Good-Enough Brain Model: Challenges, Algorithms and Discoveries in Multi-Subject Experiments

Evangelos E. Papalexakis
Carnegie Mellon University
epapalex@cs.cmu.edu

Alona Fyshe
Carnegie Mellon University
afyshe@cs.cmu.edu

Nicholas D. Sidiropoulos
University of Minnesota
nikos@ece.umn.edu

Partha Pratim Talukdar
Carnegie Mellon University
partha.talukdar@cs.cmu.edu

Tom M. Mitchell
Carnegie Mellon University
tom.mitchell@cmu.edu

Christos Faloutsos
Carnegie Mellon University
christos@cs.cmu.edu

ABSTRACT

Given a simple noun such as *apple*, and a question such as *is it edible?*, what processes take place in the human brain? More specifically, given the stimulus, what are the interactions between (groups of) neurons (also known as *functional connectivity*) and how can we automatically infer those interactions, given measurements of the brain activity? Furthermore, how does this connectivity differ across different human subjects?

In this work we present a simple, novel *good-enough* brain model, or GeBM in short, and a novel algorithm Sparse-SysId, which are able to effectively model the dynamics of the neuron interactions and infer the functional connectivity. Moreover, GeBM is able to simulate basic psychological phenomena such as *habituation* and *priming* (whose definition we provide in the main text).

We evaluate GeBM by using both synthetic and real brain data. Using the real data, GeBM produces brain activity patterns that are strikingly similar to the real ones, and the inferred functional connectivity is able to provide neuroscientific insights towards a better understanding of the way that neurons interact with each other, as well as detect regularities and outliers in multi-subject brain activity measurements.

Categories and Subject Descriptors

H.2.8 [**Database Management**]: Database Applications—*Data mining*; G.3 [**Mathematics of Computing**]: Probability and Statistics—*Time series analysis*; J.3 [**Computer Applications**]: Life and Medical Sciences—*Biology and genetics*; J.4 [**Computer Applications**]: Social and Behavioral Sciences—*Psychology*

Keywords

Brain Activity Analysis; System Identification; Brain Functional Connectivity; Control Theory; Neuroscience

KDD'14, August 24–27, 2014, New York, NY, USA.
Copyright 2014 ACM 978-1-4503-2956-9/14/08 ...$15.00.
http://dx.doi.org/10.1145/2623330.2623639.

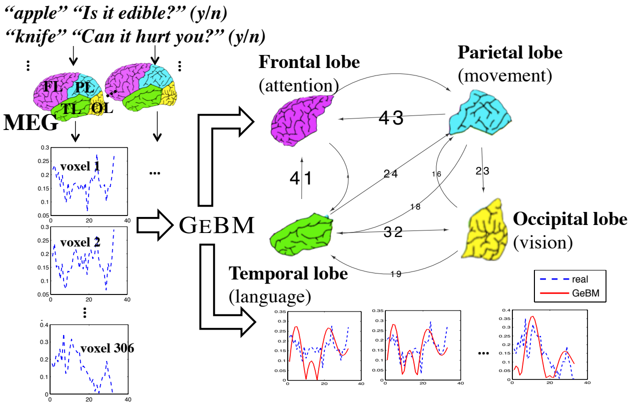


Figure 1: Big picture: our GeBM estimates the hidden functional connectivity (top right, weighted arrows indicating number of inferred connections), when given multiple human subjects (left) that respond to yes/no questions (e.g., *edible?*) for typed words (e.g., *apple*). Bottom right: GeBM also produces brain activity (in solid-red), that matches reality (in dashed-blue).

1. INTRODUCTION

Can we infer the brain activity for subject 'Alice', when she is shown the typed noun *apple* and has to answer a yes/no question, like *is it edible*? Can we infer the connectivity of brain regions, given numerous brain activity data of subjects in such experiments? These are the first two goals of this work: single-subject, and multi-subject analysis of brain activity.

The third and final goal is to develop a brain connectivity model, that can also generate activity that agrees with psychological phenomena, like *priming*[1] and *habituation*[2]

Here we tackle all these challenges. We are given *Magnetoencephalography* (MEG) brain scans for nine subjects, shown several typed nouns (*apple*, *hammer*, etc), and being requested to answer a yes/no question (*is it edible?*, *is it dangerous?*, and so on), by pressing one of two buttons.

Our approach: Discovering the multi-billion connections among

[1] *Priming* illustrates the power of context: a person hearing the word *iPod*, and then *apple*, will think of *Apple-inc*, as opposed to the fruit *apple*

[2] *Habituation* illustrates compensation: a person hearing the same word all the time, will eventually stop paying attention.

the tens of billions [23, 2] of neurons would be the holy grail, and clearly outside the current technological capabilities. How close can we approach this ideal? We propose to use a *good-enough* approach, and try to explain as much as we can, by assuming a small, manageable count of neuron-regions and their interconnections, and trying to guess the connectivity from the available MEG data. In more detail, we propose to formulate the problem as 'system identification' from control theory, and we develop novel algorithms to find *sparse* solutions.

We show that our *good-enough* approach is a very good first step, leading to a tractable, yet effective model (GEBM), that can answer the above questions. Figure 1 gives the high-level overview: at the bottom-right, the blue, dashed-line time sequences correspond to measured brain activity; the red lines correspond to the guess of our GEBM model. Notice the qualitative goodness of fit. At the top-right, the arrows indicate interaction between brain regions that our analysis learned, with the weight being the strength of interaction. Thus we see that the vision cortex ('occipital lobe') is well connected to the language-processing part ('temporal lobe'), which agrees with neuroscience, since all our experiments involved typed words.

Our **contributions** are as follows:

- *Novel analytical model*: We propose the GEBM model (see Section 3, and Eq (2)-(3)).
- *Algorithm*: we introduce SPARSE-SYSID, a novel, sparse, system-identification algorithm (see Section 3).
- *Effectiveness*: Our model can explain psychological phenomena, such as habituation and priming (see Section 5.4); it also gives results that agree with experts' intuition (see Section 5.1)
- *Validation*: GEBM indeed matches the given activity patterns, both on synthetic, as well as real data (see Section 4 and 5.3, resp.).
- *Multi-subject analysis*: Our SPARSE-SYSID, applied on 9 human subjects (Section 5.2), showed that (a) 8 of them had very consistent brain-connectivity patterns while (b) the outlier was due to exogenous factors (excessive road-traffic noise during his experiment).

Additionally, our GEBM highlights connections between multiple, mostly disparate areas: 1) Neuroscience, 2) Control Theory & System Identification, and 3) Psychology.

Reproducibility: Our implementation is open sourced and publicly available [3]. Due to privacy reasons, we are not able to release the MEG data, however, in the online version of the code we include the synthetic benchmarks, as well as the simulation of psychological phenomena using GEBM.

2. PROBLEM DEFINITION

As mentioned earlier, our goal is to infer the brain connectivity, given measurements of brain activity on multiple *yes/no tasks*, of multiple subjects. We define as **yes/no task** the experiment where the subject is given a yes/no question (like, *'is it edible?'*, *'is it alive?'*), and a typed English word (like, *apple*, *chair*), and has to decide the answer.

Throughout the entire process, we attach m sensors that record brain activity of a human subject. Here we are using Magnetoencephalography (MEG) data, although our GEBM model could be applied to any type of measurement (fMRI, etc). In Section 5.4 we provide a more formal definition of the measurement technique.

Thus, in a given experiment, at every time-tick t we have m measurements, which we arrange in an $m \times 1$ vector $\mathbf{y}(t)$. Additionally, we represent the stimulus (e.g. *apple*) and the task (e.g. *is it edible?*) in a time-dependent vector $\mathbf{s}(t)$, by using feature representation of the stimuli; a detailed description of how the stimulus vector is formed can be found in Section 5.4. For the rest of the paper, we shall use interchangeably the terms *sensor*, *voxel* and neuron-region.

