

---

# Region-network hierarchical sparsity priors for high-dimensional inference in brain imaging

---

Danilo Bzdok, Michael Eickenberg, Gaël Varoquaux, Bertrand Thirion

Department of Psychiatry, Psychotherapy and Psychosomatics, RWTH Aachen, Germany

INRIA, Parietal team, Saclay, France

CEA, Neurospin, Gif-sur-Yvette, France

firstname.lastname@inria.fr

## Abstract

**keywords:** Sparsity-inducing Norms, Structured Sparsity, Numerical Optimization, Systems Neuroscience, Brain Imaging, Functional Specialization, Functional Integration

## 1 Introduction

Many quantitative scientific domains underwent a recent passage from the classical regime (i.e., "long data") to the high-dimensional regime (i.e., "wide data") [33]. Also in the brain imaging domain, many contemporary methods for acquiring brain signals yield more variables per observation than total observations available in a data sample. This scenario challenges many statistical estimators from classical statistics. For instance, linear regression models without additional assumptions yield an infinity of possible coefficients and, thus, no solution. *Sparsity* assumptions have consequently been introduced and made many ill-posed estimation problems tractable [11, 25]. Domain-specific structure can be imposed on the statistical estimation, thus preassuming that variables have unequal importance and follow existing knowledge [3]. Sparsified supervised and unsupervised learning algorithms have proven to yield statistical relationships that can be readily estimated, reproduced, and interpreted [23]. Yet, what neurobiological structure suggests itself to successfully tackle the *curse of dimensionality* in functional neuroimaging research?

Concepts on human brain organization have long been torn between the two extremes *functional specialization* and *functional integration*. Functional specialization emphasizes that microscopically distinguishable brain regions probably solve distinct classes of computational processes [35]. Functional integration, in turn, emphasizes that brain function is probably enabled by complex connections between these distinct brain regions [49]. These notions were predominantly derived from invasive examination of *structure* (i.e., histological preparation), *connectivity*, (i.e., axonal tracing), and *functional properties* (i.e., single cell recordings) in same animals. Regarding functional segregation into specialized regions, early histological investigations into the microscopic heterogeneity of the human cerebral cortex have resulted in several detailed anatomical maps [9, 54]. Regarding axonal connections, each such cortical area has been observed to possess a unique set of incoming and outgoing connections [40, 56, 45]. Both local cyto- and chemoarchitectonic infrastructure and its unique global connectivity profile together realize computational processes (i.e., function). While cortical modules versus connections between those reflect functional specialization versus functional integration [21, 38], both architectural principles are conceptually inextricable when explaining the emergence of mental operations [52, 44].

Functional specialization has been explored and interpreted based on many different research methods. Single cell recordings and microscopic examination revealed, for instance, the specialization in the visual cortex into V1, V2, V3, V3A, and V4 [27, 58]. Tissue lesion of the mid-fusiform

gyrus of the visual system were frequently reported to impair recognition of others' identity from faces [28]. The localization of sensory, motor, and emotional cognitive functions to cortical areas in the living brain has later been enabled by non-invasive brain imaging with functional magnetic resonance imaging (fMRI) and positron emission tomography (PET) [20]. Further, radioactive mapping of neurotransmitter receptor distributions rendered accessible yet another local characteristic of neuronal populations [60]. In the computational era, automatic clustering methods are increasingly employed to regionally differentiate the cerebral cortex, which can partly be more fine-grained than classical cytoarchitectonic borders [7, 18]. Today, high-throughput computing enables ultra-high-resolution 3D models of brain anatomy at macroscopical to near-cellular scale [2]. As a crucial common point, all these methodological approaches yield neuroscientific findings that are naturally interpreted according to non-overlapping, discrete region compartments as the basic architectonic framework of brain organization.

It is more recent that the main interpretational focus has shifted from circumscribed regions to network stratifications in systems neuroscience [57, 51]. The higher the processing hierarchy level, the more a cortical area is known to connect to different large-scale networks [55]. Invasive axonal tracing studies in monkeys were complemented by in-vivo diffusion MRI tractography in humans as a now frequently employed method to outline fiber bundles between brain regions [29]. Besides analyses of electrophysiological oscillations [13] and graph-theoretical properties [12], studies of functional connectivity [10] and independent component analysis (ICA) [6] became the workhorses of network discovery in functional neuroimaging. These revealed the important implication of canonical brain networks across cognitive domains, including the so-called "default-mode network" [42], "salience network" [46], and "dorsal attention network" [15]. Characteristic changes in the configuration of these large-scale networks were repeatedly observed to be induced by the onset of given cognitive tasks [19]. Such task-induced mechanisms orchestrating supraordinate networks might be subserved by the right anterior insula [50] and temporo-parietal junction [14]. Ultimately, interpretation of findings from all these methods naturally embraces cross-regional integration by overlapping network compartments as the basic architectonic framework of brain organization, in stark contrast to methods examining regional specialization.

