NeuroImage 58 (2011) 416–441

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

# Full-brain auto-regressive modeling (FARM) using fMRI

Rahul Garg *, Guillermo A. Cecchi, A. Ravishankar Rao

*Computational Biology Center, IBM T.J. Watson Research Center, Yorktown Heights, NY 10598, USA*

ARTICLE INFO

*Article history:*
Received 21 July 2010
Revised 10 February 2011
Accepted 27 February 2011
Available online 22 March 2011

*Keywords:*
Granger causality
Auto-regressive modeling
Functional MRI (fMRI)
Dynamical systems
Brain dynamics
Default mode networks
Resting state analysis
Prediction

ABSTRACT

In order to fully uncover the information potentially available in the fMRI signal, we model it as a multivariate auto-regressive process. To infer the model, we do not apply any form of clustering or dimensionality reduction, and solve the problem of under-determinacy using *sparse regression*. We find that only a few small clusters (with average size of 3–4 voxels) are useful in predicting the activity of other voxels, and demonstrate remarkable consistency within a subject as well as across multiple subjects. Moreover, we find that: (a) the areas that can predict activity of other voxels are consistent with previous results related to networks activated by the specific somatosensory task, as well as networks related to the default mode activity; (b) there is a global dynamical state dominated by two prominent (although not unique) *streams*, originating in the posterior parietal cortex and the posterior cingulate/precuneus cortex; (c) these streams span default mode and task-specific networks, and interact in several regions, notably the insula; and (d) the posterior cingulate is a central node of the default mode network, in terms of its ability to determine the future evolution of the rest of the nodes.



## Introduction

### *Background*

The brain is a highly interconnected dynamical system, in which the activity and temporal evolution of its elements are primarily determined by the activity of other elements. Though the average connectivity of neurons is small relative to the total number of possible connections, it is still quite high in absolute terms, and includes long-range links that make the brain a “small-world” graph (Smith Basset and Bullmore, 2006).

The brain's response to external stimuli has been often compared to “ripples in a pond” or disturbances of an “oneiric-like state” even in wakefulness (Koch and Davis, 1995). Experimental evidence indicates that ongoing brain activity (i.e. not elicited by stimuli), the reflection of a *dynamic core* (Koch and Davis, 1995; Edelman and Tononi, 2001), has a structured temporal evolution (Arieli et al., 1996). Examples of such activity include the maintenance of a normal wakeful state and attentional and emotional biases, which are known to provide a strong top-down modulation of even early sensorial processing (Fox et al., 2005; Li et al., 2001). Moreover, an explicit consideration of neural correlations – as opposed to external correlates – can lead to significant insights, such as the presence of otherwise undetectable collective states in local as well as distributed networks (Schneidman et al., 2006; Cecchi et al., 2009).

The identification of the system known as the default mode network (Fox et al., 2005; Raichle et al., 2001) has only made more urgent the need to analyze the validity of the dynamic core hypothesis: it postulates the presence of an extended network whose components evolve relatively slowly, and can be activated or deactivated *en masse* by an attention-demanding task, implying that brain function cannot be parsed as a set of separated modules implementing functions at well-defined segments of time. The association of the task-negative default mode network, as identified by (Raichle et al., 2001), with a system of internal and external sensory monitors that is tuned down during attention (Raichle et al., 2001), is another strong indication of a non-trivial interaction between the intrinsic time-scale of the default mode (<0.1 Hz) and the less constrained one of the task.

The field of functional brain imaging is, however, largely constrained by an analytic framework that focuses on the relationship between the activation of an area and the presence or absence of a task or stimulus, while ignoring temporal and spatial correlations that may influence local responses as much as the experimental paradigm. This is the approach used by the general linear model (GLM), which currently dominates the field of fMRI image analysis. (Grinband et al., 2008) identified 170 papers published in leading journals during the first six months of 2007 using this approach alone. Though the GLM approach has demonstrated its utility, and has resulted in many insights into brain function, it has limited use in being applied across a wide variety of experimental protocols, especially in the case of resting state and similar brain states (Li et al., 2009). Recently, the introduction of machine-learning algorithms for fMRI analysis has demonstrated that there is a wealth of functional information in

* Corresponding author. Fax: +1 914 945 4217.
*E-mail addresses:* grahul@us.ibm.com (R. Garg), gcecchi@us.ibm.com (G.A. Cecchi), ravirao@us.ibm.com (A.R. Rao).

1053-8119/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2011.02.074

distributed patterns of activation, even though specific inter-voxel interactions are usually not modeled (Norman et al., 2006; Mitchell et al., 2008; Carroll et al., 2009). An alternate method for analysis is offered by functional connectivity-based methods, which have been receiving increasing attention in the field (Li et al., 2009; Eguiluz et al., 2005; Cecchi et al., 2007). In this paper we advance the state-of-the-art in functional connectivity by developing a whole-brain method based on auto-regressive modeling. Our approach captures, in a single framework, the dynamics and spatial correlations of the brain using fMRI measurements from individual voxels. It also leads to the identification of the "Granger causality" relationships among brain voxels; under the daring but falsifiable *information flow hypothesis* which states that the structure of the information flow in the fMRI data reflects, however partially, the flow of information in the underlying neural networks. Our approach provides an interpretation of functional causation based on the ability of one region of the brain to predict the future activity of another one.

Early functional connectivity methods used correlation metrics (Dodel et al., 2002; Eguiluz et al., 2005), which do not provide directional information for the links in the network. Delayed correlation analysis can be used to infer link directionality by including part of the temporal structure of voxel–voxel correlations (Cecchi et al., 2007), though this is not a robust method as it considers only pairwise interactions. A more complete model was presented in (Cecchi et al. 2008a), which introduced the idea of using a sparse regression framework to estimate an auto-regressive model that best explained the measured fMRI time sequences. The idea of Granger causality is that a variable $X$ is the cause of another $Y$ if the past of $X$ can be used to increase the predictability of the future of $Y$ (Granger, 1969); for practical reasons, auto-regressive modeling is typically utilized to infer Granger causality. One limitation for the applicability of Granger causality to fMRI data is that auto-regressive models at the voxel level are highly under-constrained, and therefore aggregation of voxels is required to reduce dimensionality and match it with the number of independent data samples (Goebel et al., 2003). We demonstrate that this aggregation process can lead to severe artifacts; here we propose the utilization of sparse (i.e. constrained) regression as a means to overcome the dimensionality problem, while preserving voxel resolution.

It is evident then that an accurate description of functional data should include not only explicit dynamical components, but also a means to *discover* dynamical interactions, as opposed to just test the hypothesis of their significance based on a priori knowledge. Most functional dynamics methods, however, are constrained by the need to aggregate voxels in pre-defined regions of interests (Goebel et al., 2003; Friston et al., 2003). Voxel aggregation potentially implies discarding useful voxel-level information, and inviting artifacts unless very carefully performed. Here we demonstrate that a voxel-wise auto-regressive model is computationally feasible, and yields results that are consistent with the current knowledge of areas involved in the task under study – finger tapping – as well as with the task-negative and task-positive systems (Fox et al., 2005); in particular, the analysis provides a novel interpretation of the previously proposed role of the precuneus as a "central node" of the default mode network. The findings validate the analytic approach, and point to new directions for how the interaction between task-driven responses and ongoing brain activity may be interpreted, by providing a model for the global dynamical state of the brain.

### *Related work*

The simplest and most popular approaches to model interactions among different brain regions are based on the notion of *functional connectivity* (Friston, 1994). Some of these approaches use pair-wise voxel correlations (Eguiluz et al., 2005) while others use more sophisticated methods (Cecchi et al., 2007; Achard et al., 2006; Bullmore and Sporns, 2009; Cecchi et al., 2008b) to generate networks of brain interactions. All these approaches suffer from the problem of spurious connections due to confounding effects.

The approaches that attempt to alleviate this problem are termed as *effective connectivity* analysis methods (Friston, 1994). These include dynamic casual modeling (Friston et al., 2003; Penny et al., 2004), structural equation models (SEM) (Pearl, 1998; Mcintosh and Gonzalez-Lima, 1994; Buchel and Friston, 1997), graphical models (Eichler, 2005; Dahlhaus and Eichler, 2003) and Granger causality analysis (Granger, 1969; Seth, 2005; Sato et al., 2007). The structural equation models are based on a priori specification of all the potential interactions among the regions in the brain. Dynamic causal modeling, which is based on a non-linear model of neuronal interactions, also requires a priori specification of the experimental hypothesis. To minimize the confounding effects, approaches such as partial correlations (Salvador et al., 2005; Marrelec et al., 2006) and partial mutual information (Sun et al., 2004; Salvador et al., 2007) have also been used for fMRI data analysis. Due to the problem of dimensionality, these approaches also need to aggregate the data into a small number of pre-specified regions of interest.

Granger causality analysis is almost synonymous with multivariate auto-regressive modeling since the Granger causality relationships are mostly determined by solving the model identification problem for a suitably defined multivariate auto-regressive model (see Auto-regressive modeling section). It can be applied in a purely data driven manner, but requires very good temporal data resolution. For this reason, it has mostly been applied to modalities such as EEG (Brovelli et al., 2004; Pereda et al., 2005; Anderson et al., 1998), MEG (Darvas and Leahy, 2007) and local field potential recordings (Brovelli et al., 2004; Kamiński et al., 2001). The application of Granger causality to fMRI data has either been restricted to pair-wise causality (bi-variate analysis) as in (Goebel et al. 2003) and (Roebroeck et al. 2005) or analysis using a small number of pre-specified regions of interest (Seth, 2005; Chen et al., 2009; Deshpande et al., 2009; Harrison et al., 2003). Thus, for functional imaging data, the analysis either becomes susceptible to confounding effects or needs accurate specification of the key regions of interest and experimental hypothesis.

The work by (Valdes-Sosa et al., 2005) solved the model identification problem in Granger causality analysis using a variety of methods, including the use of the lasso regression (Tibshirani, 1996) which we have chosen to follow. They concluded that the performance of different techniques, as measured by the ROC curve area, does not differ significantly. While applying this technique to functional imaging, the fMRI data were aggregated into 116 regions — a step that was not really required for using their technique. Therefore it was not possible to uncover certain key findings that we report in this paper. Since we consider all the brain voxels at every step without carrying out any data aggregation, we get a more accurate and parsimonious model of the underlying brain dynamics, as described in subsequent sections.

### *Organization of the paper*

In the Auto-regressive modeling section, we present our methodology of modeling neuronal interactions using fMRI data, with our solution of the model and simulations demonstrating that our approach does recover the true model satisfactorily. In the Model interpretation section, we present two approaches to visualize and interpret the model. Finally, in the Evaluation on a sample data set section we present the results of our analysis on a simple finger-tapping experiment. We conclude in the Conclusions section along with directions for further work.

## Auto-regressive modeling

### *Granger causality using multivariate auto-regressive model*

The metaphysical concept of causality has been a matter of intense debate and discussion for ages (Machamer and Wolters, 2007).

However, Granger defined causality for economic processes in strictly mathematical terms, using the notion of predictability of stationary stochastic processes (Granger, 1969). In this approach, time series data from various sources are modeled as stationary stochastic processes. Process $i$ is said to have a causal influence on process $j$ if the past values of process $i$ can improve the prediction accuracy of process $j$ in the presence of past data from all the other processes in the universe.