We are interested in two problems: the first is to understand how the brain works, given a single subject. The second problem is to do cross-subject analysis, to find commonalities (and deviations) in a group of several human subjects. Informally, we have:

INFORMAL PROBLEM 1 (SINGLE SUBJECT). *Definition:*

- **Given***: The input stimulus; and a sequence of $m \times T$ brain activity measurements for the m voxels, for all timeticks $t = 1 \cdots T$*
- **Estimate***: the functional connectivity of the brain, i.e. the strength and direction of interaction, between pairs of the m voxels, such that*
    1. *we understand how the brain-regions collaborate, and*
    2. *we can effectively simulate brain activity.*

For the second problem, informally we have:

INFORMAL PROBLEM 2 (MULTI-SUBJECT). *Definition:*

- **Given***: Multi-subject experimental data (brain activity for 'yes/no tasks')*
- **Detect***: Regularities, commonalities, clusters of subjects (if any), outlier subjects (if any).*

For the particular experimental setting, prior work [15] has only considered transformations from the space of noun features to the voxel space and vice versa, as well as word-concept specific prediction based on estimating the covariance between the voxels [7].

Next we formalize the problems, we show some straightforward (but unsuccessful) solutions, and finally we give the proposed model GEBM, and the estimation algorithm.

3. PROBLEM FORMULATION AND PROPOSED METHOD

There are two over-arching assumptions:

- linearity: linear models are good-enough
- stationarity: the connectivity of the brain does not change, at least for the time-scales of our experiments.

Non-linear/sigmoid models is a natural direction for future work; and so is the study of neuroplasticity, where the connectivity changes. However, as we show later, linear, static, models are "*good-enough*" to answer the problems we listed, and thus we stay with them.

However, we have to be careful. Next we list some natural, but unsuccessful models, to illustrate that we did do 'due dilligence', and to highlight the need for our slightly more complicated, GEBM model. The conclusion is that the hasty, Model$_0$, below, leads to poor behavior, as we show in Figure 2 (red, and black, lines), completely missing all the trends and oscillations of the real signal (in dotted-blue line). In fact, the next subsection may be skipped, at a first reading.

3.1 First (unsuccessful) approach: Model$_0$

Given the linearity and static-connectivity assumptions above, a natural additional assumption is to postulate that the $m \times 1$ brain activity vector $\mathbf{y}(t+1)$ depends *linearly*, on the activities of the previous time-tick $\mathbf{y}(t)$, and, of course, the input stimulus, that is, the $s \times 1$ vector $\mathbf{s}(t)$.

[3] http://www.cs.cmu.edu/~epapalex/src/GeBM.zip

Formally, in the absence of input stimulus, we expect that

$$\mathbf{y}(t+1) = \mathbf{A}\mathbf{y}(t).$$

where $\mathbf{A}$ is the $m \times m$ connectivity matrix of the m brain regions. Including the (linear) influence of the input stimulus $\mathbf{s}(t)$, we reach the MODEL$_0$:

$$\mathbf{y}(t+1) \quad = \quad \mathbf{A}_{[m\times m]} \times \mathbf{y}(t) \; + \; \mathbf{B}_{[m\times s]} \times \mathbf{s}(t) \qquad (1)$$

The $\mathbf{B}_{[m\times s]}$ matrix shows how the s input signals affect the m brain-regions.

To solve for $\mathbf{A}, \mathbf{B}$, notice that: $\mathbf{y}(t+1) = \begin{bmatrix}\mathbf{A} & \mathbf{B}\end{bmatrix} \begin{bmatrix}\mathbf{y}(t) \\ \mathbf{s}(t)\end{bmatrix}$ which eventually becomes $\mathbf{Y}' = \begin{bmatrix}\mathbf{A} & \mathbf{B}\end{bmatrix} \begin{bmatrix}\mathbf{Y} \\ \mathbf{S}\end{bmatrix}$

In the above equation, we arranged all the measurement vectors $\mathbf{y}(t)$ in matrices: $\mathbf{Y} = \begin{bmatrix}\mathbf{y}(1) & \cdots & \mathbf{y}(T-1)\end{bmatrix}$,
$\mathbf{Y}' = \begin{bmatrix}\mathbf{y}(2) & \cdots & \mathbf{y}(T)\end{bmatrix}$, and $\mathbf{S} = \begin{bmatrix}\mathbf{s}(1) & \cdots & \mathbf{s}(T-1)\end{bmatrix}$

This is a well-known, least squares problem. We can solve it 'as is'; we can ask for a low-rank solution; or for a sparse solution - none yields a good result, but we briefly describe each, next.

- **Least Squares (LS)**: The solution is unique, using the Moore-Penrose pseudo-inverse, i.e. $\begin{bmatrix}\mathbf{A} & \mathbf{B}\end{bmatrix}_{LS} = \mathbf{Y}' \times \begin{bmatrix}\mathbf{Y} \\ \mathbf{S}\end{bmatrix}^{\dagger}$.
- **Canonical Correlation Analysis (CCA)**: The reader may be wondering: *what if we have over-fitting here - why not ask for a low-rank solution*. This is exactly what CCA does [14]. It solves for the same objective function as in LS, further requesting low rank r for $\begin{bmatrix}\mathbf{A} & \mathbf{B}\end{bmatrix}$
- Sparse solution: what if we solve the least squares problem, further requesting a sparse solution? We tried that, too, with ℓ_1 norm regularization.

None of the above worked. Fig. 2 shows the real brain activity (dotted-blue line) and predicted activity, using LS (pink) and CCA (black), for a particular voxel. The solutions completely fail to match the trends and oscillations. The results for the ℓ_1 regularization, and for several other voxels, are similar to the one shown, and omitted for brevity.

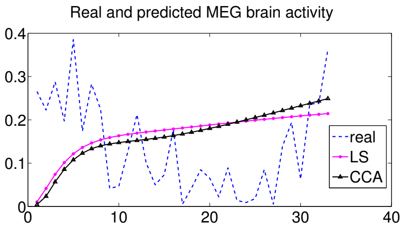


Figure 2: MODEL$_0$ fails: True brain activity (dotted blue) and the model estimate (pink, and black, resp., for the least squares, and for the CCA variation).

The conclusion of this subsection is that we need a more complicated model, which leads us to GEBM, next.

3.2 Proposed approach: GeBM

Before we introduce our proposed model, we should introduce our notation, which is succinctly shown in Table 1.

Formulating the problem as MODEL$_0$ is not able to meet the requirements for our desired solution. However, we have not exhausted the space of possible formulations that live within our set of simplifying assumptions. In this section, we describe GEBM, our proposed approach which, under the assumptions that we have already made in Section 2, is able to meet our requirements remarkably well.

Symbol	Definition
n	number of hidden neuron-regions
m	number of MEG sensors/voxels we observe (306)
s	number of input signals (40 questions)
T	time-ticks of each experiment (340 ticks, of 5msec each)
$\mathbf{x}(t)$	vector of neuron activities at time t
$\mathbf{y}(t)$	vector of voxel activities at time t
$\mathbf{s}(t)$	vector of input-sensor activities at time t
$\mathbf{A}_{[n\times n]}$	connectivity matrix between neurons (or neuron regions)
$\mathbf{C}_{[m\times n]}$	summarization matrix (neurons to voxels)
$\mathbf{B}_{[n\times s]}$	perception matrix (sensors to neurons)
$\mathbf{A}_v$	connectivity matrix between voxels
REAL	real part of a complex number
IMAG	imaginary part of a complex number
$\mathbf{A}^{\dagger}$	Moore-Penrose Pseudoinverse of $\mathbf{A}$

Table 1: Table of symbols

In order to come up with a more accurate model, it is useful to look more carefully at the actual system that we are attempting to model. In particular, the brain activity vector $\mathbf{y}$ that we observe is simply the collection of values recorded by the m sensors, placed on a person's scalp. In MODEL$_0$, we attempt to model the dynamics of the sensor measurements directly. However, by doing so, we are directing our attention to an *observable proxy* of the process that we are trying to estimate (i.e. the functional connectivity). Instead, it is more beneficial to model the direct outcome of that process. Ideally, we would like to capture the dynamics of the internal state of the person's brain, which, in turn, cause the effect that we are measuring with our MEG sensors.

