mathcal{N}(\alpha^{i,k}, \beta^{i,k}) \tag{4}$$

Lastly, plate properties can be further exploited to reduce the size of the parameterization and increase convergence speed ([22]). To this end, we decompose the parameters $\phi^{i,k}$ controlling the surrogate posterior $q^{i,k}$ into two sets of parameters $(\psi^i, \boldsymbol{E}^{i,k})$. ψ^i are the weights shared by all the ground RVs $\omega^{i,k}$: this factorization is referred to as *plate amortization*([22]). The parameters $\boldsymbol{E}^{i,k}$, i.e. the *encodings*, capture the specificity of the ground RV $\omega^{i,k}$ among all its siblings $(\omega^{i,m})_{m\neq k}$. Thus, Eq. 4 becomes:

$$\omega^{i,k} \underset{q^{i,k}}{\sim} \mathcal{N}(\alpha^i(\psi^i, \boldsymbol{E}^{i,k}), \beta^i(\psi^i, \boldsymbol{E}^{i,k})), \tag{5}$$

where the mean α^i and the standard deviation β^i are functions parameterized by the weights ψ^i (e.g., a neural network) and conditioned by the encodings $\boldsymbol{E}^{i,k}$. Finally, $q^{i,k}_{\phi^{i,k}}$ is trained over the observed data using stochastic VI over the plates [13,22].

3.3 Datasets and preprocessing

Our work is based on two datasets: the HCP S1200 dataset [29] and the CamCAN dataset [25]. We relied on the original pre-processing for the former and on fMRIPrep [8]

pipeline for the latter. We performed additional steps to produce the functional connectomes for every scan, as detailed in the supplementary material, Appendix C.

Note that the connectivity, measured as the Pearson correlation between two signals, was computed between every surface vertex and every region defined by the DiFuMo atlas [5] (with a number of regions $d = 128$). In addition, to make the parcellation [27] robust with respect to the noise and make the data as site-invariant as possible, we binarized the connectome by keeping the top 20% connectivity values.

3.4 Predicting behavioral scores as validation task

Without a ground truth, validating the obtained result becomes challenging. We choose an extrinsic validation involving downstream tasks, more adapted to the large-dimensionality regime, and more data-driven [4]. This task is based on predicting individual behavioral test scores provided along with the experimental datasets. The regression (using ridge regression) uses as input features the inferred RVs obtained with our method presented in section 3.2, such as the individual-specific RSFC profiles or parcellation. We measure prediction accuracy with the Pearson correlation coefficient r between the predicted values and the ground truth on a validation set of subjects. To avoid bias due to validation set selection, we perform a 10-fold cross-validation repeated 20 times with different foldings. This prediction accuracy is systematically computed for 13 behavioral measures [17] for the HCP dataset and 13 other behavioral measures for the CamCAN dataset [7] (see supplementary material, Appendix D).

We use the global accuracy, averaged across measures, as an indicator to assess the performance of an inference, considered as the predictive power of the inferred individual variables. As a first step, we compute the learning curves of our model for different sizes of cohorts and different input features (see Fig. 2). We compare the performance to a mere baseline, that is, the mean RSFC profiles over the regions provided by Yeo's atlas [27], a discrete population-level atlas based on 1,000 subjects. Then, to assess TL efficiency, we drew the same learning curves for the individual soft parcellation, with and without knowledge transfer.

3.5 Assessing Transfer Learning performance

As mentioned in section 2, TL with Bayesian graphical models involves sharing part or all of the parameters inferred in the source domain to initialize a new model in the target domain. Here, we transfer all the $(\psi^i, \boldsymbol{E}^{i,k})$ parameterizing the variational families $q^{i,k}_{\phi^{i,k}}$. Besides this warm start, to avoid a catastrophic forgetting [15] that would hamper generalization, we preserve the prior information contained in the population-level RVs by freezing the global RSFC profiles $\mu^g_{l,d}$ and soft parcellation θ^n_l. It is important to note that this operation may be sensitive to domain shift. Still, we consider the population-level topography as site-agnostic, and we recall that we eliminate, during pre-processing, amplitude information in connectivity values to reduce the impact of the acquisition site on the functional connectomes.

To assess TL performance, as shown in Fig. 1, we divided 1,008 subjects of the HCP dataset into a large set of 768 individuals with 4 scans per subject (the source dataset)

and smaller ones ranging from 48 to 240 subjects with 1 scan per subject, to simulate the features of a clinical dataset (the target dataset). We pre-trained a model with the source dataset and used the inferred features to transfer the knowledge to a new model trained on the out-of-sample dataset. We repeated the whole process with the CamCAN dataset as the target domain.