Building on these interpretational traditions, the present study proposes to incorporate known architectonic aspects from both anatomical segregation and integration into unified classification and regression algorithms by means of structured sparsity assumptions. The "true" relative importance of local region compartments and global network compartments is typically unknown but probably varying in degree across diverse neuroscientific questions. Recent advances in multivariate statistical learning techniques enable such adaptive estimators with neurobiologically plausible region and network priors. Using a large reference dataset, we demonstrated that domain-informed supervised models gracefully handle the curse of dimensionality, yield many more human-interpretable results, and generalize better to new samples than domain-naïve models.

higher-order prior knowledge

structured sparsity extend trend in statistical modelling from processing of auditory signals [17], natural images [24] and videos [34, 37] to genetics [43, 36], astrophysics [53], and conformational dynamics of protein complexes [32].

structured-sparsity-inducing penalty

to brain imaging domain.

## 2 Methods

**Rationale** we need to inject domain knowledge into statistical estimations to harness the curse of dimensionality. two neurobiological design principles

imposing parsimony integrative processes

This L1/L2 norm for group lasso has been extended to a more general setting to designed groups the child nodes enter the set of relevant inputs only if its parent node does.

should be able to estimate voxel level while taking into account known supravoxel structure. is instrumental in Developmentally, such large-scale networks emerge during late fetal growth (Doria et al., 2010), before cognitive capacities mature in childhood.

In adults, nodes of a same cohesive network have more similar functional profiles than nodes from different networks (Anderson et al., 2013).

data exhibit natural correlations between neighboring voxels forming clusters

representing some phenomenon with as few variables as possible

neurobiologically motivated restrictions to complexity circumvented the curse of dimensionality three-dimensional spatial arrangement that respects the functional anatomy of the brain not ignore the spatial configuration

incorporate rich prior knowledge

If meaningful structures exist, we show that one can take advantage of such structures

Statistically,  $\ell_1$  and  $\ell_2$  are local sparsity priors - $\ell_i$  resulting sparsity does yield structure we want to privilege representations with structure

= $\ell_i$  a biologically and statistically desirable bias

**Problem formulation** Sparse linear models encode geometric prior information topology local sets of voxels

Group-sparsity is a first step towards the more general idea that a regularization function can encourage sparse solutions with a particular structure.

it is not realistic to assume that all of the tasks share the same set of relevant inputs as in the L1/L2-regularized regression. A subset of highly related outputs may share a common set of relevant inputs, whereas weakly related outputs are less likely to be affected by the same inputs.

structured regularization We might therefore gain in the quality of the factors induced by enforcing directly this a priori

groups at multiple granularity

tree-guided group lasso

encourage structured shrinkage effect

$\ell_1$  = unstructured sparsity-inducing penalty

Our method extends the L1/L2 penalty to the tree-lasso penalty by letting the hierarchically-defined groups overlap. the tree lasso is a special case of overlapping group lasso

for every column  $u$  of  $U$ , it compute a column  $v$  of  $V$  solving

we aim at learning a weight vector  $w \in \mathbb{R}^p$  and an intercept  $b \in \mathbb{R}$  such that the prediction of  $y$  can be based on the value of  $w^T x + b$ .

We omit a bias term, since the data were mean-centered and unit-variance scaled. The scalar  $b$  is not particularly informative

however the vector  $w$  corresponds to a volume that can be represented in brain space as a volume

hierachial tree = more generally into a directed acyclic graph

more precisely, we denote by  $X \in \mathbb{R}^{n \times p}$  the design matrix assembled from  $n$  fMRI volumes and by  $y \in \mathbb{R}^n$  the corresponding  $n$  targets. In other words, each row of  $X$  is a  $p$ -dimensional sample, i.e., an activation map of  $p$  voxels related to one stimulus presentation. for visualization of the predictive pattern of voxels.

Learning the parameters ( $w, b$ ) remains challenging since the number of features (104 to 105 voxels) exceeds by far the number of samples (a few hundreds of volumes).

The scalar  $b$  is not particularly informative, however the vector  $w$  corresponds to a volume that can be represented in brain space as a volume for visualization of the predictive pattern of voxels.

each row of  $\mathbf{X}$  is a  $p$ -dimensional sample, i.e., an activation map of  $p$  voxels related to one stimulus presentation.

To address this issue, dimensionality reduction attempts to find a low dimensional subspace that concentrates as much of the predictive power of the original set as possible for the problem at hand. - $\zeta$  we do not want to do preliminary feature selection or dimensionality reduction or feature agglomeration because we want to fit one model parameter to each brain voxel for maximal interpretability This corresponds to discarding some columns of  $\mathbf{X}$ .

The essential shortcoming of the Elastic net is that it does not take into account the spatial structure of the data, which is crucial in this context

Craddock clusters are often used for feature agglomeration into parcels - $\zeta$  exploits only a part of the data

dual-level spatial structure sparse hierarchical regularization structured sparsity-inducing regularization the root of the tree  $T$  is the unique cluster that gathers all the voxels,

It is a generalization of the traditional  $\ell_1$ -norm  $\Omega(\mathbf{w}) = \sum_{j=1}^p |\mathbf{w}_j|$  ignores structure

[31]

*structured sparsity*

[26, 39, 30]

a node  $j$  of  $\mathcal{T}$ , we denote by  $g_j \subseteq \{1, \dots, q\}$  the set of indices that record all the descendants of  $j$  in  $\mathcal{T}$

the family of sparsity-inducing norms has recently been extended by hierarchical sparsity penalty terms [59].

$$\Omega(\mathbf{w}) = \sum_{g \in G} \|\mathbf{w}_g\|_2 = \sum_{g \in G} \sqrt{\sum_{j \in g} \mathbf{w}_j^2} \quad (1)$$

For example, when  $G$  is the set of all singletons, is the usual  $l_1$  norm (assuming that all the weights are equal to 1).