Let us assume that there are n *hidden* (hyper-)regions of the brain, which interact with each other, causing the activity that we observe in $\mathbf{y}$. We denote the vector of the hidden brain activity as $\mathbf{x}$ of size $n \times 1$. Then, by using the same idea as in MODEL$_0$, we may formulate the temporal evolution of the hidden brain activity as:

$$\mathbf{x}(t+1) = \mathbf{A}_{[n\times n]} \times \mathbf{x}(t) \; + \; \mathbf{B}_{[n\times s]} \times \mathbf{s}(t)$$

A subtle issue that we have yet to address is the fact that $\mathbf{x}$ is *not observed* and we have no means of measuring it. We propose to resolve this issue by modelling the measurement procedure itself, i.e. model the transformation of a hidden brain activity vector to its observed counterpart. We assume that this transformation is linear, thus we are able to write

$$\mathbf{y}(t) = \mathbf{C}_{[m\times n]}\mathbf{x}(t)$$

Putting everything together, we end up with the following set of equations, which constitute our proposed model GEBM:

$$\boxed{\begin{aligned} \mathbf{x}(t+1) &= \mathbf{A}_{[n\times n]} \times \mathbf{x}(t) \; + \; \mathbf{B}_{[n\times s]} \times \mathbf{s}(t) && (2) \\ \mathbf{y}(t) &= \mathbf{C}_{[m\times n]} \times \mathbf{x}(t) && (3) \end{aligned}}$$

The key concepts behind GEBM are:

- **(Latent) Connectivity Matrix**: We assume that there are n regions, each containing 1 or more neurons, and they are connected with an $n \times n$ adjacency matrix $\mathbf{A}_{[n\times n]}$. We only observe m voxels, each containing multiple regions, and we record the activity (eg., magnetic activity) in each of them; this is the total activity in the constituent regions

- **Measurement Matrix**: Matrix $\mathbf{C}_{[m\times n]}$ is an $m \times n$ matrix, with $c_{i,j}$ =1 if voxel i contains region j
- **Perception Matrix**: Matrix $\mathbf{B}_{[n\times s]}$ shows the influence of each sensor to each neuron-region. The input is denoted as $\mathbf{s}$, with s input signals
- **Sparsity**: We require that our model's matrices are *sparse*; only few sensors are responsible for a specific brain region. Additionally, the interactions between regions should not form a complete graph, and finally, the perception matrix should map only few activated sensors to neuron regions at every given time.

3.3 Algorithm

Our solution is inspired by control theory, and more specifically by a sub-field of control theory, called *system identification*. In the appendix, we provide an overview of how this can be accomplished. However, the matrices we obtain through this process are usually dense, counter to GeBM's specifications. We, thus, need to refine the solution until we obtain the desired level of sparsity. In the next few lines, we show why this sparsification has to be done carefully, and we present our approach.

Crucial to GeBM's behavior is the spectrum of its matrices; in other words, any operation that we apply on any of GeBM's matrices needs to preserve the eigevnalue profile (for matrix $\mathbf{A}$) or the singular values (for matrices $\mathbf{B}, \mathbf{C}$). Alterations thereof may lead GeBM to instabilities. From a control theoretic and stability perspective, we are mostly interested in the eigenvalues of $\mathbf{A}$, since they drive the behavior of the system. Thus, in our experiments, we heavily rely on assessing how well we estimate these eigenvalues.

Sparsifying a matrix while preserving its spectrum can be seen as a similarity transformation of the matrix to a sparse subspace. The following lemma sheds more light towards this direction.

Lemma 1. *System identification is able to recover matrices* $\mathbf{A}$*,* $\mathbf{B}$*,* $\mathbf{C}$ *of* GeBM *up to rotational/similarity transformations.*

Proof. See the appendix. □

An important corollary of the above lemma (also proved in the appendix) is the fact that pursuing sparsity only on, say, matrix $\mathbf{A}$ is not well defined. Therefore, since all three matrices share the same similarity transformation freedom, we have to sparsify all three.

In Sparse-SysId, we propose a fast, greedy sparsification scheme which can be seen as approximately applying the aforementioned similarity transformation to $\mathbf{A}, \mathbf{B}, \mathbf{C}$, without calculating or applying the transformation itself. Iteratively, for all three matrices, we delete small values, while maintaining ther spectrum within ϵ from the one obtained through system identification. Additionally, for $\mathbf{A}$, we also do not allow eigenvalues to switch from complex to real and vice versa. This scheme works very well in practice, providing very sparse matrices, while respecting their spectrum. In Algorithm 1, we provide an outline of the algorithm.

So far, GeBM as we have described it, is able to give us the hidden functional connectivity and the measurement matrix, but does not directly offer the voxel-to-voxel connectivity, unlike Model$_0$, which models it explicitly. However, this is by no means a weakness of GeBM, since there is a simple way to obtain the voxel-to-voxel connectivity (henceforth referred to as $\mathbf{A}_v$) from GeBM's matrices.

Lemma 2. *Assuming that* $m > n$*, the voxel-to-voxel functional connectivity matrix* $\mathbf{A}_v$ *can be defined and is equal to* $\mathbf{A}_v = \mathbf{CAC}^\dagger$

Algorithm 1: Sparse-SysId: Sparse System Identification of GeBM

```
Input: Training data in the form $\{\mathbf{y}(t), \mathbf{s}(t)\}_{t=1}^T$, number of
    hidden states $n$.
Output: GeBM matrices $\mathbf{A}$ (hidden connectivity matrix), $\mathbf{B}$
    (perception matrix), $\mathbf{C}$ (measurement matrix), and $\mathbf{A}_v$
    (voxel-to-voxel matrix).
$\{\mathbf{A}^{(0)}, \mathbf{B}^{(0)}, \mathbf{C}^{(0)}\} = \text{SysId}\left(\{\mathbf{y}(t), \mathbf{s}(t)\}_{t=1}^T, n\right)$
$\mathbf{A} = \text{EigenSparsify}(\mathbf{A}^{(0)})$
$\mathbf{B} = \text{SingularSparsify}(\mathbf{B}^{(0)})$
$\mathbf{C} = \text{SingularSparsify}(\mathbf{C}^{(0)})$
$\mathbf{A}_v = \mathbf{CAC}^\dagger$
```

Algorithm 2: EigenSparsify: Eigenvalue Preserving Sparsification of System Matrix $\mathbf{A}$.

```
Input: Square matrix $\mathbf{A}^{(0)}$.
Output: Sparsified matrix $\mathbf{A}$.
$\boldsymbol{\lambda}^{(0)}$ =Eigenvalues($\mathbf{A}^{(0)}$)
Initialize $\mathbf{d}_R^{(0)} = \mathbf{0}$, $\mathbf{d}_I^{(0)} = \mathbf{0}$. Vector $\mathbf{d}_R^{(i)}$ holds the
    element-wise difference of the real part of the eigenvalues of
    $\mathbf{A}^{(i)}$. Similarly for $\mathbf{d}_I^{(i)}$ and the imaginary part.
Set vector $\mathbf{c}$ as a boolean vector that indicates whether the $j$-th
    eigenvalue in $\boldsymbol{\lambda}^{(0)}$ is complex or not. One way to do it is to
    evaluate element-wise the following boolean expression:
    $\mathbf{c} = \left(\text{Imag}(\boldsymbol{\lambda}^{(0)}) \neq 0\right)$.
Initialize $i = 0$
while $\mathbf{d}_R^{(i)} \leq \epsilon$ and $\mathbf{d}_I^{(i)} \leq \epsilon$ and $\left(\text{Imag}(\boldsymbol{\lambda}^{(i)}) \neq 0\right) == \mathbf{c}$
do
    Initialize $\mathbf{A}^{(i)} = \mathbf{A}^{(i-1)}$
    $\{v_i^*, v_j^*\} = \arg\min_{v_i,v_j} |\mathbf{A}^{(i-1)}(v_i, v_j)|$
        s.t. $\mathbf{A}^{(i-1)}(v_i, v_j) \neq 0$.
    Set $\mathbf{A}^{(i)}(v_i^*, v_j^*) = 0$
    $\boldsymbol{\lambda}^{(i)}$ =Eigenvalues($\mathbf{A}^{(i)}$)
    $\mathbf{d}_R^{(i)} = |\text{Real}(\boldsymbol{\lambda}^{(i)}) - \text{Real}(\boldsymbol{\lambda}^{(i-1)})|$
    $\mathbf{d}_I^{(i)} = |\text{Imag}(\boldsymbol{\lambda}^{(i)}) - \text{Imag}(\boldsymbol{\lambda}^{(i-1)})|$
end while
$\mathbf{A} = \mathbf{A}^{(i-1)}$
```