4 Results – Discussion

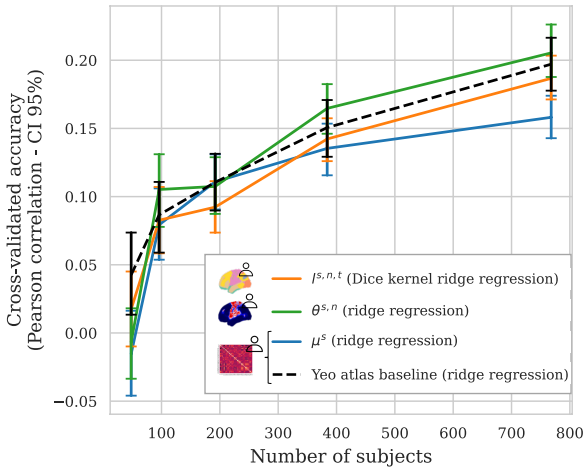

Fig. 2. Learning curves w.r.t. the size of a cohort sampled from the source dataset (from 48 to 768 patients). The solid lines represent 3 individual features inferred from our method: the discrete and probabilistic parcellations (we fixed the number of networks to 17 [27]), and the RSFC profiles at subject and scan levels. The error bars show the accuracy variability w.r.t. subsample composition (10 iterations were performed for each sample size).

Probabilistic network topography is more predictive than connectivity strength. The learning curves in Fig. 2 show the paramount importance of dataset size in inferring meaningful features. With less than 100 subjects, predictions are not connected with the ground truth. Then, the accuracy increases with the size of the cohort, but more in a logarithmic rate than a linear one. The TL is expected to overcome this significant difficulty. RSFC profiles and spatial parcellations exhibit predictive power of identical magnitude, in line with the Yeo atlas' baseline, showing that both aspects of the generative model learn inter-subject variability. Network topography, though, displays higher accuracy than inferred connectivity strength, especially when considering the probabilistic parcellation. It might be due to this feature being more informative, e.g., by capturing regions overlapping [2]. Interestingly, we note the good performance of a mere baseline based on RSFC profiles extracted from a proven population-level atlas. Even if this technique does not yield subject-level parcellation, it still can remain relevant for downstream tasks leveraging functional connectomes.

Transfer learning improves the predictive power of inferred features. As plotted in Fig.1B, the effects of the TL are mainly twofold. For a small dataset with less than 200 subjects, the variability related to cohort composition shrinks, proving a better generalization. Second, it clearly improves the predictive power of the inferred features. For the HCP dataset and 200 subjects, with 1 scan per subject in the target domain, the

performance with pre-training is twice as good as without pre-training, and scores are comparable with those of the source domain dealing with 4 scans. For the CamCAN dataset, the improvement is even better: e.g., the performance with 50 subjects with pre-training is comparable to the one with 150 subjects without. Nevertheless, the discrepancy between with and without TL scores tends to reduce along with the sample size, owing to the effect of the Bayesian regularization via the global RVs μ^g and $\theta^{s,n}$.

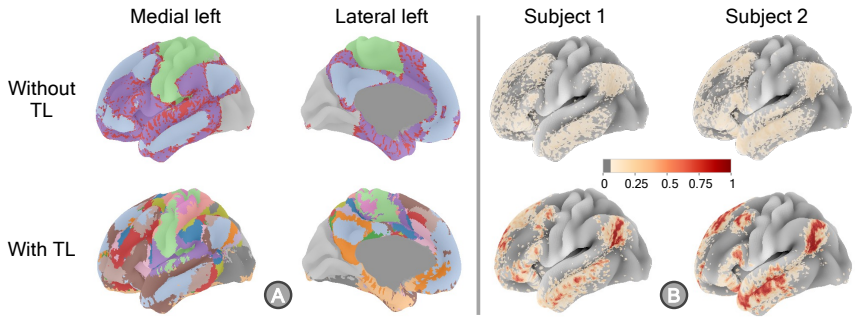

Fig. 3. Panel A - Effect of the TL on the discrete parcellation of a given subject's left hemisphere (HCP target dataset). TL enables the differentiation of networks within larger networks, reflecting the hierarchical organization of the brain [21]. **Panel B** - Example of soft parcellation by focusing on a default mode network and by comparing two subjects, with and without TL. We see here that the network differentiation leads to a more precise localization of the networks and a better inter-subject separation.