$l_1/l_2$  mixed norm is convex

Discarding coefficients belonging to a network group will naturally enforce discarding the coefficients belonging to each of its descendent region groups. Conversely, variable selection of a network group will also enforce selection of all voxel of its descendent group regions. Single region groups can however be set to zero (unselected) or non-zero (selected) without analogous effect on the parent network group.

At the between-group level,...  $\Omega$  exerts  $\ell_1$ -like variable selection on the  $(-\mathbf{w}_g\|_2)_g$  groups, yielding a maximum of  $g \in G$  to be zeroed out [30]. The important consequence is that also all descendents of such a zeroed group  $g \in G$  will be discarded. Conversely, if one group  $g$  is selected, then all the ancestral groups will also be selected. Thus, statistical estimation will be improved by enticing entire voxel sets to be selected or discarded as predictive, although one individual coefficient is computed for each voxel.

- $\zeta$  it is a  $(\ell_1, \ell_2)$ -mixed norm - $\zeta$  between-group sparsity effect by  $l_1$  - $\zeta$  within-group shrinkage effect by  $l_2$

$$\Omega(\mathbf{w}) = \sum_{g \in G} \eta_g \|\mathbf{w}_g\|_2 \quad (2)$$

$(\eta_g)_{g \in G}$  are positive weights for the groups

fit to the data is measured through a convex loss function  $(\mathbf{w}, \mathbf{b}) \rightarrow L(\mathbf{y}, \mathbf{X}, \mathbf{w}, \mathbf{b})$   $R+$ .

**Classification** logistic loss function

$$P(y = k|x, W, b) = \frac{\exp\{x^T w^k + b_k\}}{\sum_{m=1}^c \exp\{x^T w^m + b_m\}}$$

$$\arg \min \frac{1}{2} \|u - v\|_2^2 + \lambda \Omega(\mathbf{w}) \quad (3)$$

$$\lambda > 0.$$

bias is omitted because  $\mathbf{X}$  and  $\mathbf{y}$  are mean-centered and unit-variance scaled.

In this setting, and given a new fMRI volume  $\mathbf{x}$ , we make predictions by choosing the label that maximizes the class-conditional probabilities (3.1), that is,  $\text{argmax}_{k=1,\dots,c} \text{Prob}(y = k | \mathbf{x}; \mathbf{W}, \mathbf{b})$

One-versus-rest scheme

**Regression** squared error as loss

$$\arg \min \frac{1}{2} \|\mathbf{y} - \mathbf{X}\mathbf{w}\|_2^2 + \lambda \Omega(\mathbf{w}) \quad (4)$$

$$\lambda > 0.$$

Prediction for a new fMRI volume  $\mathbf{x}$  is then simply performed by computing the dot product  $\mathbf{x}^T \mathbf{w}$

**Numerical optimization** Difficult because high-dimensional setting

empirical risk minimization was performed by

The intercept  $b$  is left unregularized

**Implementation.** The analyses were performed in Python. We used *nilearn* to handle the large quantities of neuroimaging data [1] and *Theano* for automatic, numerically stable differentiation of symbolic computation graphs [5, 8]. All Python scripts that generated the results are accessible online for reproducibility and reuse (<http://github.com/banilo/nips2015>).

**Data.** As the currently biggest openly-accessible reference dataset, we chose resources from the Human Connectome Project (HCP) [4]. Neuroimaging task data with labels of ongoing cognitive processes were drawn from 500 healthy HCP participants (cf. Appendix for details on datasets). 18 HCP tasks were selected that are known to elicit reliable neural activity across participants (Table 1). In sum, the HCP task data incorporated 8650 first-level activity maps from 18 diverse paradigms administered to 498 participants (2 removed due to incomplete data). All maps were resampled to a common  $60 \times 72 \times 60$  space of 3mm isotropic voxels and gray-matter masked (at least 10% tissue probability). The supervised analyses were thus based on labeled HCP task maps with 79,941 voxels of interest representing z-values in gray matter.

These labeled data were complemented by unlabeled activity maps from HCP acquisitions of unconstrained resting-state activity [47]. These reflect brain activity in the absence of controlled thought. In sum, the HCP rest data concatenated 8000 unlabeled, noise-cleaned rest maps with 40 brain maps from each of 200 randomly selected participants.

We were further interested in the utility of the optimized low-rank projection in one task dataset for dimensionality reduction in another task dataset. To this end, the HCP-derived network decompositions were used as preliminary step in the classification problem of another large sample. The ARCHI dataset [41] provides activity maps from diverse experimental tasks, including auditory and visual perception, motor action, reading, language comprehension and mental calculation. Analogous to HCP data, the second task dataset thus incorporated 1404 labeled, grey-matter masked, and z-scored activity maps from 18 diverse tasks acquired in 78 participants.