Algorithm 3: SingularSparsify: Singular Value Preserving Sparsification

```
Input: Matrix $\mathbf{M}^{(0)}$.
Output: Sparsified matrix $\mathbf{M}$
$\boldsymbol{\lambda}^{(0)}$ =SingularValues($\mathbf{A}^{(0)}$)
Initialize $\mathbf{d}_R^{(0)} = \mathbf{0}$ which holds the element-wise difference
    of the singular values of $\mathbf{A}^{(i)}$.
Initialize $i = 0$
while $\mathbf{d}_R^{(i)} \leq \epsilon$ do
    Initialize $\mathbf{M}^{(i)} = \mathbf{M}^{(i-1)}$
    $\{v_i^*, v_j^*\} = \arg\min_{v_i,v_j} |\mathbf{M}^{(i-1)}(v_i, v_j)|$
        s.t. $\mathbf{M}^{(i-1)}(v_i, v_j) \neq 0$.
    Set $\mathbf{M}^{(i)}(v_i^*, v_j^*) = 0$
    $\boldsymbol{\lambda}^{(i)}$ =SingularValues($\mathbf{M}^{(i)}$)
    $\mathbf{d}_R^{(i)} = |\boldsymbol{\lambda}^{(i)} - \boldsymbol{\lambda}^{(i-1)}|$
end while
$\mathbf{M} = \mathbf{M}^{(i-1)}$
```

Proof. The observed voxel vector can be written as

$$\mathbf{y}(t+1) = \mathbf{C}\mathbf{x}(t+1) = \mathbf{C}\mathbf{A}\mathbf{x}(t) + \mathbf{C}\mathbf{B}\mathbf{s}(t)$$

Matrix $\mathbf{C}$ is tall (i.e. $m > n$), thus we can write: $\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) \Leftrightarrow \mathbf{x}(t) = \mathbf{C}^{\dagger}\mathbf{y}(t)$ Consequently, $\mathbf{y}(t+1) = \mathbf{C}\mathbf{A}\mathbf{C}^{\dagger}\mathbf{y}(t) + \mathbf{C}\mathbf{B}\mathbf{s}(t)$ Therefore, it follows that $\mathbf{C}\mathbf{A}\mathbf{C}^{\dagger}$ is the voxel-to-voxel matrix $\mathbf{A}_v$. □

Finally, an interesting aspect of our proposed model GeBM is the fact that if we ignore the notion of the summarization, i.e. matrix $\mathbf{C} = \mathbf{I}$, then our model is reduced to the simple model Model$_0$. In other words, GeBM contains Model$_0$ as a special case. This observation demonstrates the importance of hidden states in GeBM.

4. EVALUATION

4.1 Implementation Details

The code for Sparse-SysId has been written in Matlab. For the system identification part, initially we experimented with Matlab's System Identification Toolbox and the algorithms in [10]. These algorithms worked well for smaller to medium scales, but were unable to perform on our full dataset. Thus, in our final implementation, we use the algorithms of [20]. Our code is publicly available at `http://www.cs.cmu.edu/~epapalex/src/GeBM.zip`.

4.2 Evaluation on synthetic data

In lieu of ground truth in our real data, we generated synthetic data to measure the performance of Sparse-SysId.

The way we generate the ground truth system is as follows: First, given fixed n, we generate a matrix $\mathbf{A}$ that has 0.25 on the main diagonal, 0.1 on the first upper diagonal (i.e. the $(i, i+1)$ elements), -0.15 on the first lower diagonal (i.e., the $(i-1, i)$ elements), and 0 everywhere else. We then create randomly generated sparse matrices $\mathbf{B}$ and $\mathbf{C}$, varying s and m respectively.

After we generate a synthetic ground truth model, we generate Gaussian random input data to the system, and we obtain the system's response to that data. Consequently, we use the input/output pairs with Sparse-SysId, and we assess our algorithm's ability to recover the ground truth. Here, we show the *noiseless* case due to space restrictions. In the noisy case, estimation performance is slowly degrading when n increases, however this is expected from estimation theory.

We evaluate Sparse-SysId's accuracy with respect to the following aspects:

Q1: How well can Sparse-SysId recover the true hidden connectivity matrix $\mathbf{A}$?

Q2: How well can Sparse-SysId recover the voxel-to-voxel connectivity matrix $\mathbf{A}_v$?

Q3: Given that we know the the true number of hidden states n, how does Sparse-SysId behave as we vary the n used for GeBM?

In order to answer **Q1**, we measure how well (in terms of RMSE) Sparse-SysId recovers the eigenvalues of $\mathbf{A}$. We are mostly interested in recovering perfectly the real part of the eigenvalues, since even small errors could lead to instabilities. Figure 4(a) shows our results: We observe that the estimation of the real part of the eigenvalues of $\mathbf{A}$ is excellent. We are omitting the estimation results for the imaginary parts, however they are within the ϵ we selected in our sparsification scheme of Sparse-SysId. Overall, Sparse-SysId is able to recover the true GeBM, for various values of m and n.

With respect to **Q2**, it suffices to measure the RMSE of the true $\mathbf{A}_v$ and the estimated one, since we have thoroughly tested the system's behavior in **Q1**. Figure 4(b) shows that the estimation of the voxel-to-voxel connectivity matrix using Sparse-SysId is highly accurate. Additionally, for ease of exposition, in Figure 3 we show an example a true matrix $\mathbf{A}_v$, and its estimation through Sparse-SysId; it is impossible to tell the difference between the two matrices, a fact also corroborated by the RMSE results.

The third dimension of Sparse-SysId's performance is its sensitivity to the selection of the parameter n; In order to test this, we generated a ground truth GeBM with a known n, and we varied our selection of n for Sparse-SysId. The result of the experiment is shown in Fig. 4(c). We observe that for values of n smaller than the real one, Sparse-SysId's performance is increasingly good, and still, for small values of n the estimation quality is good. When n exceeds the value of the real n, the performance starts to degrade, due to *overfitting*. This provides an insight on how to choose n for Sparse-SysId in order to fit GeBM: it is better to under-estimate n rather than over-estimate it, thus, it is better to start with a small n and possibly increase it as soon as performance (e.g. qualitative assessment of how well the estimated model predicts brain activity) starts to degrade.

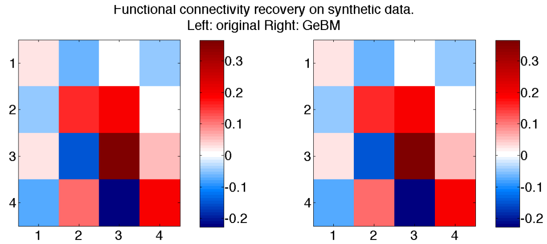


Figure 3: Q2: Perfect estimation of $\mathbf{A}_v$:Comparison of true and estimated $\mathbf{A}_v$, for $n = 3$ and $m = 4$. We can see, qualitatively, that GeBM is able to recover the true voxel-to-voxel functional connectivity.

5. GeBM AT WORK

This section is focused on showing different aspects of GeBM at work. In particular, we present the following discoveries:

D1: We provide insights on the obtained functional connectivity from a Neuroscientific point of view.

D2: Given multiple human subjects, we discover regularities and outliers, with respect to functional connectivity.