Transfer learning enhances the quality of individual parcellation. As illustrated in Fig.3, inference with pre-training provides finer-grained parcellation. The networks are now clearly drawn, showing less noise and more inter-individual contrast. These results, linking with the hierarchical brain organization, have to be further investigated and might represent an interesting research avenue for our method.

5 Conclusion

As mentioned in section 3.1, we essentially rely on the Bayesian graphical model, a suited representation to design and share generative models. Ours, derived from the MS-HBM, is composed of latent variables that are easily interpretable, such as the strength of the connectivity between parts of the cortex or brain topography. Therefore, every transfer of knowledge can be neurologically motivated and consequently tested with our method. However, this inductive bias driven by the generative model can also show limits. For example, the population-level parcellation θ_l^n, which provides regularisation over the individual parcellations, might be counterproductive if one learns it on healthy subjects and applies it to patients with, e.g., brain tumors. Even though our method, thanks to its flexibility, can fit new models in an automatized way and, with the Bayesian framework, allows the investigator to transfer the information he considers relevant to the most appropriate place in the target model.

References

1. Ardalan, Z., Subbian, V.: Transfer Learning Approaches for Neuroimaging Analysis: A Scoping Review. Frontiers in Artificial Intelligence **5**, 780405 (Feb 2022)
2. Bijsterbosch, J.D., Beckmann, C.F., Woolrich, M.W., Smith, S.M., Harrison, S.J.: The relationship between spatial configuration and functional connectivity of brain regions revisited. eLife **8**, e44890 (May 2019)
3. Blei, D.M., Kucukelbir, A., McAuliffe, J.D.: Variational Inference: A Review for Statisticians. Journal of the American Statistical Association **112**(518), 859–877 (Apr 2017)
4. Bzdok, D., Yeo, B.T.: Inference in the age of big data: Future perspectives on neuroscience. NeuroImage **155**, 549–564 (Jul 2017)
5. Dadi, K., Varoquaux, G., Machlouzarides-Shalit, A., Gorgolewski, K.J., Wassermann, D., Thirion, B., Mensch, A.: Fine-grain atlases of functional modes for fMRI analysis. NeuroImage **221**, 117126 (Nov 2020)
6. Eickhoff, S., Yeo, B.T., Genon, S.: Imaging-based parcellations of the human brain. Nature Reviews Neuroscience **19** (Oct 2018)
7. Engemann, D.A., Kozynets, O., Sabbagh, D., Lemaître, G., Varoquaux, G., Liem, F., Gramfort, A.: Combining magnetoencephalography with magnetic resonance imaging enhances learning of surrogate-biomarkers. eLife **9**, e54055 (May 2020)
8. Esteban, O., Markiewicz, C.J., Blair, R.W., Moodie, C.A., Isik, A.I., Erramuzpe, A., Kent, J.D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S.S., Wright, J., Durnez, J., Poldrack, R.A., Gorgolewski, K.J.: fMRIPrep: A robust preprocessing pipeline for functional MRI. Nature Methods **16**(1), 111–116 (Jan 2019)
9. Fang, Y., Wang, M., Potter, G.G., Liu, M.: Unsupervised cross-domain functional MRI adaptation for automated major depressive disorder identification. Medical Image Analysis **84**, 102707 (Feb 2023)
10. Gao, Y., Zhang, Y., Cao, Z., Guo, X., Zhang, J.: Decoding Brain States From fMRI Signals by Using Unsupervised Domain Adaptation. IEEE Journal of Biomedical and Health Informatics **24**(6), 1677–1685 (Jun 2020)
11. Gao, Y., Zhang, Y., Wang, H., Guo, X., Zhang, J.: Decoding Behavior Tasks From Brain Activity Using Deep Transfer Learning. IEEE Access **7**, 43222–43232 (2019)
12. Gelman, A.: Bayesian Data Analysis. Chapman & Hall/CRC Texts in Statistical Science, CRC Press, Boca Raton, third edition edn. (2014)
13. Hoffman, M., Blei, D.M., Wang, C., Paisley, J.: Stochastic Variational Inference (Apr 2013)
14. Huang, S., Li, J., Chen, K., Wu, T., Ye, J., Wu, X., Yao, L.: A transfer learning approach for network modeling. IIE Transactions **44**(11), 915–931 (Nov 2012)
15. Iman, M., Arabnia, H.R., Rasheed, K.: A Review of Deep Transfer Learning and Recent Advancements. Technologies **11**(2), 40 (Mar 2023)
16. Koller, D., Friedman, N.: Probabilistic Graphical Models: Principles and Techniques. Adaptive Computation and Machine Learning, MIT Press, Cambridge, MA (2009)
17. Kong, R., Li, J., Orban, C., Sabuncu, M.R., Liu, H., Schaefer, A., Sun, N., Zuo, X.N., Holmes, A.J., Eickhoff, S.B., Yeo, B.T.T.: Spatial Topography of Individual-Specific Cortical Networks Predicts Human Cognition, Personality, and Emotion. Cerebral Cortex **29**(6), 2533–2551 (Jun 2019)
18. Kong, R., Yang, Q., Gordon, E., Xue, A., Yan, X., Orban, C., Zuo, X.N., Spreng, N., Ge, T., Holmes, A., Eickhoff, S., Yeo, B.T.T.: Individual-Specific Areal-Level Parcellations Improve Functional Connectivity Prediction of Behavior. Cerebral Cortex **31**(10), 4477–4500 (Aug 2021)
19. Kouw, W.M., Ørting, S.N., Petersen, J., Pedersen, K.S., de Bruijne, M.: A cross-center smoothness prior for variational Bayesian brain tissue segmentation (2019)