### Theory

Consider a universe $U$ of $n$ discrete-time stationary stochastic processes represented as $X_1, X_2, \ldots X_n$. Let $X_i(t)$ represent the random variable corresponding to the stochastic process $X_i$ at time $t$ and $X(t)$ represent the $n$-dimensional vector $(X_1(t), X_2(t), \ldots, X_n(t))$. Let $\overline{X}(t)$ represent the set of past random variables in $X$ i.e., $\overline{X}(t) = \{X(t-j) : j = 1, 2, \ldots \infty\}$. Let $P(A|\overline{B})$ represent the optimal unbiased least variance predictor of the random variable $A$ using only the random variables in the set $\overline{B}$. Thus $P(X_i(t)|\overline{X}_i(t))$ represent the optimal unbiased least variance predictor of $X_i(t)$ using only the past values of $X_i(t)$. Let $\sigma^2(A|\overline{B})$ be the variance of the stochastic process $A - P(A|\overline{B})$. Since, the analysis is restricted to stationary processes, for convenience of notation, we drop the time indices while writing $\sigma^2$. The process $X_j$ is said to have a causal influence on the process $X_i$ in the context of processes $X$, if

$$\sigma^2(X_i|\overline{X}) < \sigma^2\left(X_i|\overline{X}\backslash\overline{X}_j\right).$$

Stated simply, removing the past values of $X_j$ from the universe of past data available to predict $X_i$ increases the prediction error of $X_i$. Therefore, $X_j$ contains some unique information about the future of $X_i$ which is not present in the rest of the data.

The above definition of causality is very general. It is defined in the existential form, for arbitrary stationary stochastic processes, using the abstract notion of "optimal unbiased least variance predictor", without the specification of any method to compute it. Also, the above definition of causality is critically dependent on the context of the stochastic processes being studied. As Granger pointed out himself, the apparent causality relationship between two stochastic processes may disappear after the inclusion of a confounding process in the context ($X$) of processes being considered (see also the illustrations in Some example networks). Granger causality is a mathematical definition of a concept which may or may not reflect the metaphysical concept of cause and effect.

### The linear simplification

For application of the above definition of causality in practice, Granger resorted to a linear simplification where the "optimal unbiased least variance predictor" was restricted to an "optimal unbiased *linear* least variance predictor". With this simplification, the causality relationships can be determined by solving a suitable multivariate auto-regressive model (MAR) of the form given below:

$$X(t) = \sum_{\tau=1}^{k} A(\tau)X(t-\tau) + E(t). \quad (1)$$

The parameter $k$ in the above equation is called the *model order*; $A(\tau)_{\tau=1\ldots k}$ are the model parameters in the form of $k$ matrices of size $n \times n$ (with coefficients $a_{ij}(\tau)$), $E(t)$ is a $n$-dimensional vector of noise with zero mean (and a covariance equal to $C$). For any $t_1 \neq t_2$, $E(t_1)$ and $E(t_2)$ are identically distributed and uncorrelated.

Now, the model coefficients $a_{ij}(\tau)$ encode the causality relationships among the processes in $X$. If $a_{ij}(\tau) > 0$ for some $\tau$, then the past values of $X_j$ improve the predictability of $X_i$ and therefore, $X_j$ is said to have a causal influence on $X_i$. The parameter $\tau$ is called the causality lag between $X_j$ and $X_i$.

### Some example networks

In the example network of Fig. 1, the exogenous input to the system $s(t)$ drives the node $v_1$ as $v_1(t) = s(t) * h(t)$ where $h(t)$ is the convolution kernel (such as the hemodynamic response function) translating the input into the response. The node $v_1$ has an excitatory influence on $v_2$ and $v_3$; $v_3$ and $v_4$ have excitatory and inhibitory influences on each other respectively. The equations governing the temporal dynamics of the system are given in Fig. 2, where $\eta_i(t)$ represents the combination of noise and endogenous activities at the respective nodes in the network. Fig. 3 shows the evolution of the system for a sample input.

The goal is to recover the qualitative model of Fig. 1 (and possibly the quantitative model of Fig. 2) using the observed system behavior of Fig. 3.

The general linear model (GLM) is a standard approach in functional neuroimaging that finds the nodes that are activated in response to the exogenous input $s(t)$ (Friston et al., 2007). For the example above, nodes $v_1$, $v_2$ and $v_3$ are found active as a response to the input $s(t)$ as shown in Fig. 4(a).

Methods of functional network identification based on voxel–voxel correlations, or similar pair-wise relational measures, are also used in many studies (Eguiluz et al., 2005; Dodel et al., 2002; Stam et al., 2007). The correlation matrix for the above example is shown in Fig. 4(d). Thresholding this matrix (using a value of 0.25) gives the network shown in Fig. 4(b). Another approach to identify functional networks is the use of delayed cross correlations (Cecchi et al., 2007). The lag-1 correlations in the above example are shown in Fig. 4(e) which in turn gives the directed network of Fig. 4(c) (using a threshold of 0.29). None of these approaches is able to identify the correct system dynamics, primarily because they consider one voxel pair at a time and ignore the confounding effects of other voxels. The multivariate auto-regressive modeling of the system described in this paper and in (Cecchi et al., 2008a) identifies the correct relationships between the nodes.

Fig. 5 represents a three layer network where an exogenous input drives a single node in the first layer which, in turn, drives four nodes in the second layer. The nodes in the second layer drive the third layer as shown in the figure. If the input signal has large lag-1 and lag-2 auto-correlation (such as the fMRI BOLD signal), the GLM analysis identifies all the nodes in the three layers as active in response to the input. The functional connectivity analysis, as shown in Fig. 5(c), identifies not only the links driving the above system, but also many spurious links due to the confounding effects of a common source. The cross-correlation analysis correctly identifies the direction of the information flow (as shown in Fig. 5(d)), but also suffers from the problem of a common source. The pairwise Granger causality analysis (Goebel et al., 2003; Roebroeck et al., 2005) has the same problem as it

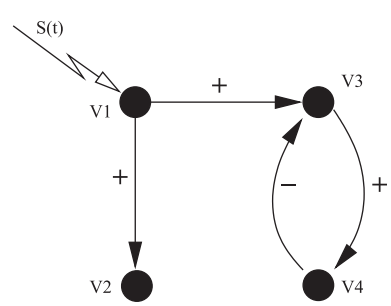



**Fig. 1.** An example network.

$$
\begin{aligned}
v_1(t) &= s(t) * h(t) \\
v_2(t) &= v_1(t-1) + 0.5\ \eta_2(t) \\
v_3(t) &= v_1(t-1) - 0.5\ v_4(t-1) + 0.5\ \eta_3(t) \\
v_4(t) &= 0.5\ v_3(t-1) + 0.5\ \eta_4(t)
\end{aligned}
$$

**Fig. 2.** Equations governing system dynamics.

does not consider the simultaneous interactions of all the voxels. The full-brain auto-regressive modeling (FARM) proposed in this paper correctly identifies all the interactions of this system.

A commonly encountered problem in the auto-regressive modeling of fMRI data is the lack of adequate temporal resolution for reliably uncovering the interactions between every pair of voxels in the brain (see The problem of model identification). One approach to address this problem is to identify a small number of regions of interest (ROIs) in the brain and then find causality relationships across the ROIs (Friston et al., 2003; Penny et al., 2004; Buchel and Friston, 1997). This is usually done by aggregating the data within each ROI. Such an approach works very well if the experimental hypothesis is narrowly defined and the ROIs are accurately identified. In general, since data aggregation loses information, this approach may give incorrect results if the ROIs are incorrectly defined. Fig. 6 shows a hypothetical system driven by a periodic input $s(t)$. If the nodes $v_1 \ldots v_{10}$ on the left are aggregated into a single node (using signal averaging), then an incorrect auto-regressive model as shown in Fig. 6(c) will be inferred, whereas the desired model after aggregation is shown in Fig. 6(b). Here, averaging the responses in the nodes $v_1 \ldots v_{10}$ leads to the loss of the signal due to destructive interference, thereby giving incorrect results.

In the example network of Fig. 1, the equations governing system dynamics (Fig. 2) can be modeled as a multivariate auto-regressive model of the form (1). Thus, the Granger causality relationships correspond precisely to the arrows in Fig. 1.

### *The problem of model identification*

For the example system of Fig. 1, the problem of model identification is to infer the system dynamics of Fig. 2 using the observed system behavior of Fig. 3. The auto-regressive model may be inferred from the observations using standard techniques in statistics and time series analysis (Brockwell and Davis, 1986; Box et al., 2008; Zivot and Wang, 2006). The most common technique is the method of least squares, which gives the maximum likelihood estimate (MLE) of the model parameters $a_{ij}(\tau)$ to be tested for statistical significance.

Let $\{x(t)\}_{t=1\ldots T}$ be a $T$-step realization of the stochastic process $X$ and $\{e(t)\}_{t=1\ldots T}$ be a realization of $E$. This realization must satisfy

$$x(t) = \sum_{\tau=1}^{k} A(\tau)x(t-\tau) + e(t) \tag{2}$$

for all $t \in [k+1, \ldots T]$. The maximum likelihood estimate (assuming iid Gaussian distribution of $e(t)$) of the model parameters $a_{ij}(\tau)$ is obtained by solving the following least squares problem:

$$\min_{a_{ij}(\tau)} \sum_{i=1}^{n} \sum_{t=k+1}^{T} \left[ x_i(t) - \sum_{\tau=1}^{k} \sum_{j=1}^{n} a_{ij}(\tau) x_j(t-\tau) \right]^2. \tag{3}$$

The above set of equations can be written in a compact matrix form as:

$$W_{MLE} = \arg\min_{W} \|Y - ZW\|_2^2 \tag{4}$$

where $Y$ is an $n(T-k)$-dimensional vector obtained by stacking the $n$-dimensional vectors $(x(k+1), x(k+2), \ldots x(T))$, $W$ is a $n^2k$-dimensional vector of the model parameters $a_{ij}(\tau)$, $Z$ is a suitable matrix obtained from Eq. (3) and $\|v\|_2^2$ represents the square of the $\ell_2$ norm of the vector $v$ (i.e., the sum of square of all the components of the vector $v$).

The above system has $n^2k$ unknowns. The first order conditions for the optimality of the above system give only $n(T-k)$ linearly independent equations (Strang, 1988). Thus, the system cannot be solved reliably unless $n(T-k) \gg n^2k$. Almost all of the classical literature on the subject is applicable only when the number of temporal observations of the system is much larger than the product of the number of nodes in the network and the model order (i.e., $T \gg nk$) (Box et al., 2008; Zivot and Wang, 2006).

In a typical fMRI experiment, there is no hope of solving the problem of model identification using existing techniques, since $n$ is in the range of 20,000 voxels whereas $T$ is in the range of only about 1000 temporal observations. Even for model order 1, the system becomes under-determined with infinite solutions, with each solution over-fitting the data.