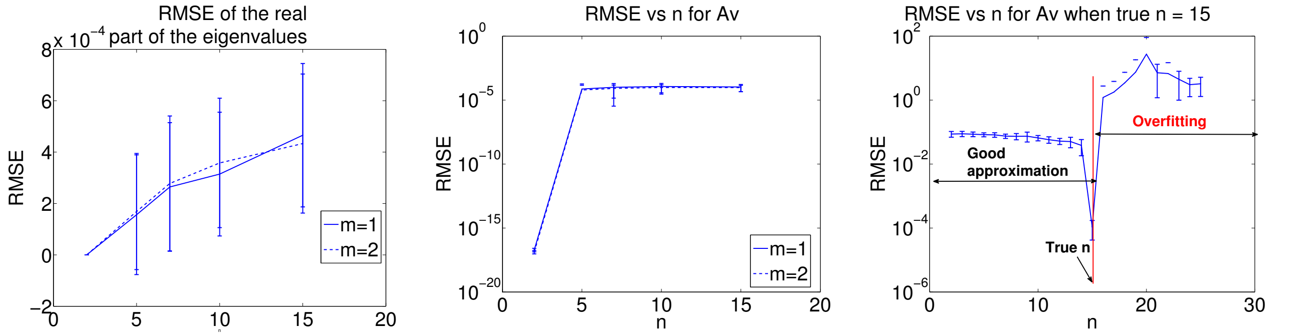
D3: We demonstrate GeBM's ability to simulate brain activity.

D4: We show how GeBM is able to capture two basic psychological phenomena.

Dataset Description & Formulation.

We are using real brain activity data, measured using MEG. MEG (Magnetoencephalography) measures the magnetic field caused by many thousands of neurons firing together, and has good time resolution (1000 Hz) but poor spatial resolution. fMRI (functional Magnetic Resonance Imaging) measures the change in blood oxygenation that results from changes in neural activity, and has good spatial resolution but poor time resolution (0.5-1 Hz). Since we are interested in the temporal dynamics of the brain, we choose to operate on MEG data.

All experiments were conducted at the University of Pittsburgh Medical Center (UPMC) Brain Mapping Center. The MEG machine consists of $m = 306$ sensors, placed uniformly across the subject's scalp. The temporal granularity of the measurements is 5ms, resulting in $T = 340$ time points; after experimenting with different aggregations in the temporal dimension, we decided to use



(a) **Q1**: RMSE of REAL(eigenvalues) of $\mathbf{A}$ vs. n.

(b) **Q2**: Estimation RMSE for $\mathbf{A}_v$ vs. n.

(c) **Q3**: Estimation RMSE for $\mathbf{A}_v$ vs. n, where the true n is fixed and equal to 15.

Figure 4: Sub-figure (a) refers to **Q1**, and show sthat SPARSE-SYSID is able to estimate matrix $\mathbf{A}$ with high accuracy, in control-theoretic terms. Sub-figure (b) illustrates SPARSE-SYSID's ability to accurately estimate the voxel-to-voxel functional connectivity matrix $\mathbf{A}_v$. Finally, sub-figure (c) shows the behavior of SPARSE-SYSID with respect to parameter n, when the true value of this parameter is known; the message from this graph is that as long as n is under-estimated, SPARSE-SYSID's performance is steadily good and is not greatly influenced by the particular choice of n.

50ms of time resolution, because this yielded the most interpretable results.

For the experiments, nine right handed human subjects were shown a set of 60 concrete English nouns (*apple, knife* etc), and for each noun 20 simple yes/no questions (*Is it edible? Can you buy it?* etc). The subject were asked to press the right button if their answer to each question was 'yes', or the left button if the answer was 'no'. After the subject pressed the button, the stimulus (i.e. the noun) would disappear from the screen. We also record the exact time that the subject pressed the button, relative to the appearance of the stimulus on the screen. A more detailed description of the data can be found in [15].

In order to bring the above data to the format that our model expects, we make the following design choices: In lack of sensors that measure the response of the eyes to the shown stimuli, we represent each stimulus by a set of semantic features for that specific noun. This set of features is a superset of the 20 questions that we have already mentioned; the value for each feature comes from the answers given by Amazon Mechanical Turk workers. Thus, from time-tick 1 (when the stimulus starts showing), until the button is pressed, all the features that are active for the particular stimulus are set to 1 on our stimulus vector $\mathbf{s}$, and all the rest features are equal to 0; when the button is pressed, all features are zeroed out. On top of the stimulus features, we also have to incorporate the task information in $\mathbf{s}$, i.e. the particular question shown on the screen. In order to do that, we add 20 more rows to the stimulus vector $\mathbf{s}$, each one corresponding to every question/task. At each given experiment, only one of those rows is set to 1 for all time ticks, and all other rows are set to 0. Thus, the number of input sensors in our formulation is $s = 40$ (i.e. 20 neurons for the noun/stimulus and 20 neurons for the task).

As a last step, we have to incorporate the button pressing information to our model; to that end, we add two more voxels to our observed vector $\mathbf{y}$, corresponding to left and right button pressing; initially, those values are set to 0 and as soon as the button is pressed, they are set to 1.

Finally, we choose $n = 15$ for all the results we show in the following lines; particular choice of n did not incur qualitative changes in the results, however, as we highlight in the previous section, it is better to under-estimate n, and therefore we chose $n = 15$ as an adequately small choice which, at the same time, produces interpretable results.

5.1 D1: Functional Connectivity Graphs

The primary focus of this work is to estimate the functional connectivity of the human brain, i.e. the interaction pattern of groups of neurons. In the next few lines, we present our findings in a concise way and provide Neuroscientific insights regarding the interaction patterns that GEBM was able to infer.

In order to present our findings, we post-process the results obtained through GEBM in the following way: The data we collect come from 306 sensors, placed on the human scalp in a uniform fashion. Each of those 306 sensors is measuring activity from one of the four main regions of the brain, i.e.

- **Frontal Lobe**, associated with attention, short memory, and planning.
- **Parietal Lobe**, associated with movement.
- **Occipital Lobe**, associated with vision.
- **Temporal Lobe**, associated with sensory input processing, language comprehension, and visual memory retention.

Even though our sensors offer within-region resolution, for exposition purposes, we chose to aggregate our findings per region; by doing so, we are still able to provide useful neuroscientific insights.

Figure 5 shows the functional connectivity graph obtained using GEBM. The weights indicate the strength of the interaction, measured by the number of distinct connections we identified. These results are consistent with current research regarding the nature of language processing in the brain. For example, Hickock and Poeppel [9] have proposed a model of language comprehension that includes a "dorsal" and "ventral" pathway. The ventral pathway takes the input stimuli (spoken language in the case of Hickock and Poeppel, images and words in ours) and sends the information to the temporal lobe for semantic processing. Because the occipital cortex is responsible for the low level processing of visual stimuli (including words) it is reasonable to see a strong set of connections between the occipital and temporal lobes. The dorsal pathway sends processed sensory input through the parietal and frontal lobes where they are processed for planning and action purposes. The task performed during the collection of our MEG data required that subjects consider the meaning of the word in the context of a semantic question. This task would require the recruitment of the dorsal pathway (occipital-parietal and parietal-frontal connections). In addition, frontal involvement is indicated when the task performed by the subject requires the selection of semantic information [3], as in our question answering paradigm. It is interesting that the number of connections from parietal to occipital cortex is larger than from occipital to parietal, considering the flow of information is likely occipital to parietal. This could, however, be indicative of what is termed "top down" processing, wherein higher level cognitive processes can work to focus upstream sensory processes. Perhaps the semantic task causes the subjects to focus in anticipation of the upcoming word while keeping the semantic question in mind.

5.2 D2: Cross-subject Analysis

In our experiments, we have 9 participants, all of whom have undergone the same procedure, being presented with the same stimuli, and asked to carry out the same tasks. Availability of such a rich, multi-subject dataset inevitably begs the following question: are there any differences across people's functional connectivity? Or is everyone, more or less, wired equally, at least with respect to the stimuli and tasks at hand?

By using GeBM, we are able (to the extent that our model is able to explain) to answer the above question. We trained GeBM for each of the 9 human subjects, using the entire data from all stimuli and tasks, and obtained matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}$ for each person. For the purposes of answering the above question, it suffices to look at matrix $\mathbf{A}$ (which is the hidden functional connectivity), since it dictates the temporal dynamics of the brain activity. At this point, we have to note that the exact location of each sensor can differ between human subjects, however, we assume that this difference is negligible, given the current voxel granularity dictated by the number of sensors.