20. Mandonnet, E., Vincent, M., Valero-Cabré, A., Facque, V., Barberis, M., Bonnetblanc, F., Rheault, F., Volle, E., Descoteaux, M., Margulies, D.S.: Network-level causal analysis of set-shifting during trail making test part B: A multimodal analysis of a glioma surgery case. Cortex **132**, 238–249 (Nov 2020)
21. Meunier, D.: Hierarchical modularity in human brain functional networks. Frontiers in Neuroinformatics **3** (2009)
22. Rouillard, L., Bris, A.L., Moreau, T., Wassermann, D.: PAVI: Plate-Amortized Variational Inference (2023)
23. Schaefer, A., Kong, R., Gordon, E.M., Laumann, T.O., Zuo, X.N., Holmes, A.J., Eickhoff, S.B., Yeo, B.T.T.: Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. Cerebral Cortex **28**(9), 3095–3114 (Sep 2018)
24. Suder, P.M., Xu, J., Dunson, D.B.: Bayesian Transfer Learning (2023)
25. Taylor, J.R., Williams, N., Cusack, R., Auer, T., Shafto, M.A., Dixon, M., Tyler, L.K., Cam-CAN, Henson, R.N.: The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) data repository: Structural and functional MRI, MEG, and cognitive data from a cross-sectional adult lifespan sample. NeuroImage **144**, 262–269 (Jan 2017)
26. Thomas, A.W., Müller, K.R., Samek, W.: Deep Transfer Learning For Whole-Brain fMRI Analyses (2019)
27. Thomas Yeo, B.T., Krienen, F.M., Sepulcre, J., Sabuncu, M.R., Lashkari, D., Hollinshead, M., Roffman, J.L., Smoller, J.W., Zöllei, L., Polimeni, J.R., Fischl, B., Liu, H., Buckner, R.L.: The organization of the human cerebral cortex estimated by intrinsic functional connectivity. Journal of Neurophysiology **106**(3), 1125–1165 (Sep 2011)
28. Valverde, J.M., Imani, V., Abdollahzadeh, A., De Feo, R., Prakash, M., Ciszek, R., Tohka, J.: Transfer Learning in Magnetic Resonance Brain Imaging: A Systematic Review. Journal of Imaging **7**(4), 66 (Apr 2021)
29. Van Essen, D., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T., Bucholz, R., Chang, A., Chen, L., Corbetta, M., Curtiss, S., Della Penna, S., Feinberg, D., Glasser, M., Harel, N., Heath, A., Larson-Prior, L., Marcus, D., Michalareas, G., Moeller, S., Oostenveld, R., Petersen, S., Prior, F., Schlaggar, B., Smith, S., Snyder, A., Xu, J., Yacoub, E.: The Human Connectome Project: A data acquisition perspective. NeuroImage **62**(4), 2222–2231 (Oct 2012)
30. Wang, M., Zhang, D., Huang, J., Yap, P.T., Shen, D., Liu, M.: Identifying Autism Spectrum Disorder With Multi-Site fMRI via Low-Rank Domain Adaptation. IEEE Transactions on Medical Imaging **39**(3), 644–655 (Mar 2020)
31. Yeo, B.T.T., Eickhoff, S.B.: A modern map of the human cerebral cortex. Nature **536**(7615), 152–154 (Aug 2016)
32. Zhang, H., Chen, P.H., Ramadge, P.J.: Transfer learning on fMRI datasets. pp. 595–603 (Jan 2018)

Improving Individual-Specific Functional Parcellation Through Transfer Learning - Supplementary Material

Appendix A - Description of the random variables in the generative model MS-HBM

Spatial RVs $\theta_l^{s,n}$ and θ_l^n represent soft parcellation, that is, the probability of a voxel belonging to a network: $\theta_l^{s,n}$ is the probability that the voxel n of the subject s belongs to the network l. The network label $l^{s,t,n}$ is a discrete RV sampled from the logits $\theta_l^{s,n}$.