Most of the work in applying the concept of Granger causality to fMRI data has resorted to decreasing the size of data (i.e., $n$) by either aggregating the data into a small number of regions (Friston et al., 2003; Penny et al., 2004; Buchel and Friston, 1997) or by considering only pair-wise interactions among the voxels (Goebel et al., 2003; Roebroeck et al., 2005). In contrast, our approach considers all the voxels in deriving the causality relationships and carries out no aggregation of the data. Instead, it uses the technique of *sparse*

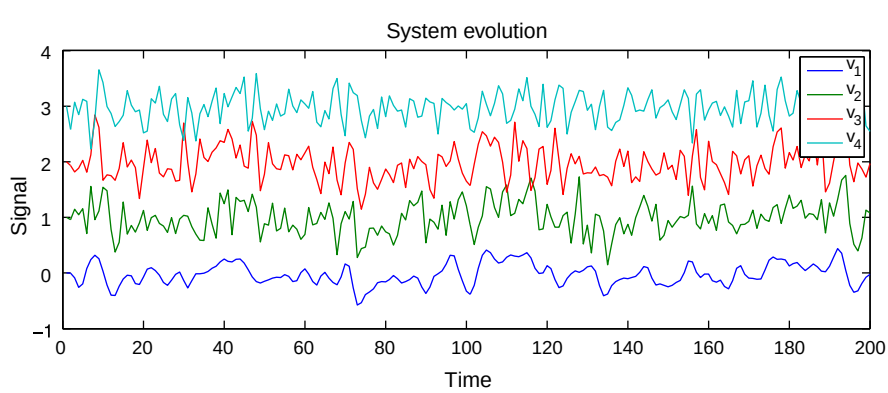
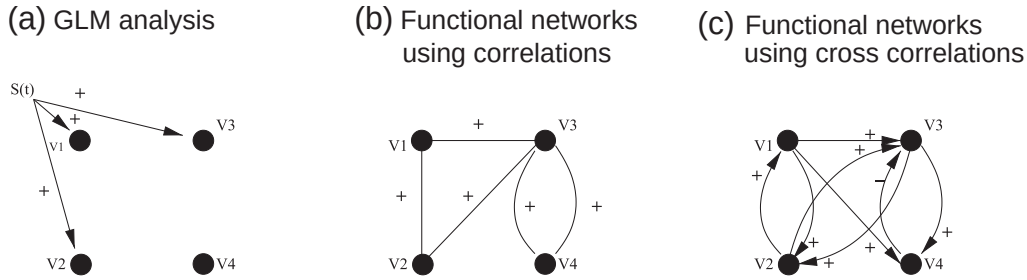



**Fig. 3.** Temporal evolution of the system.




(d) Correlations matrix

| | $v_1$ | $v_2$ | $v_3$ | $v_4$ |
|---|---|---|---|---|
| $v_1$ | 1 | 0.60 | 0.47 | 0.18 |
| $v_2$ | 0.60 | 1 | 0.34 | 0.19 |
| $v_3$ | 0.47 | 0.34 | 1 | 0.25 |
| $v_4$ | 0.18 | 0.19 | 0.25 | 1 |

(e) Lag-1 cross correlations matrix

| | $v_1$ | $v_2$ | $v_3$ | $v_4$ |
|---|---|---|---|---|
| $v_1$ | 0.81 | 0.42 | 0.19 | 0.10 |
| $v_2$ | 0.69 | 0.47 | 0.33 | 0.22 |
| $v_3$ | 0.60 | 0.32 | 0.21 | -0.29 |
| $v_4$ | 0.31 | 0.26 | 0.61 | -0.02 |

**Fig. 4.** Model recovery using different methods.

*regression* from the machine learning literature (Tibshirani, 1996; Garg and Khandekar, 2009; Candès, 2008), recently introduced to brain imaging (Carroll et al., 2009), to solve this problem.

*Sparse regression for model identification*

We solve the problem of model identification using the techniques of sparse regression from machine learning (Tibshirani, 1996; Garg and Khandekar, 2009). Consider a multivariate linear regression model of the form $Y = ZW + e$ where $Y$ is a known $n_1 \times 1$ vector of observations (in our case, the measurements of the evolution of the system), $Z$ is a known $n_1 \times n_2$ regressor matrix (the time-lagged cross-covariance of the data) and $W$ is the unknown model vector of size $n_2 \times 1$ (the coefficients $a_{ij}(\tau)$) to be determined using the observations $Y$ and regressor $Z$. If $n_1 < n_2$ then the system is under-determined and has an infinite number of solutions.

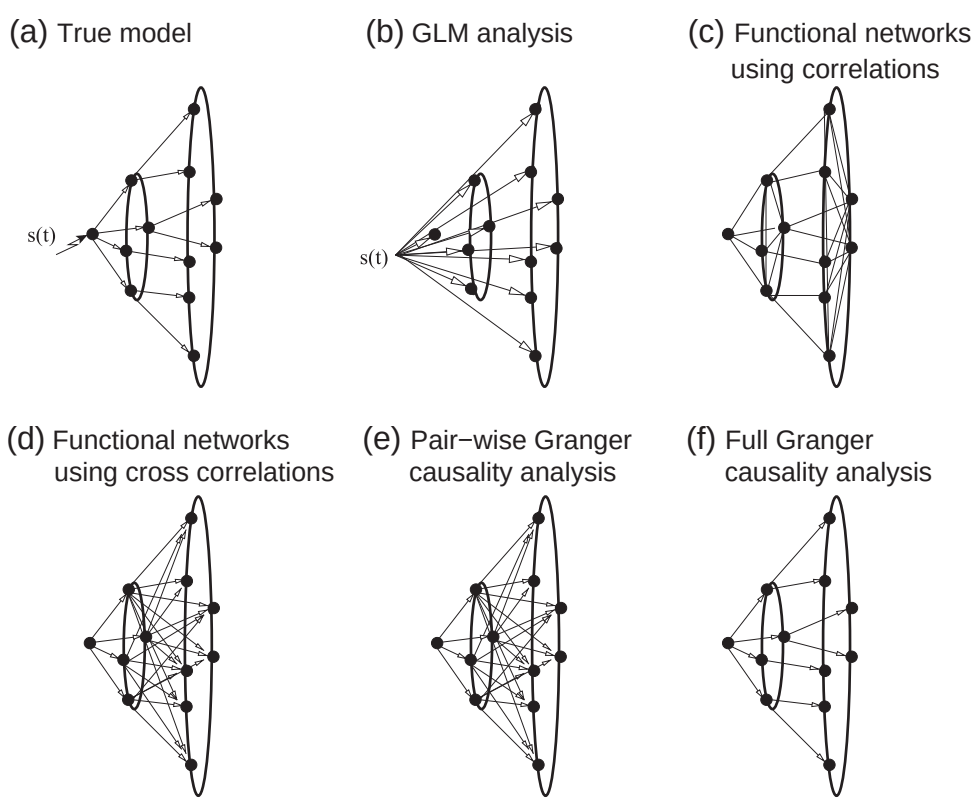



**Fig. 5.** Analysis of a layered network using different methods.

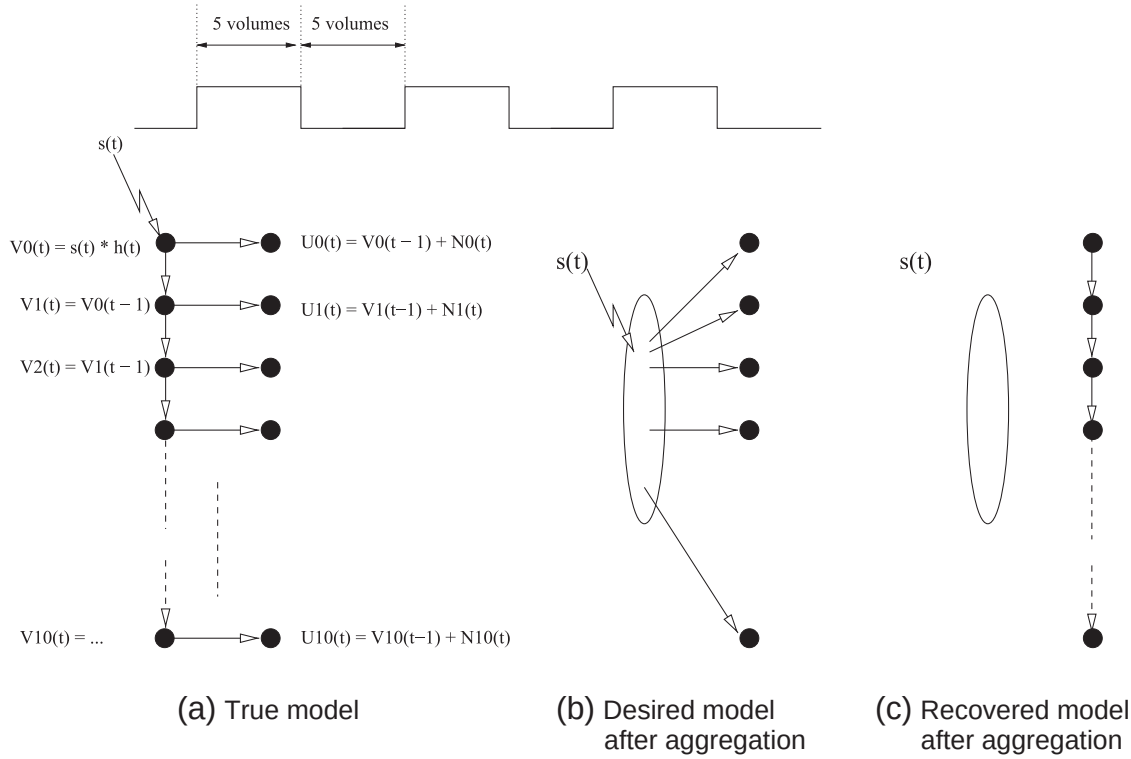



**Fig. 6.** Aggregation of data loses information and leads to an incorrect model recovery.

In the case of fMRI data, the matrices $A(\tau)$ represent the temporal interactions between the neurons contained in pairs of voxels in the brain. Such interactions are expected to be sparse (i.e. not every voxel is expected to interact with every other voxel). Therefore, a small number of non-zero entries are expected in the unknown matrices $A(\tau)$ and hence in the vector $W$.

The problem of sparse regression is to find a solution $W$ to the above system that best matches the observations $Y$ and has a small number (say $s \ll n_1$) of non-zero elements. Finding the sparse solution is NP hard in general (Natarajan, 1995; Neylon, 2006), and therefore computationally intractable. Fortunately, there is a growing body of literature that demonstrates that under a variety of conditions, the problem of sparse regression can be solved using the following $\ell_1$-regularized formulation (Tibshirani, 1996; Candès, 2008):

$$\min_{W} \|Y - ZW\|_2^2 + \lambda \|W\|_1 \tag{5}$$

where $\|v\|_2^2$ represents the square of $\ell_2$ norm (i.e., sum of square of all the components) of the vector $v$ and $\|v\|_1$ represents the $\ell_1$ norm (i.e., sum of the absolute value of all the components) of the vector $v$. The parameter $\lambda$ is called the regularization parameter which is carefully chosen to balance the trade-off between the sparsity of the solution (i.e. the number of non-zero entries) and its match with the observations $Y$.

There are several methods to solve the sparse regression problem (Chen et al., 2001; Candès and Romberg, 2004; Efron et al., 2004). The most notable method is a modification to the least angle regression (LARS) (Efron et al., 2004) that provides a computationally efficient method to solve Eq. (5). We solve the model identification problem by solving the suitably defined program of Eq. (5), using the lasso modification to the LARS (Efron et al., 2004) on the IBM Blue Gene supercomputer (IBM Blue Gene Team, 2008). The reader is referred to (Cecchi et al., 2008a) and (Garg et al., 2009) for more details about the problem formulation and parallelization on Blue Gene.