In this multi-subject study we have two very important findings:

- **Regularities**: For 8 out of 9 human subjects, we identified almost identical GeBM instances, both with respect to RMSE and to spectrum. In other words, for 8 out of 9 subjects in our study, the inferred functional connectivity behaves almost identically. This fact most likely implies that for the particular set of stimuli and assorted tasks, the human brain behaves similarly across people.
- **Anomaly**: One of our human subjects (#3) deviates from the aforementioned regular behavior.

In Fig. 6(a) & (b) we show the real and imaginary parts of the eigenvalues of $\mathbf{A}$. We can see that for 8 human subjects, the eigenvalues are almost identical. This finding agrees with neuroscientific results on different experimental settings [18], further demonstrating GeBM's ability to provide useful insights on multi-subject experiments. For subject #3 there is a deviation on the real part of the first eigenvalue, as well as a slightly deviating pattern on the imaginary parts of its eigenvalues. Figures 6(c) & (d) compare matrix $\mathbf{A}$ for subjects 1 and 3. Subject 3 negative value on the diagonal (blue square at the $(8, 8)$ entry), a fact unique to this specific person's connectivity.

Moreover, according to the person responsible for the data collection of Subject #3:

> *There was a big demonstration outside the UPMC building during the scan, and I remember the subject complaining during one of the breaks that he could hear the crowd shouting through the walls.*

This is a plausible explanation for the deviation of GeBM for Subject #3.

5.3 D3: Brain Activity Simulation

An additional way to gain confidence on our model is to assess its ability to simulate/predict brain activity, given the inferred functional connectivity. In order to do so, we trained GeBM using data from all but one of the words, and then we simulated brain activity time-series for the left-out word. In lieu of competing methods, we compare our proposed method GeBM against our initial approach (whose unsuitability we have argued for in Section 3, but we use here in order to further solidify our case). As an initial state for GeBM, we use $\mathbf{C}^{\dagger}\mathbf{y}(0)$, and for Model$_0$, we simply use $\mathbf{y}(0)$. The final time-series we show, both for the real data and the estimated ones are normalized to unit norm, and plotted in absolute values. For exposition purposes, we sorted the voxels according to the ℓ_2 norm of their time series vector, and we are displaying the high ranking ones (however, the same pattern holds for all voxels)

In Fig. 7 we illustrate the simulated brain activity of GeBM (solid red), compared against the ones of Model$_0$ (using LS (dash-dot magenta) and CCA (dashed black)), as well as the original brain activity time series (dashed blue) for the four highest ranking voxels. Clearly, the activity generated using GeBM is far more realistic than the results of Model$_0$.

5.4 D4: Explanation of Psychological Phenomena

As we briefly mentioned in the Introduction, we would like our proposed method to be able to capture some of the psychological phenomena that the human brain exhibits. We, by no means, claim that GeBM is able to capture convoluted and still under heavy investigation psychological phenomena, however, in this section we demonstrate GeBM's ability to simulate two very basic phenomena, *habituation* and *priming*. Unlike the previous discoveries, the following experiments are on synthetic data and their purpose is to showcase GeBM's additional strengths.

Habituation In our simplified version of habituation, we observe the demand behaviour: Given a repeated stimulus, the neurons initially get activated, but their activation levels decline ($t = 60$ in Fig. 8) if the stimulus persists for a long time ($t = 80$ in Fig. 8). In Fig. 8, we show that GeBM is able to capture such behavior. In particular, we show the desired input and output for a few (observed) voxels, and we show, given the functional connectivity obtained according to GeBM, the simulated output, which exhibits the same, desired behavior.

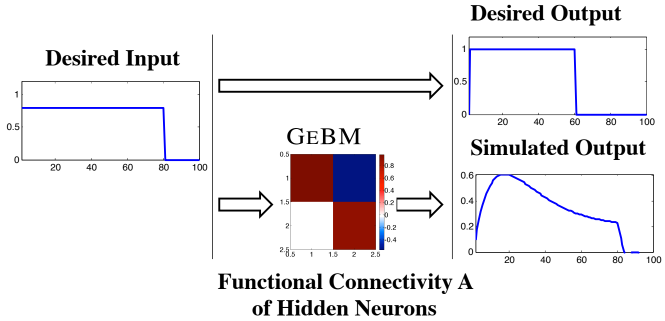


Figure 8: GeBM captures *Habituation*: Given repeated exposure to a stimulus, the brain activity starts to fade.

Priming In our simplified model on priming, first we give the stim-

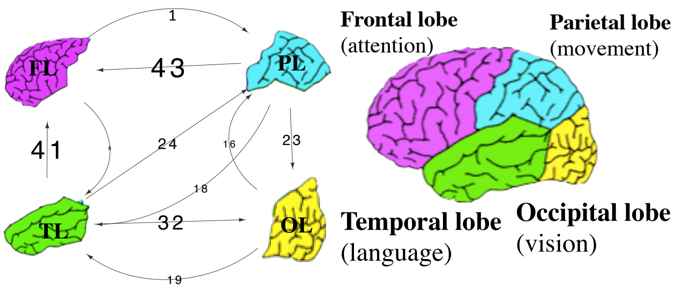
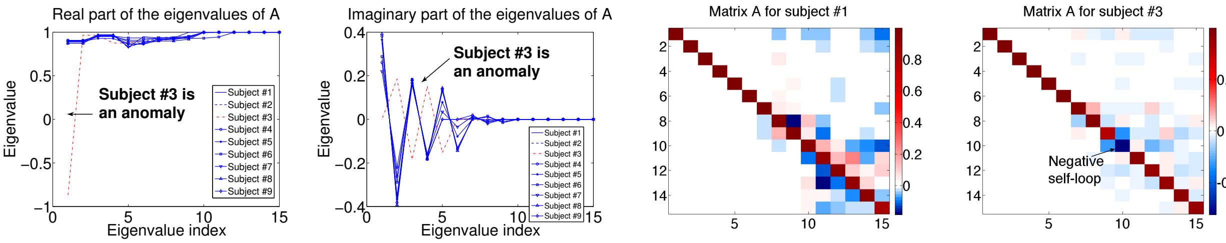


Figure 5: The functional connectivity derived from GeBM. The weights on the edges indicate the number of inferred connections. Our results are consistent with research that investigates natural language processing in the brain.



(a) Real part of eigenvalues

(b) Imaginary part of eigenvalues

(c) Subject #1

(d) Subject #3

Figure 6: Multi-subject analysis: Sub-figures (a) and (b), show the real and imaginary parts of the eigenvalues of matrix $\mathbf{A}$ for each subject. For all subjects but one (subject #3) the eigenvalues are almost identical, implying that the GeBM that captures their brain activity behaves more or less in the same way. Subject #3 on the other hand is an outlier; indeed, during the experiment, the subject complained that he was able to hear a demonstration happening outside of the laboratory, rendering the experimental task assigned to the subject more difficult than it was supposed to be. Sub-figures (c) and (d) show matrices $\mathbf{A}$ for subject #1 and #3. Subject #3's matrix seems sparser and most importantly, we can see that there is a negative entry on the diagonal, a fact unique to subject #3.

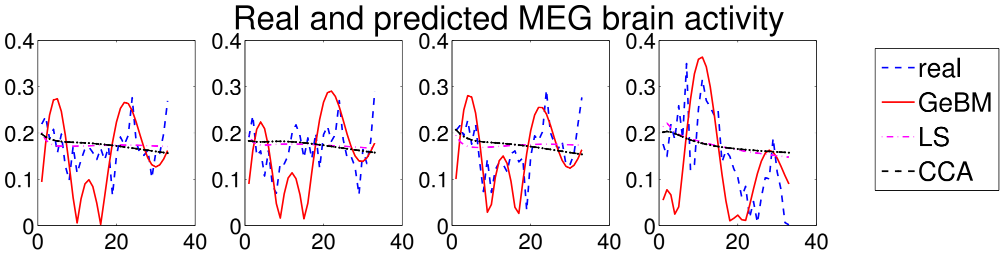


Figure 7: Effective brain activity simulation: Comparison of he real brain activity and the simulated ones using GeBM and Model$_0$, for the first four high ranking voxels (in the ℓ_2 norm sense).

ulus *apple*, which sets off neurons that are associated with the fruit 'apple', as well as neurons that are associated with Apple inc. Consequently, we are showing a stimulus such as *iPod*; this predisposes the regions of the brain that are associated with Apple inc. to display some small level of activation, whereas suppressing the regions of the brain that are associate with apple (the fruit). Later on, the stimulus *apple* is repeated, which, given the aforementioned predisposition, activates the voxels associated with Apple (company) and suppresses the ones associated with the homonymous fruit.