Cognitive Task	Stimuli	Instruction for participants
1 Reward	Card game	Guess the number of a mystery card for gain/loss of money
2 Punish		
3 Shapes	Shape pictures	Decide which of two shapes matches another shape geometrically
4 Faces	Face pictures	Decide which of two faces matches another face emotionally
5 Random		
6 Theory of mind	Videos with objects	Decide whether the objects act randomly or intentionally
7 Mathematics	Spoken numbers	Complete addition and subtraction problems
8 Language	Auditory stories	Choose answer about the topic of the story
9 Tongue movement		Move tongue
10 Food movement		Squeezing of the left or right toe
11 Hand movement		Tapping of the left or right finger
12 Matching	Shapes with textures	Decide whether two objects match in shape or texture
13 Relations		Decide whether object pairs differ both along either shape or texture
14 View Bodies	Pictures	Passive watching
15 View Faces	Pictures	Passive watching
16 View Places	Pictures	Passive watching
17 View Tools	Pictures	Passive watching
18 Two-Back	Various pictures	Indicate whether current stimulus is the same as two items earlier

Table 1: Description of psychological tasks to predict.

sparse statistical models have only few nonzero parameters

### 3 Experimental Results

**Serial versus parallel structure discovery and classification.**

**Serial versus parallel structure discovery and classification.**

**Serial versus parallel structure discovery and classification.** DATARATIO

tell maps per percent

### 4 Discussion

exploit known structure for variable selection and prediction in classification and regression

Among a collection of plausible models, the simplest one is often preferred

Functional specialization and functional integration in the human brain are probably both mechanistically relevant for emerging behavior. However, there is a scarcity of statistical methods that concomitantly relate brain regions and networks to an organism's repertoire of mental operations.

However, in the problem of multi-task regression, where the output is a multivariate vector with an internal sparsity structure, the estimation of the regression parameters can potentially benefit from taking into account this sparsity structure in the estimation process. This will allow the strongly related output variables to be mapped to the input factors in a synergistic way, which is not possible in the standard lasso.

We assume regions and networks but can estimate importance of each voxel.

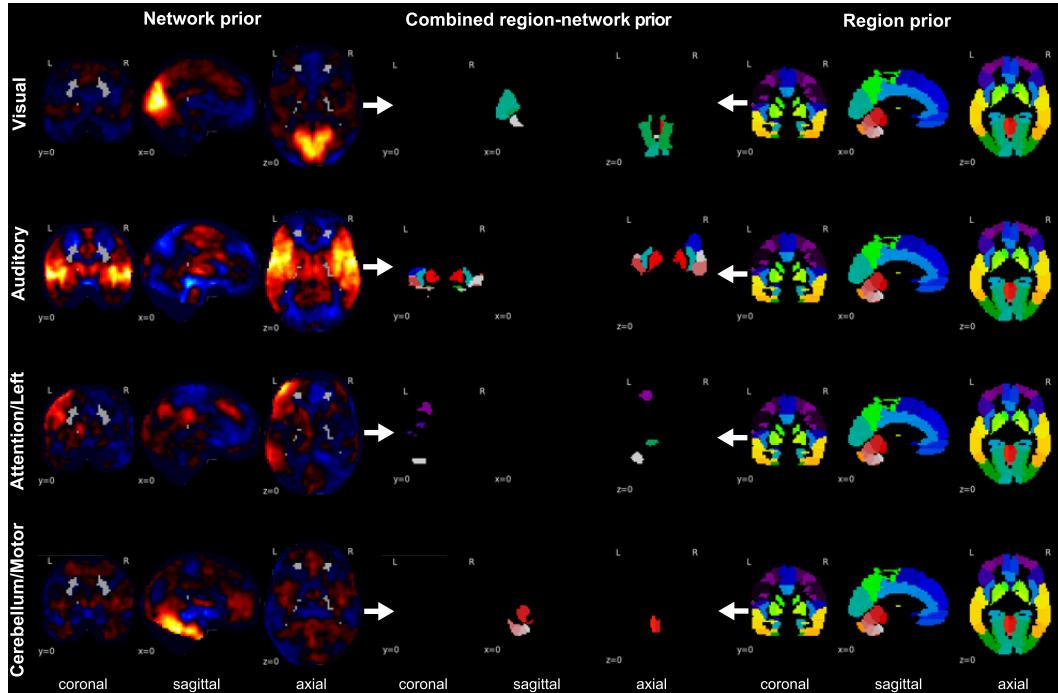
neurobiologically motivated restrictions to complexity circumvented the curse of dimensionality and allowed for useful, simplified views on brain function.

The same classification performance can be reached by many different models - we naturally want to privilege those among the equally performing models that are closest to what we know about the brain