*Modeling neuronal activity using a multivariate auto-regressive model*

Let $x_i(t)$ in Eq. (2) represent some measure of the aggregate neuronal activity in voxel $i$ at time $t$. Since we are only interested in the temporal dynamics of the system, $x_i(t)$ may be normalized to zero mean for the duration of the experiment. Eq. (2) represents a linear multivariate auto-regressive model of order $k$, capturing aggregate temporal interactions among the neurons of different voxels, where $a_{ij}(\tau)$ is the entry in $i$th row of $j$th column of the $n \times n$ matrix $A(\tau)$, representing the influence of aggregate neuronal activity of voxel $j$ on aggregate neuronal activity of voxel $i$ after $\tau$ time steps; $e(t)$ is an $n \times 1$ vector with entries $e_i(t)$ representing the aggregate unexplained neuronal activity at voxel $i$.

The fMRI BOLD signal is an indirect measure of the neuronal activity. It is believed that the neuronal activity, which is supported by aerobic metabolism which results in the consumption of oxyhemoglobin (which is diamagnetic) and release of deoxyhemoglobin (which is paramagnetic). However, this demand for energy leads to an inflow of blood carrying oxyhemoglobin, that far exceeds the requirements. Thus, the fMRI BOLD signal shows an initial dip corresponding to the consumption of oxyhemoglobin followed by a larger increase in the signal (corresponding to the over-compensated flow of blood carrying oxyhemoglobin) reaching its peak value in 4–5 s, which returns to its baseline level in 18–20 s (Handwerker et al., 2004; Aguirre et al., 1998; Menon and Kim, 1999). This process is characterized by hemodynamic response function (HRF) $r_h(t)$.

The neuro-vascular coupling is usually modeled as a linear time invariant (LTI) system (Friston et al., 1995, 2007). In this case, the aggregate neuronal activity $x_i(t)$ at voxel $i$ is related to the fMRI BOLD response as $\hat{y}_i(t) = r_h(t) * x_i(t)$, where $\hat{y}_i(t)$ represents the BOLD response of voxel $i$ and $*$ represents the convolution operator. Although the LTI assumption is inaccurate (Friston et al., 2000;

Vazquez and Noll, 1998), it has served well for a variety of experimental paradigms and has become a standard assumption in most of the fMRI research literature. Let the vector $Y(t) = \hat{Y}(t) + \eta(t)$, represent the BOLD response obtained from the experiment after suitably processing the measurements, and $\eta(t)$ represent the noise introduced in the measurement and processing steps. We make the simplifying assumption that the hemodynamic response function is the same for every voxel (see Variability in hemodynamic response function and Impact of the variability in the hemodynamic response function for detailed discussions). Convolving Eq. (2) with the hemodynamic response function $r_h$ gives:

$$\hat{Y}(t) = \sum_{\tau=1}^{k} A(\tau)\hat{Y}(t-\tau) + r_h(t) * e(t). \tag{6}$$

Therefore,

$$Y(t) = \sum_{\tau=1}^{k} A(\tau)Y(t-\tau) + r_h(t) * e(t) + \eta(t) - \sum_{\tau=1}^{k} A(\tau)\eta(t-\tau). \tag{7}$$

Let $\mu(t) = r_h(t) * e(t) + \eta(t) - \sum_{\tau=1}^{k} A(\tau)\eta(t-\tau)$. As a first step in the analysis, we make the simplifying assumptions that the vector $\mu(t)$ is stationary, and spatially and temporally uncorrelated for the duration of the experiment. Formally,

$$\text{Exp.}\left(\mu(t_1)^{\top}\mu(t_2)\right) = \begin{cases} \text{diag}(\mu_1, \mu_2, \ldots, \mu_n) & \text{if } t_1 = t_2 \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

where Exp.($X$) represents the expectation of the random variable $X$ and diag($\mu_1$, $\mu_2$,…,$\mu_n$) represents a $n \times n$ diagonal matrix with entries $\mu_1$, $\mu_2$,…,$\mu_n$.

Before proceeding, it is pertinent to discuss some implications of our modeling assumptions. In the first place, let us notice that even though Eq. (2) and those derived from it do not seem to take into account zero-lag correlations, i.e., correlations between $x_i(t)$ and $x_j(t)$ for $i \neq j$, such correlations are implicit in the model. Significant correlations between $x_i(t)$ and $x_j(t)$ may be present even without any zero-lag correlation in the residual term $e(t)$ (see also Fig. 18 in Prediction power maps are highly localized). Secondly, even though the model doesn't seem to explicitly account for instantaneous Granger causality, i.e., explaining $x_i(t)$ by $x_j(t) \forall j \neq i$, such relationships are implicit. This can be seen if we express the auto-regression as:

$$x(t) = A(0)x(t) + \sum_{\tau=1}^{k} A(\tau)x(t-\tau) + e(t)$$

$$x(t) = (I - A(0))^{-1}\left(\sum_{\tau=1}^{k} A(\tau)x(t-\tau) + e(t)\right)$$

where $I$ is the identity matrix. The form of the noise term in the above expression implies that our assumption of a diagonal noise (Eq. (8)) is problematic. It would be possible, in principle, to extend our approach by iteratively estimating the noise matrix using the Lyapunov equation (Gajic and Qureshi, 1995), but that would make for a much more computationally demanding algorithm. One expected consequence of our assumptions is that, due to the limited temporal resolution of the data, the algorithm will be less sensitive to zero-lag correlations that cannot be explained by causal interactions. We will return to this issue when we analyze the result of applying the method to real data.

Stationarity of neural activity is another assumption which may be inaccurate for arbitrary brain processes. However, most of the fMRI experiment designs employ significant amounts of repeatability in the experimental conditions. The stationarity assumption is expected to be valid if the aggregate statistical properties, such as mean inter-stimulus interval in even-based design, block size in block designs, remain unchanged for the duration of the experiment. In such cases, the stationary components of the brain processes are likely to be discovered by the model whereas the non-stationary components are likely to be treated as noise.

With these assumptions, the spatio-temporal dynamics of the BOLD response becomes an auto-regressive process identical to the spatio-temporal dynamics of the aggregate neuronal activity Eq. (2) with $x(t)$ replaced with $Y(t)$ and $e(t)$ replaced with $\mu(t)$. As a result, the model $A(\tau)$ for aggregate neuronal interactions may be inferred by recovering the auto-regressive model for the BOLD response. We solve for the matrices $A(\tau)$ using the lasso regression on the normalized post-processed BOLD response as described briefly in Sparse regression for model identification and, in detail in (Cecchi et al., 2008a) and (Garg et al., 2009). Although the assumption in Eq. (9) on the property of $\mu(t)$ is not optimal, our simulations indicate that the true model can be estimated very well qualitatively, even with these simplifying assumptions. The estimated model parameters $\hat{a}_{ij}(\tau)$ show a small "shrinkage" bias towards the origin due to large auto-correlations introduced by the convolutional kernel $r_h(t)$ (see Recovering the true model).

### *Variability in hemodynamic response function*

A large body of literature shows that the hemodynamic response function (HRF) is not fixed. It varies across different population groups (D'Esposito et al., 2003; Thomason et al., 2005), across different subjects of the same population (Aguirre et al., 1998), across different brain regions of the same subject (Handwerker et al., 2004), across different sessions (Neumann et al., 2003) and across different trials of the same session (Duann et al., 2002). The inter-subject variability has been found to be much more than the variability across different brain regions of the same subject (Handwerker et al., 2004).

Most of the studies demonstrating the variability in HRF have followed the paradigm of the general linear model (GLM), which although not entirely accurate, was the only method to reliably analyze the fMRI data until the dawn of newer techniques based on machine learning (Norman et al., 2006; Mitchell et al., 2008) and connectivity (Li et al., 2009; Eguiluz et al., 2005).

According to the GLM assumption, the aggregate neuronal activity in any brain region is a weighted linear sum of the experimental conditions. Thus, if $s_j(t)$ represents the state of the $j$th experimental condition at time $t$, then the aggregate neuronal activities are modeled as $x_i(t) = \sum_j \beta_{ij} s_j(t)$. The voxel $i$ is said to be "active" in response to the experimental condition $j$ if $\beta_{ij} > 0$ (and de-active if $\beta_{ij} < 0$) at high levels of statistical significance. This assumption worked well for block-based designs. Improved temporal resolution and rapid event-based experimental designs exposed not only the variability of HRF, but also the shortcomings of the linearity assumption (Friston et al., 2000; Vazquez and Noll, 1998; Glover, 1999; Gitelman et al., 2003).

According to a more accurate model (Friston et al., 2000; Buxton et al., 2004), the fMRI BOLD response has two components: (a) the *neurodynamic response function* $r_n(t)$ which maps the experimental condition to the aggregate neuronal activity and (b) the *hemodynamic response function* $r_h(t)$ which maps the neuronal activity to the fMRI BOLD signal. Under the LTI assumptions, the neuronal activity is related to the experimental condition through the neurodynamic response function as

$$x_i(t) = \sum_j \beta_{ij} r_n(t) * s_j(t).$$

The fMRI BOLD signal is modeled as follows

$$\begin{aligned} y_i(t) &= r_h(t) * x_i(t) \\ &= \textstyle\sum_j \beta_{ij} r_h(t) * r_n(t) * s_j(t) \\ &= \textstyle\sum_j \beta_{ij} r_b(t) * s_j(t) \end{aligned}$$

where $r_b(t) = r_n(t) * r_h(t)$, represents the fMRI *BOLD response function* to the experimental conditions.

Until the advent of simultaneous fMRI and neuronal recordings (Lippert et al., 2010; Logothetis, 2008), it was impossible to segregate the neuronal activity from the fMRI BOLD response. Therefore, most of the papers reporting variability in the HRF (Handwerker et al., 2004; Aguirre et al., 1998; D'Esposito et al., 2003; Thomason et al., 2005; Neumann et al., 2003; Duann et al., 2002) implicitly assumed that (as in the general linear model) the *neurodynamic response function* was identity. Thus, the fMRI BOLD response function $r_b$, which was found to be variable, became synonymous with the hemodynamic response function $r_h$. Most of the researchers recognize this fact and acknowledge the possibility that the observed variability in the HRF could have its origin in neuronal activity differences. In other words, most of the literature concludes that the observed inter/intra subject variability in the fMRI BOLD response function $r_b$ could stem either from the variability in neurodynamic response function $r_n$, or from the hemodynamic response function $r_h$ differences, or from a combination of the two.

Studies using simultaneous EEG and fMRI recordings have reported "hemodynamic response functions" that are very different from the canonical HRFs used in the conventional GLM analysis (see Fig. 3 in (Lu et al., 2006). Such studies actually report the *fMRI BOLD response function* which is a convolution of the neurodynamic response function of the EEG signal and the hemodynamic response function of the respective group of neurons, under the LTI assumption. It is unlikely that regional differences in neuro-vascular coupling alone can explain the differences observed in (Lu et al., 2006). There is no established methodology to segregate the observed response into the above two components.