Figure 9 displays is a pictorial description of the above example of priming; given desired input/output pairs, we derive a model that obeys GeBM, such that we match the priming behavior.

6. RELATED WORK

Brain Functional Connectivity Estimating the brain's functional connectivity is an active field of study of computational neuroscience. Examples of works can be found in [13, 8, 7]. There have been a few works in the data mining community as well: In [16], the authors derive the brain region connections for Alzheimer's patients, and recently [5] that leverages tensor decomposition in order to discover the underlying network of the human brain. Most related to the present work is the work of Valdes et al [19], wherein the authors propose an autoregressive model (similar to Model$_0$) and solve it using regularized regression. However, to the best of our knowledge, this work is the first to apply system identification concepts to this problem.

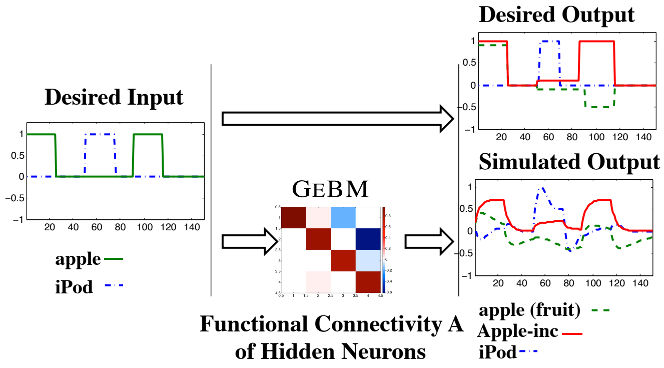


Figure 9: GEBM captures *Priming*: When first shown the stimulus *apple*, both neurons associated with the fruit 'apple' and Apple inc. get activated. When showing the stimulus *iPod* and then *apple*, *iPod* predisposes the neurons associated with Apple inc. to get activated more quickly, while suppressing the ones associated with the fruit.

Psychological Phenomena A concise overview of literature pertaining to habitutation can be found in [17]. A more recent study on habitutation can be found in [12]. The definition of priming, as we describe it in the lines above concurs with the definition found in [6]. Additionally, in [11], the authors conduct a study on the effects of priming when the human subjects were asked to write sentences. The above concepts of priming and habituation have been also studied in the context of spreading activation [1, 4] which is a model of the cognitive process of memory.

Control Theory & System Identification System Identification is a field of control theory. In the appendix we provide more theoretical details on subspace system identification, however, [10] and [21] are the most prominent sources for system identification algorithms.

Network Discovery from Time Series Our work touches upon discovering underlying network structures from time series data; an exemplary work related to the present paper is [22] where the authors derive a who-calls-whom network from VoIP packet transmission time series.

7. CONCLUSIONS

The list of our contributions is:

- **Analytical model** : We propose GEBM, a novel model of the human brain functional connectivity.
- **Algorithm**: We introduce SPARSE-SYSID, a novel sparse system identification algorithm that estimates GEBM
- **Effectiveness**: GEBM simulates psychological phenomena (such as habituation and priming), as well as provides valuable neuroscientific insights.
- **Validation**: We validate our approach on synthetic data (where the ground truth is known), and on real data, where our model produces brain activity patterns, remarkably similar to the true ones.
- **Multi-subject analysis**: We analyze measurements from 9 human subjects, identifying a consistent connectivity among 8 of them; we successfully identify an outlier, whose experimental procedure was compromised.

Acknowledgements

Research was supported by grants NSF IIS-1247489, NSF IIS-1247632, NSF CDI 0835797, NIH/NICHD 12165321, DARPA FA87501320005, IARPA FA865013C7360, and by Google. This work was also supported in part by a fellowship to Alona Fyshe from the Multimodal Neuroimaging Training Program funded by NIH grants T90DA022761 and R90DA023420. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding parties. We would also like to thank Erika Laing and the UPMC Brain Mapping Center for their help with the data and visualization of our results. Finally, we would like to thank Gustavo Sudre for collecting and sharing the data, and Dr. Brian Murphy for valuable conversations.

8. REFERENCES

[1] J.R. Anderson. A spreading activation theory of memory. *Journal of verbal learning and verbal behavior*, 22(3):261–95, 1983.

[2] Frederico AC Azevedo, Ludmila RB Carvalho, Lea T Grinberg, José Marcelo Farfel, Renata EL Ferretti, Renata EP Leite, Roberto Lent, Suzana Herculano-Houzel, et al. Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *Journal of Comparative Neurology*, 513(5):532–541, 2009.

[3] Jeffrey R Binder and Rutvik H Desai. The neurobiology of semantic memory. *Trends in cognitive sciences*, 15(11):527–36, November 2011.

[4] Allan M. Collins and Elizabeth F. Loftus. A spreading-activation theory of semantic processing. *Psychological Review*, 82(6):407–428, 1975.

[5] Ian Davidson, Sean Gilpin, Owen Carmichael, and Peter Walker. Network discovery via constrained tensor analysis of fmri data. In *ACM SIGKDD*, pages 194–202. ACM, 2013.

[6] Angela D Friederici, Karsten Steinhauer, and Stefan Frisch. Lexical integration: Sequential effects of syntactic and semantic information. *Memory & Cognition*, 27(3):438–453, 1999.

[7] Alona Fyshe, Emily B Fox, David B Dunson, and Tom M Mitchell. Hierarchical latent dictionaries for models of brain activation. In *AISTATS*, pages 409–421, 2012.

[8] Michael D Greicius, Ben Krasnow, Allan L Reiss, and Vinod Menon. Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *PNAS*, 100(1):253–258, 2003.

[9] Gregory Hickok and David Poeppel. Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition*, 92(1-2):67–99, 2004.

[10] Lennart Ljung. *System identification*. Wiley Online Library, 1999.

[11] Martin J Pickering and Holly P Branigan. The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language*, 39(4):633–651, 1998.

[12] Catharine H Rankin, Thomas Abrams, Robert J Barry, Seema Bhatnagar, David F Clayton, John Colombo, Gianluca Coppola, Mark A Geyer, David L Glanzman, Stephen Marsland, et al. Habituation revisited: an updated and revised description of the behavioral characteristics of habituation. *Neurobiology of learning and memory*, 92(2):135–138, 2009.

[13] Vangelis Sakkalis. Review of advanced techniques for the estimation of brain connectivity measured with eeg/meg.

Computers in biology and medicine, 41(12):1110–1117, 2011.

[14] Ioannis D Schizas, Georgios B Giannakis, and Zhi-Quan Luo. Distributed estimation using reduced-dimensionality sensor observations. *IEEE TSP*, 55(8):4284–4299, 2007.

[15] G. Sudre, D. Pomerleau, M. Palatucci, L. Wehbe, A. Fyshe, R. Salmelin, and T. Mitchell. Tracking neural coding of perceptual and semantic features of concrete nouns. *NeuroImage*, 2012.

[16] Liang Sun, Rinkal Patel, Jun Liu, Kewei Chen, Teresa Wu, Jing Li, Eric Reiman, and Jieping Ye. Mining brain region connectivity for alzheimer's disease study via sparse inverse covariance estimation. In *ACM SIGKDD*, pages 1335–1344. ACM, 2009.

[17] Richard F Thompson and William A Spencer. Habituation: a model phenomenon for the study of neuronal substrates of behavior. *Psychological review*, 73(1):16, 1966.

[18] Dardo Tomasi and Nora D Volkow. Resting functional connectivity of language networks: characterization and reproducibility. *Molecular psychiatry*, 17(8):841–854, 2012.

[19] Pedro A Valdés-Sosa, Jose M Sánchez-Bornot, Agustín Lage-Castellanos, Mayrim Vega-Hernández, Jorge Bosch-Bayard, Lester Melie-García, and Erick Canales-Rodríguez. Estimating brain functional connectivity with sparse multivariate autoregression. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1457):969–981, 2005.