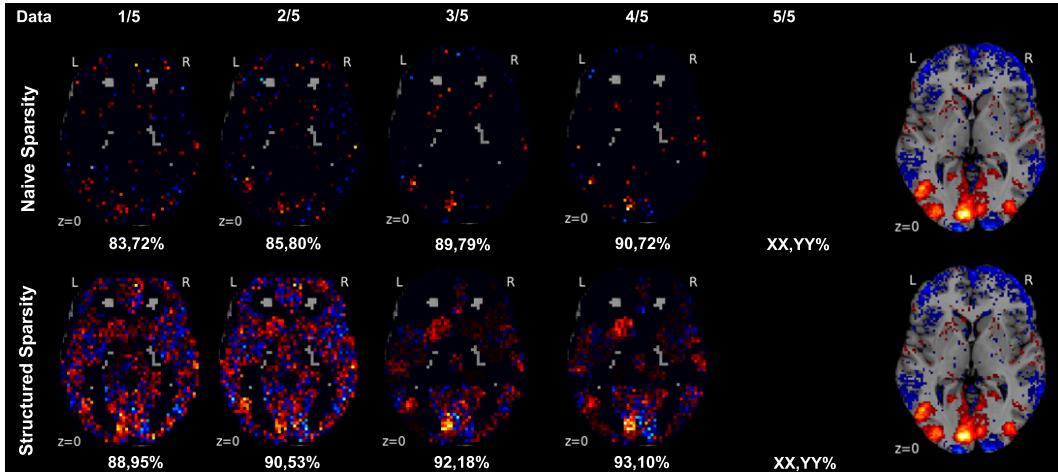
- interpretability and accuracy typically exclude each other (cf. Hastie)

alternatives would be dimensionality reduction and dictionary learning

Among a collection of plausible models, the simplest one should be preferred



**Figure 1: Building blocks of the region-network tree.** Depicts neurobiological priors introduced into supervised estimation by means of hierarchical structured sparsity. *Left:* Continuous, partially overlapping brain network priors (*hot-colored*, taken from [48]) accommodate the functional integration perspective of functional brain organization. *Right:* Discrete, non-overlapping brain region priors (*single-colored*, taken from [16]) accommodate the functional segregation perspective. *Middle:* These predefined voxel groups are incorporated into hierarchical priors of brain networks with their assigned brain regions. *Top to bottom:* Four exemplary region-network priors are shown. They include the early cortex that processes visual and sound information from the environment, a well-known attentional circuit in the left brain hemisphere, and the cerebellum that realizes motor behavior.



**Figure 2: Ordinary versus structured sparse model estimation as a function of training set size.** Ordinary  $\ell_1$ -penalized logistic regression (*upper row*) is compared to logistic regression estimation structured by hierarchical tree prior (*lower row*) with increasing fraction (*columns*) of the available training data. Unthresholded coronal maps of model weights are shown for comparison against the class-wise mean neural activity in the *rightmost column*. The 18-class out-of-sample performance in percent. In the data-scarce scenario, typical for common brain imaging studies, hierarchical tree sparsity achieves better support recovery with the biggest difference in model performance comparing to  $\ell_1$ -penalized logistic regression. In the data-rich scenario, neurobiologically informed logistic regression profits more from the increased information quantities than neurobiologically naive logistic regression.

curse of dimensionality has classically been circumvented by treating a high-dimensional problem as a mass-univariate problem. Problem was postponed to the inference step this is why we never solved the multiple comparisons problem in fMRI

regularization with sparsity penalties restrictions to complexity according to neurobiological facts  
 $\rightarrow$  bias-variance tradeoff: negligible increase in bias but large decrease in variance  
 $\rightarrow$  VC: decrease complexity capacity  
 $\rightarrow$  better chance for generalization out of sample

multiple comparisons problem has the curse of dimensionality as its close correspondence in statistical learning sparsity  
 $\rightarrow$  solves the curse at the estimation level

OUTLOOK: robust high-dimensional inference will be key to prediction in single individual, such as for personalized medicine [22]

More generally, our regularization scheme could also be used for various learning tasks, as soon as prior knowledge on the structure of the sparse representation is available, e.g., for multiple kernel learning (Micchelli and Pontil, 2006), multi-task learning (Argyriou et al., 2008; Obozinski et al., 2009; Kim and Xing, 2009) and sparse matrix factorization problems (Mairal et al., 2010; Jenatton et al., 2010).

this type of prior knowledge is of clear interest for the spatially and temporally structured data typical in bioinformatics, computer vision and neuroscience applications

structured sparsity can even be used for unsupervised data modelling, such as in PCA [32]

Differences to Bertrands structured sparsity paper: - they derive the tree from the data, we use a priori knowledge from the existing literature -

**Acknowledgment** The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 604102 (Human Brain Project). Data were provided by the Human Connectome Project. Further support was received from the German National Academic Foundation (D.B.), the German Research Foundation (BZ2/2-1 and BZ2/3-1 to D.B.), and the MetaMRI associated team (B.T., G.V.).