Nevertheless, to examine the impact of HRF variability on model recovery, we carried out simulations experiments. Our results presented in Impact of the variability in the hemodynamic response function, demonstrate that our technique does recover true model satisfactorily as long as the variability in the HRF is less than the time scale of the information flow in the network. Our results on fMRI data, presented later in Evaluation on a sample data, lead us to enunciate the *information flow hypothesis*, which states that a part of the model discovered by the full-brain auto-regressive model indeed corresponds to the information flow in the neural networks (see also Discussion).

### Choice of model order

The choice of model order depends on three factors (a) the time scale of the information flow in the experiment (b) the temporal resolution of the fMRI data being acquired and (c) the intra-subject variability in the HRFs.

If the time scale of the information flow is much smaller than the temporal resolution (TR) of the fMRI data then it will be unreasonable to expect our technique to discover the information flow patterns. If, on the other hand, the time scale of information flow is comparable to (or greater than) the TR of the fMRI data (such as in cognitive response-time tasks (Menon and Kim, 1999; Menon et al., 1998), then our technique does hold promise. If the variability in the HRFs is also small as compared to the time scale of the information flow, then our technique can be applied to discover the information flow pattern. A natural choice of the model order in this case would be the time scale of the information flow divided by the TR of the experiment. For example, in a cognitive task with reaction time varying up to 2 s and a TR of 500 ms a suitable choice of model order will be 4.

However, as it has been suggested by the dynamic core hypothesis, and partially confirmed by studies of the default mode network, we do expect relatively slow ongoing processes to interact with faster task-related activations. Therefore, it may be possible to detect changes in the dynamical structure even when the specific tasks are faster than the TR, in a similar way as they may be measured with a well-crafted block- or even-related design using GLM. In the end, we are proposing a data-driven approach, and as such it requires an unavoidable degree of explorations. The result of applying our method on real data, Evaluation on a sample data set, will shed further light on this issue.

### Simulation methodology

We carried out simulations to evaluate how well the lasso regression is able to estimate the true model coefficients for fMRI data sizes, with the simplifying assumption of Eq. (9) (see supplementary material Section I-A).

Let $S(\alpha)$ and $\hat{S}(\alpha)$ represent the sets of voxel pairs with interaction coefficient more than $\alpha$ in the true and estimated models respectively. Formally, $S(\alpha) = \{(i, j, t) : |a_{ij}(t)| > \alpha\}$ and $\hat{S}(\alpha) = \{(i, j, t) : |\hat{a}_{ij}(t)| > \alpha\}$. The following metrics were used for comparing the true model with estimated model:

*Precision.* It is defined as the ratio of numbers of *true non-zero coefficients* estimated to the *total number of non-zero coefficients estimated*. Formally, precision $p = |S(0) \cap \hat{S}(0)| / |\hat{S}(0)|$.

*Recall.* It is defined as the ratio of the number of *true non-zero coefficients estimated* to the total number of *non-zero coefficients present in the model*. Formally, recall $r = |S(0) \cap \hat{S}(0)| / |S(0)|$.

*Thresholding.* Generally, it is more important to discover the causal relationships that are strong. This can be done by considering only the coefficients with absolute value above a given threshold. The precision and recall may respectively be defined with respect to a threshold $\alpha$ as $p(\alpha) = |S(\alpha) \cap \hat{S}(\alpha)| / |\hat{S}(\alpha)|$, $r(\alpha) = |S(\alpha) \cap \hat{S}(\alpha)| / |S(\alpha)|$.

*Correlations.* We use two measures of the Pearson correlation coefficient to assess the similarity between estimated and actual model parameters. The first measure *C(I)* is defined as the correlation between $a_{ij}(t)$ and $\hat{a}_{ij}(t)$ over the true non-zero coefficient estimates (i.e. over the *true positives* given by $S(\alpha) \cap \hat{S}(\alpha)$). The second measure *C(U)* is defined as the correlation between actual and estimated parameters over *true positives*, *false positives* and *false negatives* (i.e., over the set $S(\alpha) \cup \hat{S}(\alpha)$). Formally, $C(I) = \langle a_{ij}(t), \hat{a}_{ij}(t) \rangle_{(i,j,t) \in S(0) \cap \hat{S}(0)}$ and $C(U) = \langle a_{ij}(t), \hat{a}_{ij}(t) \rangle_{(i,j,t) \in S(0) \cup \hat{S}(0)}$.

### Recovering the true model

The example system of Fig. 1 (with the dynamics of Fig. 2) was solved using sparse regression on the observed data of Fig. 3 as described in Sparse regression for model identification. The true and reconstructed model coefficients for the order-1 auto-regressive model of the system are shown in Fig. 8. It is evident that if we ignore the coefficients with an absolute magnitude of 0.06 or less, the method recovers the qualitative relationships among the nodes exactly (for node $v_1$, the strong auto-correlation introduced by the HRF *h(t)* leads to an estimate of $\hat{a}_{11}(1) = 0.83$). Moreover, the error in estimated values of the model parameters is less than 25%. The system was then driven with a different input. The temporal evolution of the system along with the predicted response is shown in Fig. 7. It may be observed that the predicted system behavior is very close to its true behavior on the new input.

In order to examine if the model coefficients can be recovered for data of dimension comparable to fMRI data, simulations on network with 10,000 nodes and 100,000 connections were carried out, as described in supplementary material, Section I-A.

Fig. 9 shows the scatter plot of the true and estimated MAR coefficients. Note the prominent + sign at the center of the scatter plot. The points on the vertical line ($x = 0$) correspond to the false positives whereas the points on the horizontal line ($y = 0$) correspond to false negatives. The other points correspond to the true positives.

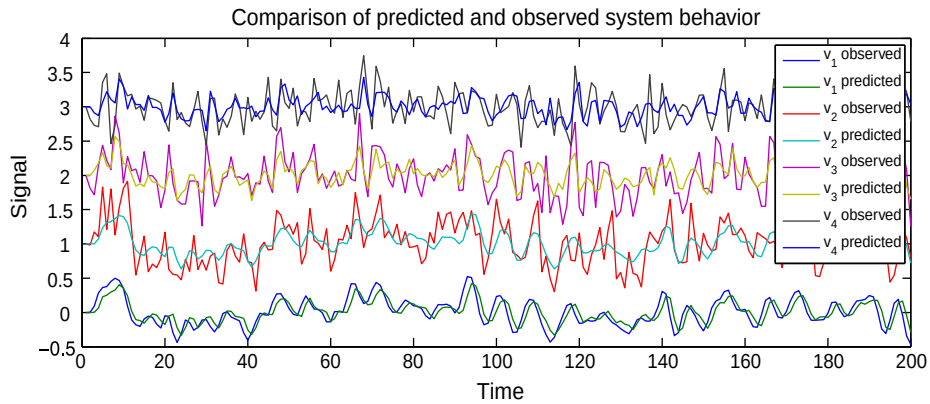



**Fig. 7.** Predicted and actual system behavior.

Note that there is hardly any point in the second and fourth quadrants (i.e. corresponding to $x<0$, $y>0$ and $x>0$, $y<0$). This indicates that most of the true positives are discovered with the correct sign. Also notice the "shrinkage" of the estimated model coefficients toward the $x$-axis (i.e., along the axis $y=0$), which is due to the fact that the lag-1 auto-correlations introduced by convolution with the HRF are assumed to be zero during model identification (see Modeling neuronal activity using a multivariate auto-regressive model, Eq. (9)). Despite various approximations, the $R^2$ of the regression line is 0.7, indicating very good model recovery. The slope of the regression line is 0.5 primarily due to the shrinkage in estimated model coefficients. If the estimated model coefficients are moved away from the $x$-axis (by adding $0.2 * \text{sign}(\hat{a}_{ij}(\tau))$ to $\hat{a}_{ij}(\tau)$), the slope of the regression line becomes 0.98 (compare with the dashed line representing $y=x$).

Fig. 10 shows the precision and recall curves ($p(\alpha)$, $r(\alpha)$), as a function of the lasso regularization penalty $\lambda$ (see Eq. (5)), for different values of $\alpha$. To account for the shrinkage in the estimated model coefficients, $S(\alpha)$ was compared with $\hat{S}(\alpha/2)$ for the computation of precision and recall values. As the penalty $\lambda$ is increased, the lasso regression estimates fewer non-zero coefficients. This improves the precision (i.e., the coefficients estimated to be non-zero are more likely to be non-zero) at the cost of recall (i.e., many non-zero coefficients are estimated to be zero). It is instructive to see that for the regularization parameter $\lambda=1/5$, the precision and recall values are between 0.4 and 0.5 for $\alpha=0$, indicating that at least 40% of the Granger causality relationships estimated by the analysis are indeed correct. In addition, the recall value indicates that the analysis discovers at least 40% of all the Granger causality relationships present in the true model. However, if only the relationships with $|a_{ij}(1)|>0.2$ are to be estimated, the figure indicates improved recovery results (precision and recall values of 0.98 and 0.7 respectively for $\alpha=0.2$ and $1/\lambda=20$). The recovery has been found to be good using other measures as well (supplementary material, Section I-B). Using Monte Carlo methods on surrogate data, it is possible to estimate suitable thresholds for identifying the statistically significant model coefficients (Small, 2005).

*Impact of the variability in the hemodynamic response function*

The fMRI BOLD response function (see Variability in hemodynamic response function) has been found to be variable not only across different subjects but also across different brain areas of the same subject. The mean and standard deviation (intra-subject) in time to peak of the hemodynamic response function have been reported to be 4.7 and 1.1 s respectively (Aguirre et al., 1998). The standard deviation in time to peak (across multiple sessions of the same subject) has been reported to be less than 1 s in most of the cases (Neumann et al., 2003). The minimum, mean and maximum times to peak value of the estimated HRFs of 20 subjects have been reported to be 2.5, 4.0 and 6.5 s respectively (Handwerker et al., 2004). This variability is due to a combination of the variabilities in the neurodynamic and the hemodynamic response functions (see

(a) True model coefficients

| | $v_1$ | $v_2$ | $v_3$ | $v_4$ |
|---|---|---|---|---|
| $v_1$ | 0 | 0 | 0 | 0 |
| $v_2$ | 1.00 | 0 | 0 | 0 |
| $v_3$ | 1.00 | 0 | 0 | -0.50 |
| $v_4$ | 0 | 0 | 0.50 | 0 |

(b) Estimated model coefficients

| | $v_1$ | $v_2$ | $v_3$ | $v_4$ |
|---|---|---|---|---|
| $v_1$ | 0.83 | -0.06 | 0 | 0 |
| $v_2$ | 0.83 | 0.01 | 0 | 0 |
| $v_3$ | 0.77 | 0 | 0 | -0.39 |
| $v_4$ | 0 | 0.02 | 0.46 | 0 |

**Fig. 8.** The estimated model coefficients are within 25% of the true model coefficients except for $\hat{a}_{11}$. The strong auto-correlation induced by the low-pass hemodynamic response function results into the incorrect estimate of $\hat{a}_{11}=0.83$.