The connectivity RV $\mu_{l,d}$ encode the RSFC profiles of every network l with d regions of interest (specifically the 128 regions of the DiFuMo atlas [5], see section 3.3) More specifically, $\mu_{l,d}^g$ stands for the population level, $\mu_{l,d}^s$ for the subject s and $\mu_{l,d}^{s,t}$ for the session t of the subject s. As for the spatial RVs, each higher-level RV regularizes the lower-level one. Lastly, the observed RV $X_d^{s,t,n}$ follows a mixture model, with the components $\mu_{l,d}^{s,t}$ and the category $l^{s,t,n}$.

The scale RVs ϵ, σ, κ and γ represent the variability of the previous RVs and follow a LogNormal distribution to ensure the positivity of the support. In particular, ϵ captures the inter-subject RSFC variability and σ the intra-subject RSFC variability.

Appendix B - Reminder about plates, RV template, and ground RVs.

As depicted in Fig. 1A, most of the RVs are located within one or several plates The plate notation indicates that the RVs within it are repeated k times, where k is the cardinality of the plate. It allows a simple representation for, e.g., hierarchical models like MS-HBM When an RV is located in multiple plates, the number of repetitions is the product of the plates' cardinality For instance, $\theta_l^{s,n}$, as indicated by the upper indices, is included in plates S (subject) and N (voxel) and thus is repeated $|S|.|N|$ times Mathematically, the plate denotes conditional independence of the repeated RVs. As a convention [16], we denote $(\omega^{i,k})_k$ the set of ground RVs, i.e., the set of k independent RVs following the same distribution as the RV template ω^i All the ω^i are grouped into the set of latent variables Ω.

Appendix C - Description of the datasets and the pre-processing steps

The HCP S1200 dataset [29] gathers imaging acquisitions from 1,200 young adults (ages 22-35). The resting-state fMRI data were collected during two sessions, one day after the other, during which two runs of 15 minutes each were recorded for every subject. We selected 1,008 subjects with four runs each and retained all the runs, regardless

of volumes with high FD or DVARS. The data underwent the HCP minimal preprocessing pipeline and the ICA-FIX denoising.

The CamCAN dataset [25] contains MRI, and cognitive-behavioural data from a large (approximately N = 700), cross-sectional adult lifespan (18–87 years old) population-based sample. The resting-state fMRI data were collected during a single session of 8 minutes and 40 seconds. The provided data were preprocessed using the fMRIPrep pipeline [8].

Volume signals from both datasets underwent additional preprocessing steps:

- Cleaning process, including detrending, low- and high-pass filtering (between 0.01 et 0.1 Hz), confounds removal and standardization.
- Smoothing with a Gaussian filter ($\mathrm{fwhm} = 4\,\mathrm{mm}$).
- Projecting the volume signal into the fsLR reference surface mesh.
- Computing the Pearson's coefficient between the 64,984 voxels' signal and the mean signal of the 128 DiFuMo regions. The obtained matrix is the functional connectome used as input feature byb our model.

Appendix D - Behavioral measures

HCP measures	CamCAN measures
Visual Episodic Memory	Benton Faces
Cognitive Flexibility (DCCS)	Fluid Intelligence
Inhibition (Flanker task)	Emotion Expression Recognition
Fluid Intelligence (PMAT)	Famous Faces
Reading (Pronunciation)	Hotel Task
Vocabulary (Picture Matching)	Picture Priming
Processing Speed	Proverb Comprehension
Delay Discounting	Sentence comprehension (unacceptable error)
Spatial Orientation	Sentence comprehension (reaction time)
Sustained Attention – Sens.	Visual short term memory (mean)
Sustained Attention – Spec.	Visual short term memory (precision)
Verbal Episodic Memory	Visual short term memory (doubt)
Working Memory (List Sorting)	Visual short term memory (MSE)

Table 1. Behavioral measures are used for prediction as a validation task. HCP measures are directly available in the dataset, while CamCAN measures require pre-processing, as detailed in Engelmann et al. [7]

Title Suppressed Due to Excessive Length