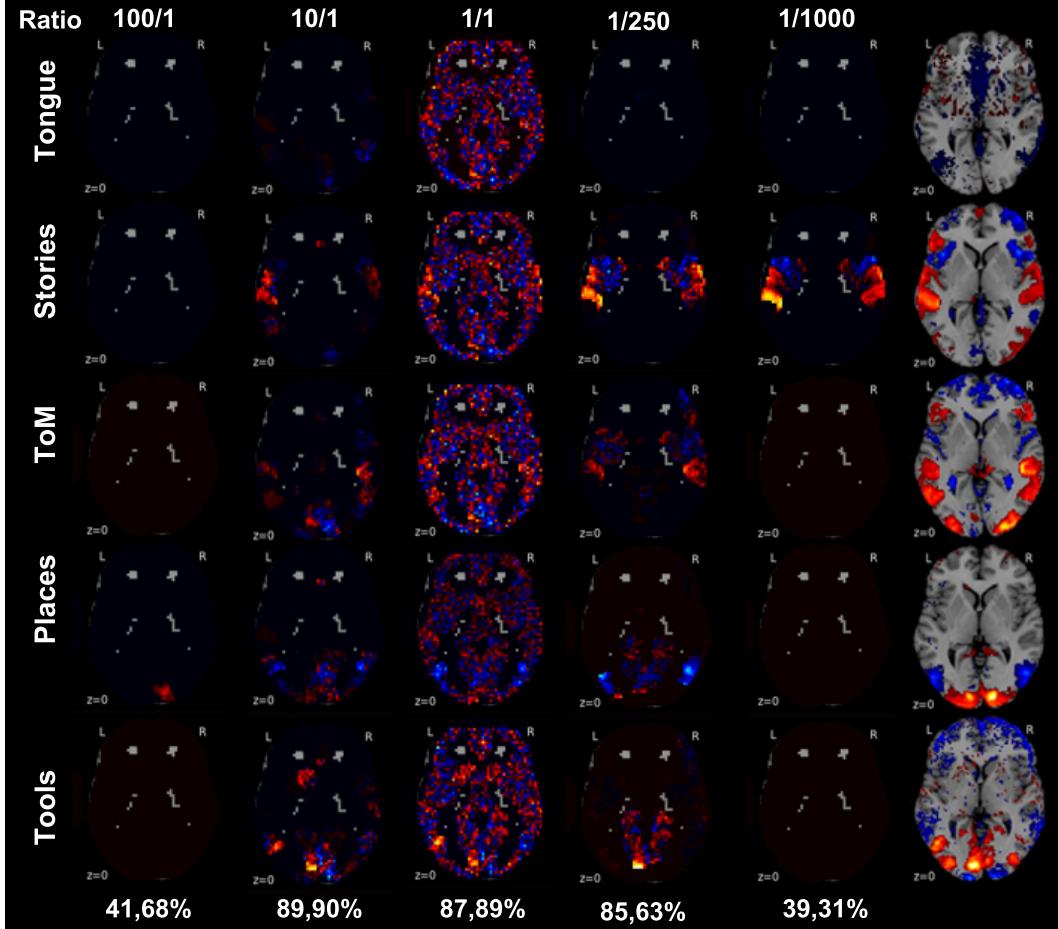


Figure 3: **Support recovery as a function of region and network emphasis.** The relative impact of the region and network priors on model selection is systematically varied against each other. This region-network ratio (*upper row*) weighted voxel groups to privilege sparse models in function space that acknowledge known data neighborhoods underlying brain region (*left columns*) or known compartments in the data that correspond to brain networks (*right columns*). Among the 18 classes, the model weights are shown for the tasks (*from top to bottom*) tongue movement, listening stories, taking somebody else’s perspective (ToM, ”theory of mind”), as well as perceiving locations and tools. The 18-class out-of-sample performance is indicated on the *bottom* and the class-wise mean neural activity in the *rightmost column*. Different emphasis on regions versus networks in hierarchical structured sparsity can yield essentially similar model performance. Privileging region versus network structure during model selection recovers complementary aspects of the brain activity pattern. Equal region and network emphasis yields more dispersed, less interpretable predictive model choices.