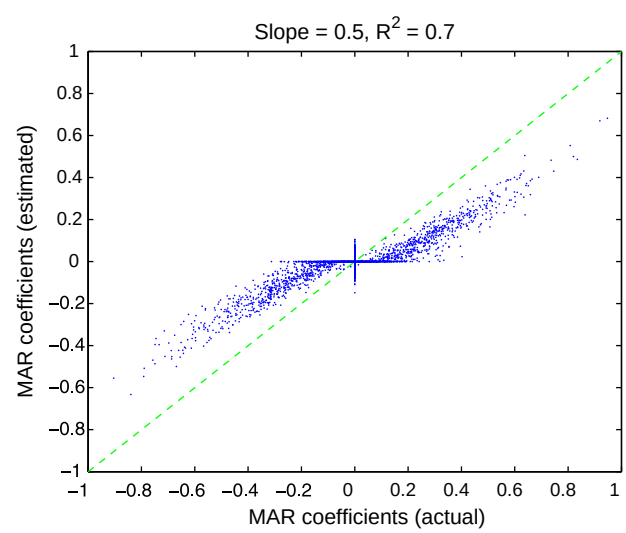



**Fig. 9.** Scatter plot of actual and estimated model parameters.

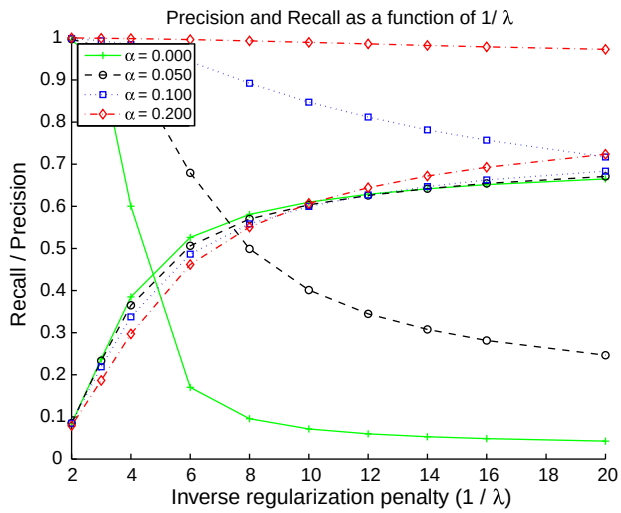



**Fig. 10.** Precision and recall as a function of λ.

Variability in hemodynamic response function). Although there is no method to reliably estimate the intra-subject variability in the HRF (in a given experiment), the above figures may serve as reasonable estimates.

We used the experimental setup of Simulation methodology for our simulations. Instead of convolving the neuronal activity with a fixed HRF, we used a unique HRF for each voxel in the brain. First, we generated a randomly distributed time to peak with a mean value of 5 s and a fixed standard deviation (referred to as the *HRF-variability*) for each of the voxel. These were used to generate the respective HRFs using the two Gamma function model of the SPM (Friston et al., 2007). These HRFs were used for convolution to generate a simulated BOLD signal as in Simulation methodology. The accuracy of model recovery was evaluated using the precision, recall and correlation metrics described earlier.

Figs. 11(a)–(d) show the precision and recall as a function of the regularization parameter λ for HRF-variability of 0.3, 1, 2, and 4 s respectively. It is evident that if the HRF-variability is 0.3 s then the recovery is almost identical to the uniform HRF case. However, if the HRF-variability is close to 1 s as reported in (Aguirre et al., 1998), then the recovery, although worse than the uniform HRF case, is satisfactory. For the value of $1/\lambda = 6$ and $\alpha = 0.05$, the precision and recall values are close to 0.5, indicating that almost 50% of the important links have been recovered by our approach. On the other extreme, if the HRF-variability is 4 s, the model recovery is very poor, with a precision and recall values of close to 0.2 for $\alpha = 0.05$. The precision-recall trade-off and the correlation metrics for these experiments are reported in the supplementary material Section I-C.

We used another metric called the *edge reversal ratio* to study the impact of HRF variability. An unidirectional edge from voxel $i$ to voxel $j$ present in the true model, is said to be reversed if it is absent in the recovered model, but an edge from voxel $j$ to voxel $i$ is present in the recovered model. The edge reversal ratio is defined as the ratio of the number of reversed edges to the number of edges present in the true model. An edge reversal ratio of zero implies that no edges are reversed in recovery whereas the ratio one implies that all the edges present in the true model are recovered in the wrong direction. Analogous to the definitions of precision and recall in Simulation methodology, the edge reversal ratio can be defined at a given level of significance $\alpha$. Figs. 12(a)–(d) show the edge reversal ratios at different levels of significance for HRF-variability of 0.3, 1, 2 and 4 s respectively. For $\alpha = 0.05$, the edge reversal ratio is less than 2% if the HRF-variability is less than 1 s. When the HRF-variability is 2 s, the edge reversal ratio increases to around 5%. When the HRF-variability reaches 4 s, the edge reversal ratio approaches 8%.

Our simulation results indicate that despite several simplifying approximations, the sparse regression methodology is able to recover the full-brain auto-regressive model satisfactorily. At the moment, the intra-subject variability in observed BOLD response cannot be attributed either to the differences in the neuronal responses or to the differences in neuro-vascular coupling. Nevertheless, our simulations show that if the time scale of the information flow is greater than the HRF variability, then more than half of the important links in the model can be correctly recovered.

## Model interpretation

The model matrices $A(\tau)$ encode the spatio-temporal interactions of the aggregate neuronal activity in different voxels. Unlike activation maps that assign a single value to each voxel (representing its activation), the multivariate auto-regressive model assigns many different values to each voxel representing its temporal interaction with other voxels. Our results indicate that these interactions not only encode the designated activity carried out by subjects during the experiment, but also other interaction patterns due to the endogenous activity (such as interactions in the default mode networks). However, it is not trivial to visualize and interpret the model data.

In this section we present two sample metrics derived from the model that make the task of visualization and interpretation easier. We will use these metrics to interpret the results on a simple finger tapping experiment. It must be noted that the metrics presented here are for the purpose of illustration. A variety of other metrics may also be derived using the model.

### *Prediction power*

The *lag-τ prediction power* of a voxel $j$ ($\pi_j(\tau)$), is defined as the sum of the absolute values of the weights of the corresponding column of matrix $A(\tau)$. Formally,

$$\pi_j(\tau) = \sum_{i=1}^{n} |a_{ij}(\tau)|.$$

The *total prediction power* of a voxel $j$ is defined as the sum of its prediction power at all lags, i.e., $\pi_j = \sum_{\tau=1}^{k} \pi_j(\tau)$. Intuitively, the prediction power of a voxel represents the total influence of the voxel in predicting the future values of other voxels.

### *Impulse response dynamics*

Let $\mathcal{I}_s(t)$ represent the impulse function of a voxel $s$ (or a set $s$ of voxels) defined as:

$$\mathcal{I}_s(t) = \begin{cases} \delta_s & \text{if } t = 0 \\ 0 & \text{otherwise} \end{cases}$$

where $\delta_s$ is a $n \times 1$ vector with value 1 at the voxel (or the set) $s$ and zero elsewhere. The *impulse response function* $\mathcal{R}$ of voxel $s$ is defined inductively as:

$$\mathcal{R}(t) = \begin{cases} \mathcal{I}_s(t) & \text{if } t \le 0 \\ \sum_{\tau=1}^{k} A(\tau) R(t-\tau) & \text{for } t > 0. \end{cases}$$

The normalized impulse response function of voxel $s$ is obtained by scaling the vectors $\mathcal{R}(t)$ to unit magnitude in a Euclidean space. The impulse response function represents the evolution of the auto-regressive system defined by the matrices $A(\tau)$ to a perturbation in the activity of a given set $s$ of voxels. To the extent that our hypothesis of linearity is valid, the dynamics of this perturbation will be

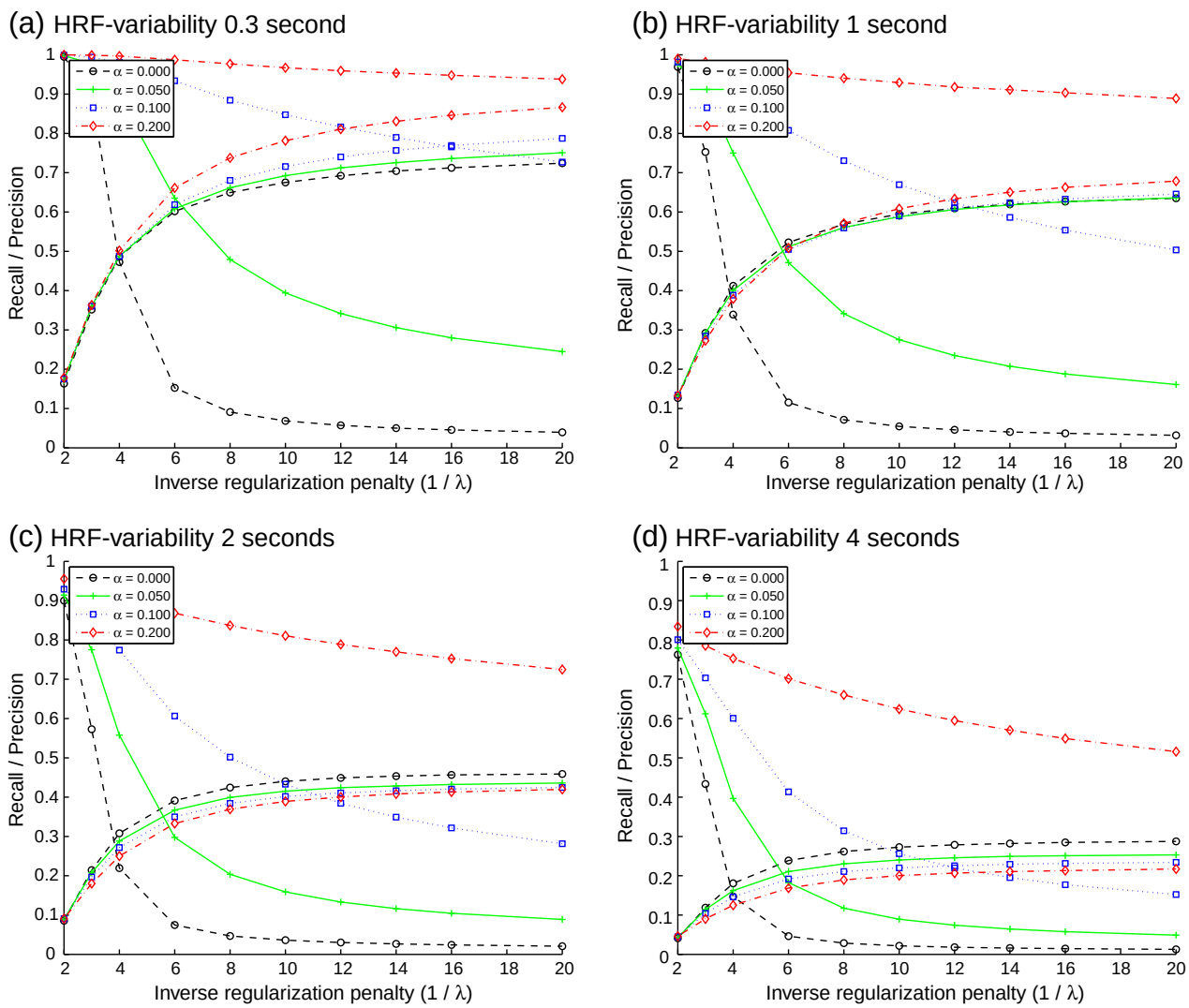



**Fig. 11.** Impact of HRF-variability on model recovery.

*embedded* in the actual, full dynamics of the system. See supplementary material, Section II for illustrative examples.