[20] Michel Verhaegen. Identification of the deterministic part of mimo state space models given in innovations form from input-output data. *Automatica*, 30(1):61–74, 1994.

[21] Michel Verhaegen and Vincent Verdult. *Filtering and system identification: a least squares approach*. Cambridge university press, 2007.

[22] Olivier Verscheure, Michail Vlachos, Aris Anagnostopoulos, Pascal Frossard, Eric Bouillet, and Philip S Yu. Finding "who is talking to whom" in voip networks via progressive stream clustering. In *IEEE ICDM*, pages 667–677. IEEE, 2006.

[23] Robert W Williams and Karl Herrup. The control of neuron number. *Annual review of neuroscience*, 11(1):423–453, 1988.

APPENDIX

A. SYSTEM IDENTIFICATION FOR GeBM

Consider again the linear state-space model of GEBM

$$\mathbf{x}(t+1) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t), \;\; \mathbf{y}(t) = \mathbf{C}\mathbf{x}(t).$$

Assuming $\mathbf{x}(0) = \mathbf{0}$, for simplicity, it is easy to see that

$$\mathbf{x}(t) = \sum_{i=0}^{t-1} \mathbf{A}^{t-1-i}\mathbf{B}\mathbf{u}(i),$$

and therefore $\mathbf{y}(t) = \sum_{i=0}^{t-1} \mathbf{C}\mathbf{A}^{t-1-i}\mathbf{B}\mathbf{u}(i)$, from which we can read out the *impulse response matrix* $\mathbf{H}(t) = \mathbf{C}\mathbf{A}^{t-1}\mathbf{B}$.

As we mentioned in Section 3, matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}$ can be identified up to a similarity transformation. In order to show this, suppose that we have access to the system's inputs $\{\mathbf{u}(t)\}_{t=0}^{T-1}$ and outputs $\{\mathbf{y}(t)\}_{t=1}^{T}$, and we wish to identify the system matrices $(\mathbf{A}, \mathbf{B}, \mathbf{C})$. A first important observation is the following. From $\mathbf{x}(t+1) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t)$ we obtain

$$\mathbf{M}\mathbf{x}(t+1) = \mathbf{M}\mathbf{A}\mathbf{x}(t) + \mathbf{M}\mathbf{B}\mathbf{u}(t) =$$

$$\mathbf{M}\mathbf{A}\mathbf{M}^{-1}\mathbf{M}\mathbf{x}(t) + \mathbf{M}\mathbf{B}\mathbf{u}(t).$$

Defining $\mathbf{z}(t) := \mathbf{M}\mathbf{x}(t)$, $\tilde{\mathbf{A}} := \mathbf{M}\mathbf{A}\mathbf{M}^{-1}$, and $\tilde{\mathbf{B}} := \mathbf{M}\mathbf{B}$, we obtain

$$\mathbf{z}(t+1) = \tilde{\mathbf{A}}\mathbf{z}(t) + \tilde{\mathbf{B}}\mathbf{u}(t),$$

and with $\tilde{\mathbf{C}} := \mathbf{C}\mathbf{M}^{-1}$, we also have

$$\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) = \mathbf{C}\mathbf{M}^{-1}\mathbf{M}\mathbf{x}(t) = \tilde{\mathbf{C}}\mathbf{z}(t).$$

It follows that $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ and $(\tilde{\mathbf{A}}, \tilde{\mathbf{B}}, \tilde{\mathbf{C}}) = (\mathbf{M}\mathbf{A}\mathbf{M}^{-1}, \mathbf{M}\mathbf{B}, \mathbf{C}\mathbf{M}^{-1})$ are indistinguishable from input-output data alone. Thus, the sought parameters can only be (possibly) identified up to a basis (similarity) transformation, in the absence of any other prior or side information.

Under what conditions can $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ be identified up to such similarity transformation? It has been shown by Kalman that if the so-called *controlability* matrix

$$\mathcal{C} := \left[\mathbf{B}, \mathbf{A}\mathbf{B}, \mathbf{A}^2\mathbf{B}, \cdots, \mathbf{A}^{n-1}\mathbf{B}\right]$$

is full row rank n, and the *observability matrix*

$$\mathcal{O} := \begin{bmatrix}\mathbf{C} & \mathbf{C}\mathbf{A} & \mathbf{C}\mathbf{A}^2 & \cdots & \mathbf{C}\mathbf{A}^{n-1}\end{bmatrix}^T$$

is full column rank n, then it is possible to identify $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ up to such similarity transformation from (sufficiently 'diverse') input-output data, and the impulse response in particular. This can be accomplished by forming a block Hankel matrix out of the impulse response, as follows,

$$\begin{bmatrix} \mathbf{H}(1) & \mathbf{H}(2) & \mathbf{H}(3) & \cdots & \mathbf{H}(n) \\ \mathbf{H}(2) & \mathbf{H}(3) & \cdots & & \\ \vdots & & & & \\ \mathbf{H}(n) & & & & \mathbf{H}(2n+1) \end{bmatrix} =$$

$$\begin{bmatrix} \mathbf{C}\mathbf{B} & \mathbf{C}\mathbf{A}\mathbf{B} & \mathbf{C}\mathbf{A}^2\mathbf{B} & \cdots & \mathbf{C}\mathbf{A}^{n-1}\mathbf{B} \\ \mathbf{C}\mathbf{A}\mathbf{B} & \mathbf{C}\mathbf{A}^2\mathbf{B} & \cdots & & \\ \vdots & & & & \\ \mathbf{C}\mathbf{A}^{n-1}\mathbf{B} & & & & \mathbf{C}\mathbf{A}^{2n-1}\mathbf{B} \end{bmatrix}.$$

This matrix can be factored into $\mathcal{O}\mathbf{M}\mathbf{M}^{-1}\mathcal{C}$, i.e., the 'true' $\mathcal{O}$ and $\mathcal{C}$ up to similarity transformation, using the singular value decomposition of the above Hankel matrix. It is then easy to recover $\tilde{\mathbf{A}}, \tilde{\mathbf{B}}, \tilde{\mathbf{C}}$ from $\mathcal{O}\mathbf{M}$ and $\mathbf{M}^{-1}\mathcal{C}$. This is the core of the Kalman-Ho algorithm for Hankel subspace-based identification [10].

This procedure enables us to identify $\tilde{\mathbf{A}}, \tilde{\mathbf{B}}, \tilde{\mathbf{C}}$, but, in our context, we are ultimately also interested in the true latent $(\mathbf{A}, \mathbf{B}, \mathbf{C})$.

PROPOSITION 1. *We can always, without loss of generality, transform $\tilde{\mathbf{A}}$ to maximal sparsity while keeping $\tilde{\mathbf{B}}$ and $\tilde{\mathbf{C}}$ dense - that is, sparsity of $\mathbf{A}$ alone does not help in terms of identification.*

PROOF. Suppose that we are interested in estimating a sparse $\mathbf{A}$, while preserving the eigenvalues of of $\tilde{\mathbf{A}}$. We therefore seek a similarity transformation (a matrix $\mathbf{M}$) that will render $\mathbf{A} = \mathbf{M}\tilde{\mathbf{A}}\mathbf{M}^{-1}$ as sparse as possible. Towards this end, assume that $\tilde{\mathbf{A}}$ has a full set of linearly independent eigenvectors, collected in matrix $\mathbf{E}$, and let Λ be the corresponding diagonal matrix of eigenvalues. Then, clearly, $\tilde{\mathbf{A}}\mathbf{E} = \mathbf{E}\Lambda$, and therefore $\mathbf{E}^{-1}\tilde{\mathbf{A}}\mathbf{E} = \Lambda$ - a diagonal matrix. Hence choosing $\mathbf{M} = \mathbf{E}^{-1}$ we make $\mathbf{A} = \mathbf{M}\tilde{\mathbf{A}}\mathbf{M}^{-1}$ maximally sparse. Note that $\mathbf{A}$ must have the same rank as $\tilde{\mathbf{A}}$, and if it has less nonzero elements it will also have lower rank. Finally, it is easy to see that if we apply this similarity transformation, the eigenvalues of $\tilde{\mathbf{A}}$ do not change. □