## References

- [1] Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., Varoquaux, G.: Machine learning for neuroimaging with scikit-learn. *Front Neuroinform* 8, 14 (2014)
- [2] Amunts, K., Lepage, C., Borgeat, L., Mohlberg, H., Dickscheid, T., Rousseau, M.E., Bludau, S., Bazin, P.L., Lewis, L.B., Oros-Peusquens, A.M., et al.: Bigbrain: an ultrahigh-resolution 3d human brain model. *Science* 340(6139), 1472–1475 (2013)
- [3] Bach, F., Jenatton, R., Mairal, J., Obozinski, G.: Optimization with sparsity-inducing penalties. *Foundations and Trends® in Machine Learning* 4(1), 1–106 (2012)
- [4] Barch, D.M., Burgess, G.C., Harms, M.P., Petersen, S.E., Schlaggar, B.L., Corbetta, M., Glasser, M.F., Curtiss, S., Dixit, S., Feldt, C.: Function in the human connectome: task-fMRI and individual differences in behavior. *Neuroimage* 80, 169–189 (2013)
- [5] Bastien, F., Lamblin, P., Pascanu, R., Bergstra, J., Goodfellow, I., Bergeron, A., Bouchard, N., Warde-Farley, D., Bengio, Y.: Theano: new features and speed improvements. arXiv preprint arXiv:1211.5590 (2012)
- [6] Beckmann, C.F., DeLuca, M., Devlin, J.T., Smith, S.M.: Investigations into resting-state connectivity using independent component analysis. *Philos Trans R Soc Lond B Biol Sci* 360(1457), 1001–13 (2005)
- [7] Behrens, T.E., Johansen-Berg, H., Woolrich, M.W., Smith, S.M., Wheeler-Kingshott, C.A., Boulby, P.A., Barker, G.J., Sillery, E.L., Sheehan, K., Ciccarelli, O., Thompson, A.J., Brady, J.M., Matthews, P.M.: Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. *Nat Neurosci* 6(7), 750–7 (2003)
- [8] Bergstra, J., Breuleux, O., Bastien, F., Lamblin, P., Pascanu, R., Desjardins, G., Turian, J., Warde-Farley, D., Bengio, Y.: Theano: a cpu and gpu math expression compiler. *Proceedings of the Python for scientific computing conference (SciPy)* 4, 3 (2010)
- [9] Brodmann, K.: Vergleichende Lokalisationslehre der Grosshirnrinde (1909)
- [10] Buckner, R.L., Krienen, F.M., Yeo, B.T.: Opportunities and limitations of intrinsic functional connectivity mri. *Nature neuroscience* 16(7), 832–837 (2013)
- [11] Bühlmann, P., Van De Geer, S.: Statistics for high-dimensional data: methods, theory and applications. Springer Science & Business Media (2011)
- [12] Bullmore, E., Sporns, O.: Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience* 10(3), 186–198 (2009)
- [13] Buzsáki, G., Draguhn, A.: Neuronal oscillations in cortical networks. *science* 304(5679), 1926–1929 (2004)
- [14] Bzdok, D., Langner, R., Schilbach, L., Jakobs, O., Roski, C., Caspers, S., Laird, A.R., Fox, P.T., Zilles, K., Eickhoff, S.B.: Characterization of the temporo-parietal junction by combining data-driven parcellation, complementary connectivity analyses, and functional decoding. *Neuroimage* 81, 381392 (2013)
- [15] Corbetta, M., Patel, G., Shulman, G.L.: The reorienting system of the human brain: from environment to theory of mind. *Neuron* 58(3), 306–24 (2008)
- [16] Craddock, R.C., James, G.A., Holtzheimer, P. E., r., Hu, X.P., Mayberg, H.S.: A whole brain fMRI atlas generated via spatially constrained spectral clustering. *Hum Brain Mapp* 33(8), 1914–28 (2012)
- [17] Daudet, L.: Sparse and structured decompositions of audio signals in overcomplete spaces. In: *In International Conference on Digital Audio Effects* (2004)
- [18] Eickhoff, S.B., Thirion, B., Varoquaux, G., Bzdok, D.: Connectivity-based parcellation: Critique and implications. *Hum Brain Mapp* (2015)
- [19] Fransson, P.: How default is the default mode of brain function? further evidence from intrinsic bold signal fluctuations. *Neuropsychologia* 44, 28362845 (2006)
- [20] Friston, K.J.: Imaging cognitive anatomy. *Trends in cognitive sciences* 1(1), 21–27 (1997)
- [21] Friston, K.: Beyond phrenology: what can neuroimaging tell us about distributed circuitry? *Annual review of neuroscience* 25(1), 221–250 (2002)
- [22] Gabrieli, J.D., Ghosh, S.S., Whitfield-Gabrieli, S.: Prediction as a humanitarian and pragmatic contribution from human cognitive neuroscience. *Neuron* 85(1), 11–26 (2015)
- [23] Giraud, C.: *Introduction to High-Dimensional Statistics*. CRC Press (2014)
- [24] Harzallah, H., Jurie, F., Schmid, C.: Combining efficient object localization and image classification. In: *Computer Vision, 2009 IEEE 12th International Conference on*. pp. 237–244. IEEE (2009)
- [25] Hastie, T., Tibshirani, R., Wainwright, M.: *Statistical Learning with Sparsity: The Lasso and Generalizations*. CRC Press (2015)