## Evaluation on a sample data set

### *Experimental protocol and fMRI data pre-processing*

The data used in this paper (and its pre-processing) are identical to (Eguiluz et al., 2005). Six right-handed subjects were scanned three times (a total of 18 sessions) in a block-based self-paced finger tapping paradigm. Each session comprised 20 blocks, each block with 10 volumes of finger tapping followed by 10 volumes of rest (with a TR of 2.5 s). In the first session subjects were instructed verbally to start and stop tapping; in the second session the start and stop cues comprised a small green or red dot on a video screen, and in the third session the cues comprised the entire screen turning green or red.

The fMRI data was acquired using a Siemens-Trio 3.0 T imaging system using a birdcage radio-frequency head coil. Blood oxygenation level-dependent single-shot echo-planar T2-weighted imaging was obtained using a scan repeat time of 2500 ms, echo time of 30 ms, flip angle of 90°, and field of 256 mm. The data were preprocessed for slice timing correction, motion correction and spatial smoothing using the FSL package (FSL Release 3.3, 2006).

### *Choice of regularization parameter using limited cross validation*

Ideally, the regularization parameter λ must be chosen through cross validation. A full cross validation could not be carried out for the eighteen sessions studied, primarily due to the excessive computational requirements. However, we performed a simplified version of cross validation described as follows.

Since the experimental protocol and the data attributes (voxel size, TR and number of volumes of data acquired) were identical for all the sessions the optimal values of λ for each of the sessions are expected to be close to each other. So we used the same value of λ for all the sessions. To determine a suitable choice of λ, we carried out cross-validation for a single session.

The data for the session were split into a training set comprising the first 350 volumes and a test set comprising the last 50 volumes. Using the simulation results, a suitable range of λ was selected. Now a separate model was learned for the session for 16 different values of λ in the range. The performance of these models was evaluated by measuring the mean prediction accuracy on the test data. The λ-value of the model giving the best prediction accuracy was used for all the sessions.

Fig. 13 shows the impact of the regularization parameter λ on the prediction power of the voxels for one of the sessions. The x-axis in the plot represents the prediction power of voxels obtained using the parameter value $\lambda = 1/3.3$. The y-axis represents the prediction power

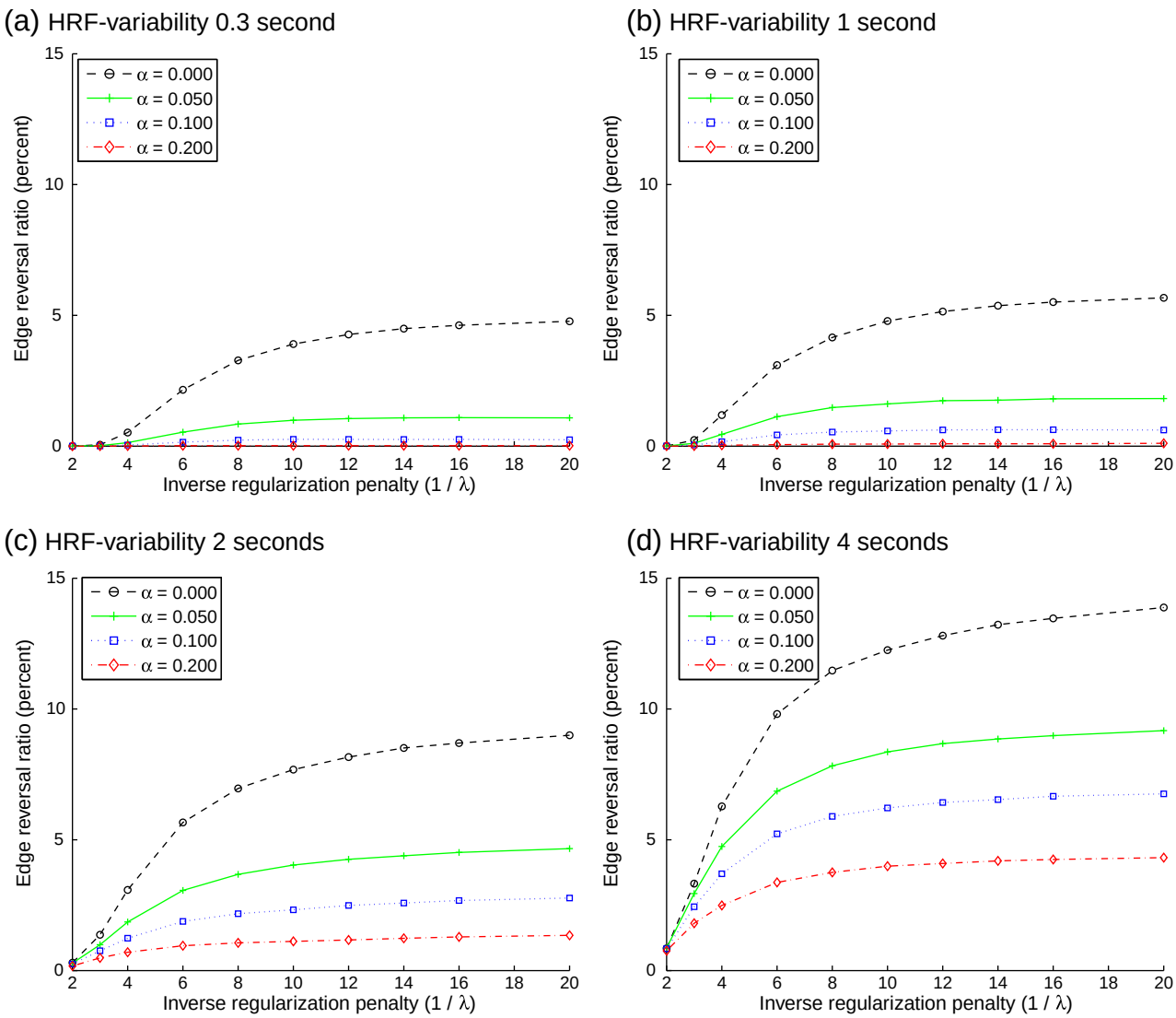



**Fig. 12.** Impact of HRF-variability on the edge reversal ratio.

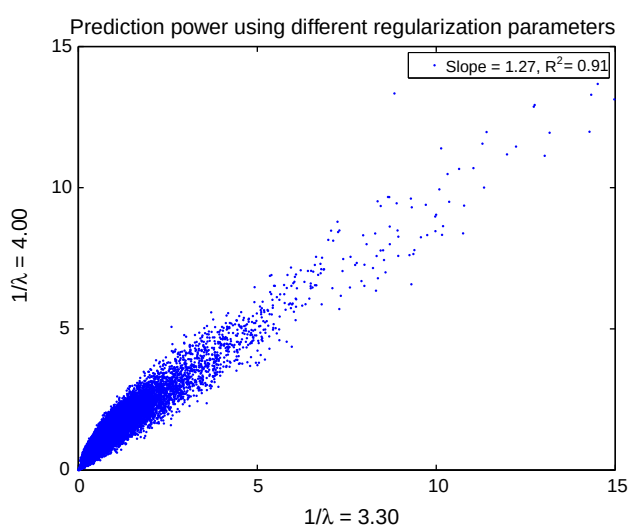



**Fig. 13.** Scatter plot showing the prediction power for two different values of the regularization parameter λ.

of the corresponding voxels obtained using $\lambda = 1/4$. The slope of the regression line is 1.27 indicating that the smaller regularization penalty λ results in larger model weights and therefore higher prediction power. However, the choice of regularization penalty does not fundamentally alter the results. The voxels with high prediction power continue to have high prediction power even with higher regularization penalty.

*Characterization of prediction power maps*

*Comparison of prediction power maps with z-statistic and link-density maps*

We first carry out a qualitative comparison of the maps derived using three different approaches — the prediction power maps as outlined above, the z-statistic maps derived using GLM analysis and link-density maps (also called the hub maps) obtained by suitably thresholding the correlation matrix as defined in (Eguiluz et al., 2005). Though all of these maps are derived from the same fMRI data, they summarize the activity in different ways.

Fig. 14 shows the activation map of one of the sessions using the standard GLM analysis. The map shows a very large region in the left motor area active in response to the experimental condition of finger tapping. Figs. 15 and 16 respectively show the link-density map and

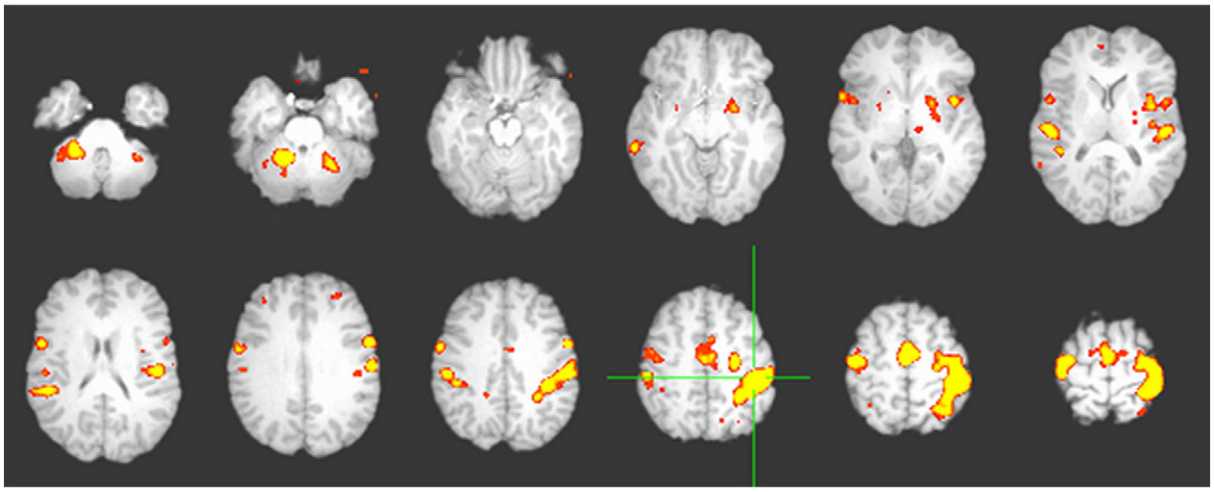


**Fig. 14.** The GLM z-statistic map of one subject representing active regions.

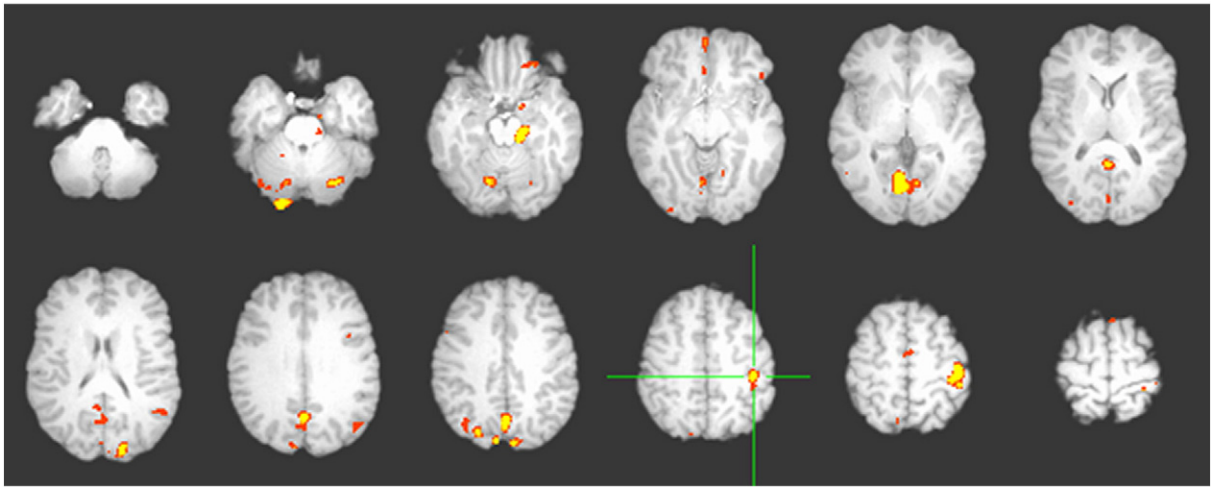


**Fig. 15.** The link-density map of one subject representing regions of high functional connectivity.

the prediction power map of the same session. Clearly, the left motor area has high connectivity in the link density map. Also evident is a small cluster in the same region with high prediction power (see the center of the cross-hair). Although these maps have some intersecting clusters, their general characteristics are very different. The GLM map has a small number of big clusters, mostly concentrated in regions