- [26] Huang, J., Zhang, T., Metaxas, D.: Learning with structured sparsity. *J Mach Learn Res* 12, 3371–3412 (2011)
- [27] Hubel, D.H., Wiesel, T.N.: Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology* 160(1), 106 (1962)
- [28] Iaria, G., Fox, C.J., Waite, C.T., Aharon, I., Barton, J.J.: The contribution of the fusiform gyrus and superior temporal sulcus in processing facial attractiveness: neuropsychological and neuroimaging evidence. *Neuroscience* 155(2), 409–22 (2008)
- [29] Jbabdi, S., Behrens, T.E.: Long-range connectomics. *Annals of the New York Academy of Sciences* 1305(1), 83–93 (2013)
- [30] Jenatton, R., Audibert, J.Y., Bach, F.: Structured variable selection with sparsity-inducing norms. *The Journal of Machine Learning Research* 12, 2777–2824 (2011)
- [31] Jenatton, R., Gramfort, A., Michel, V., Obozinski, G., Bach, F., Thirion, B.: Multi-scale mining of fmri data with hierarchical structured sparsity. In: *Pattern Recognition in NeuroImaging (PRNI), 2011 International Workshop on*. pp. 69–72. IEEE (2011)
- [32] Jenatton, R., Obozinski, G., Bach, F.: Structured sparse principal component analysis. arXiv preprint arXiv:0909.1440 (2009)
- [33] Jordan, M.I.: *Frontiers in massive data analysis*. National Academies Report (2015)
- [34] Kang, J.W.: Structured sparse representation of residue in screen content video coding. *Electronics Letters* 51(23), 1871–1873 (2015)
- [35] Kanwisher, N.: Functional specificity in the human brain: a window into the functional architecture of the mind. *Proceedings of the National Academy of Sciences* 107(25), 11163–11170 (2010)
- [36] Kim, S., Xing, E.P., et al.: Tree-guided group lasso for multi-response regression with structured sparsity, with an application to eqtl mapping. *The Annals of Applied Statistics* 6(3), 1095–1117 (2012)
- [37] Kim, T., Shakhnarovich, G., Urtasun, R.: Sparse coding for learning interpretable spatio-temporal primitives. In: *Advances in neural information processing systems*. pp. 1117–1125 (2010)
- [38] Mesulam, M.M.: From sensation to cognition. *Brain* 121, 1013–52 (1998)
- [39] Morales, J., Micchelli, C.A., Pontil, M.: A family of penalty functions for structured sparsity. In: *Advances in Neural Information Processing Systems*. pp. 1612–1623 (2010)
- [40] Passingham, R.E., Stephan, K.E., Kotter, R.: The anatomical basis of functional localization in the cortex. *Nat Rev Neurosci* 3(8), 606–16 (2002)
- [41] Pinel, P., Thirion, B., Meriaux, S., Jobert, A., Serres, J., Le Bihan, D., Poline, J.B., Dehaene, S.: Fast reproducible identification and large-scale databasing of individual functional cognitive networks. *BMC Neurosci* 8, 91 (2007)
- [42] Raichle, M.E., MacLeod, A.M., Snyder, A.Z., Powers, W.J., Gusnard, D.A., Shulman, G.L.: A default mode of brain function. *Proc Natl Acad Sci U S A* 98(2), 676–82 (2001)
- [43] Rapaport, F., Barillot, E., Vert, J.P.: Classification of arraycgh data using fused svm. *Bioinformatics* 24(13), i375–i382 (2008)
- [44] Saygin, Z.M., Osher, D.E., Koldewyn, K., Reynolds, G., Gabrieli, J.D., Saxe, R.R.: Anatomical connectivity patterns predict face selectivity in the fusiform gyrus. *Nat Neurosci* 15(2), 321–7 (2012)
- [45] Scannell, J.W., Blakemore, C., Young, M.P.: Analysis of connectivity in the cat cerebral cortex. *J Neurosci* 15(2), 1463–83 (1995)
- [46] Seeley, W.W., Menon, V., Schatzberg, A.F., Keller, J., Glover, G.H., Kenna, H., Reiss, A.L., Greicius, M.D.: Dissociable intrinsic connectivity networks for salience processing and executive control. *J Neurosci* 27(9), 2349–2356 (2007)
- [47] Smith, S.M., Beckmann, C.F., Andersson, J., Auerbach, E.J., Bijsterbosch, J., Douaud, G., Duff, E., Feinberg, D.A., Griffanti, L., Harms, M.P., et al.: Resting-state fmri in the human connectome project. *Neuroimage* 80, 144–168 (2013)
- [48] Smith, S.M., Fox, P.T., Miller, K.L., Glahn, D.C., Fox, P.M., Mackay, C.E., Filippini, N., Watkins, K.E., Toro, R., Laird, A.R., Beckmann, C.F.: Correspondence of the brain's functional architecture during activation and rest. *Proc Natl Acad Sci U S A* 106(31), 13040–5 (2009)
- [49] Sporns, O.: Contributions and challenges for network models in cognitive neuroscience. *Nat Neurosci* 17(5), 652–60 (2014)
- [50] Sridharan, D., Levitin, D.J., Menon, V.: A critical role for the right fronto-insular cortex in switching between central-executive and default-mode networks. *Proc Natl Acad Sci U S A* 105(34), 12569–74 (2008)

- [51] Stephan, K.E., Friston, K.J., Frith, C.D.: Dysconnection in schizophrenia: from abnormal synaptic plasticity to failures of self-monitoring. *Schizophr Bull* 35(3), 509–27 (2009)
- [52] Tononi, G., Edelman, G.M., Sporns, O.: Complexity and coherency: integrating information in the brain. *Trends in cognitive sciences* 2(12), 474–484 (1998)
- [53] Vinci, G., Freeman, P., Newman, J., Wasserman, L., Genovese, C.: Estimating the distribution of galaxy morphologies on a continuous space. arXiv preprint arXiv:1406.7536 (2014)
- [54] Vogt, C., Vogt, O.: Allgemeine Ergebnisse unserer Hirnforschung, vol. 21. JA Barth (1919)
- [55] Yeo, B.T., Krienen, F.M., Sepulcre, J., Sabuncu, M.R., Lashkari, D., Hollinshead, M., Roffman, J.L., Smoller, J.W., Zollei, L., Polimeni, J.R., Fischl, B., Liu, H., Buckner, R.L.: The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J Neurophysiol* 106(3), 1125–65 (2011)
- [56] Young, M.P.: The organization of neural systems in the primate cerebral cortex. *Proc Biol Sci* 252(1333), 13–8 (1993)
- [57] Yuste, R.: From the neuron doctrine to neural networks. *Nat Rev Neurosci* 16(8), 487–497 (2015)
- [58] Zeki, S.M.: Functional specialisation in the visual cortex of the rhesus monkey. *Nature* 274(5670), 423–428 (1978)
- [59] Zhao, P., Rocha, G., Yu, B.: The composite absolute penalties family for grouped and hierarchical variable selection. *The Annals of Statistics* pp. 3468–3497 (2009)
- [60] Zilles, K., Amunts, K.: Receptor mapping: architecture of the human cerebral cortex. *Current opinion in neurology* 22(4), 331–339 (2009)