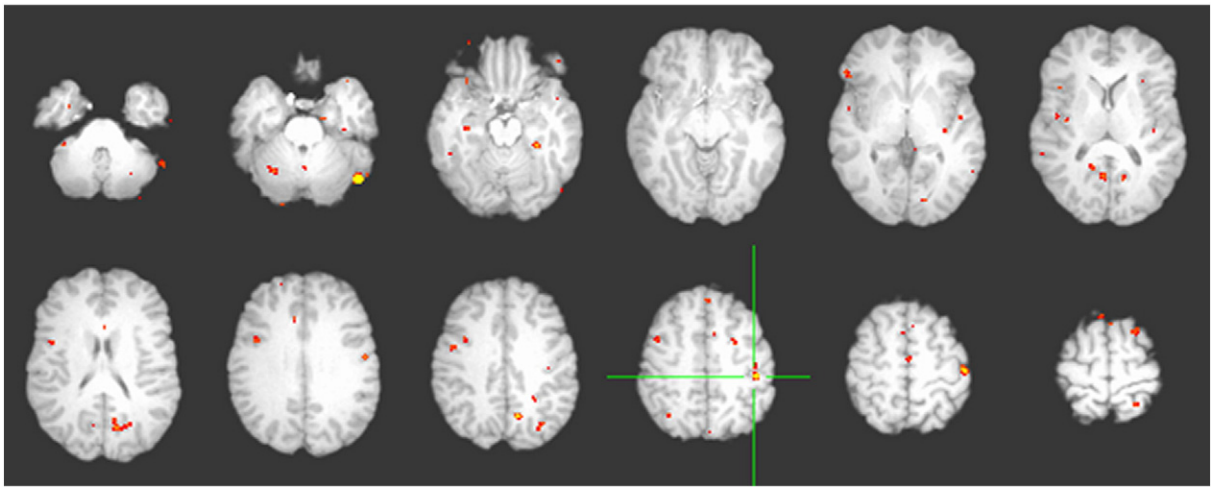


**Fig. 16.** The map of one subject representing regions of high prediction power.

**Table 1**
Comparison of the three maps using the cluster cover metric.

| | Covered map | | | |
|---|---|---|---|---|
| | Cluster cover mean (±std.) | GLM | Link-density | Prediction power |
| Covering map | GLM | 100(±0) | 62(±24) | 35(±14) |
| | Link-density | 88(±8) | 100(±0) | 65(±7) |
| | Prediction power | 91(±6) | 95(±3) | 100(±0) |

related to the task. In contrast, the prediction power map has a large number of very small clusters distributed throughout the brain.

In order to assess the sizes of overlapping clusters in these maps, we use the weighted cluster cover metric as described in supplementary material Section III-B. Table 1 shows the mean (and standard deviation over the eighteen sessions studied) cluster cover of the three maps using the top 5% of voxels. Notable is the fact that 91% of voxels found active by GLM analysis are contained in clusters that intersect with clusters of high prediction power. On the other hand, only 35% of voxels with high prediction power belong to the clusters that intersect with GLM-based clusters of active voxels. This suggests that in addition to encoding the task related activity, the prediction power maps may contain more information about the brain function that is not accessible through the GLM analysis.

*Prediction power maps are highly localized*

Unlike GLM activation maps where activity is typically spread over a large number of voxels in a cluster, the prediction power maps have very small clusters (in many cases, comprising just a single voxel) of high prediction power. Fig. 17 shows the prediction power maps of two sessions each of two subjects (aligned to the MNI atlas). From the magnified portions of these maps, it is evident that (a) despite small sizes, the clusters in the two sessions of the same subject remain aligned, indicating that these maps are highly consistent over different runs of the same subject and (b) due to small cluster sizes, functional differences between the brains of different subjects may have become more apparent in the prediction power maps.

Fig. 18 suggests one possible explanation of this phenomenon. Here, an exogenous signal $S(t)$ drives the response of voxel $v_0$ as $v_0(t) = S(t) * h(t)$ where $h(t)$ is the HRF of voxel $v_0$. The voxel $v_0$ in turn, drives other voxels as follows $v_i(t) = e^{-r_i^2/\sigma^2} v_0(t-1) + \eta_i(t)$, where $r_i$ is the Euclidean distance between voxel $v_i$ and voxel $v_0$. This network was simulated to generate a realization (data) of suitable size which was then subjected to the three analysis methods.

As shown in the figure, the GLM analysis shows a big cluster of activation as a response to the signal $S(t)$ (primarily due to the large lag-1 auto-correlations induced by the HRF $h(t)$). Similarly the hub map shows high connectivity between voxels close of $v_0$ (because of confounding effect of the influence of voxel $v_0$ in its vicinity). The full-brain auto-regressive modeling correctly shows only one voxel with high prediction power. Thus, the full-brain auto-regressive model may be able to localize the seed of functional activity much more accurately than the GLM or other methods. The rest of this section formally characterizes the above properties of the prediction power maps.

*The prediction power distribution is heavy tailed*

Fig. 19 shows the probability distributions of the prediction power maps along with GLM and link-density maps for the fMRI data described in Experimental protocol and fMRI data pre-processing. The GLM maps show an exponentially decaying tail whereas the link-density maps and the prediction power maps are heavy tailed. Moreover, the asymptotic decay of the prediction power seems to be slower than that of the link-density maps. This indicates that there are very few voxels that have extremely high prediction power.

*Comparison of the cluster sizes across multiple methods*

In Fig. 20, we compare the distribution of cluster sizes for the different analysis techniques described in this paper. We first binarize each original 3D image by setting the top 5% of the voxels in this image to 1, and the rest to 0. We perform 3D connected component analysis to identify the 3D clusters, and compute their sizes. For each given analysis mode, the size distributions are aggregated over all subjects and for three conditions per subject. The aggregate cluster size distribution is shown in Fig. 20(a) for prediction power maps, in Fig. 20(b) for the GLM z-statistic maps, and in Fig. 20(c) for the link-density maps.

Note that the $x$-axis as well as the $y$-axis is in log scale in these histograms. The prediction power maps have more than one thousand clusters of size one to two voxels, whereas GLM maps have only one hundred and the hub maps have about four hundred such clusters. On the other extreme, GLM has more than ten clusters of size more than 512 voxels, whereas the largest cluster size in the prediction power maps is less than 128 voxels. Fig. 21 shows the number of clusters of the three maps for all the sessions of the experiment. The number of clusters in each of the sessions is fairly consistent. The prediction power maps have an order of magnitude more number of clusters per session than the GLM-based activation maps.

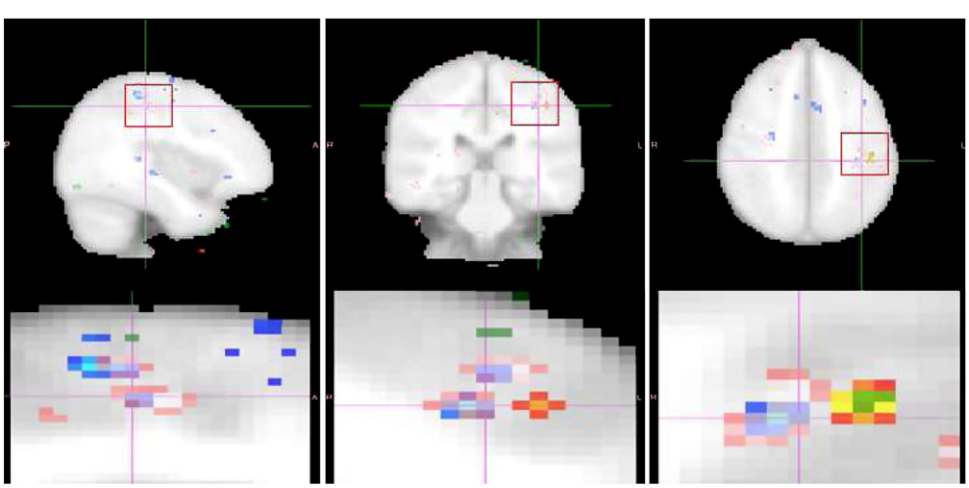


**Fig. 17.** Prediction power maps of two sessions of two subjects. The figures at the bottom are magnified views of the respective boxes on the top figure. Color code: Subject 1 — pink/blue and Subject 2 — green/red-yellow.

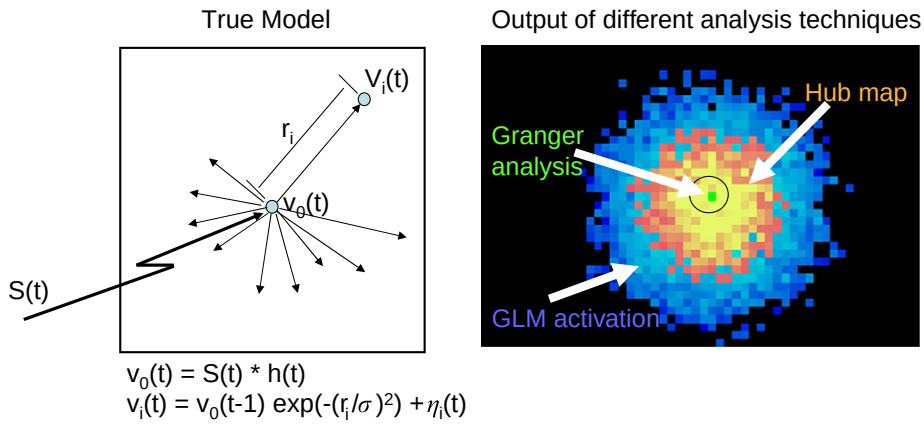



**Fig. 18.** The activity of the voxel in the center ($v_0(t)$) is being driven by an exogenous input $S(t)$. The voxel $v_0$ is the driving rest of the voxels as shown in the figure. The GLM map (shown in blue on the right) shows activity in a large cluster around the voxel $v_0$. The link-density map (shown in red/yellow) indicates high connectivity of voxels close to $v_0$. The full-brain auto-regressive model indicates significant influence only from $v_0$, i.e., it recovers the true model.

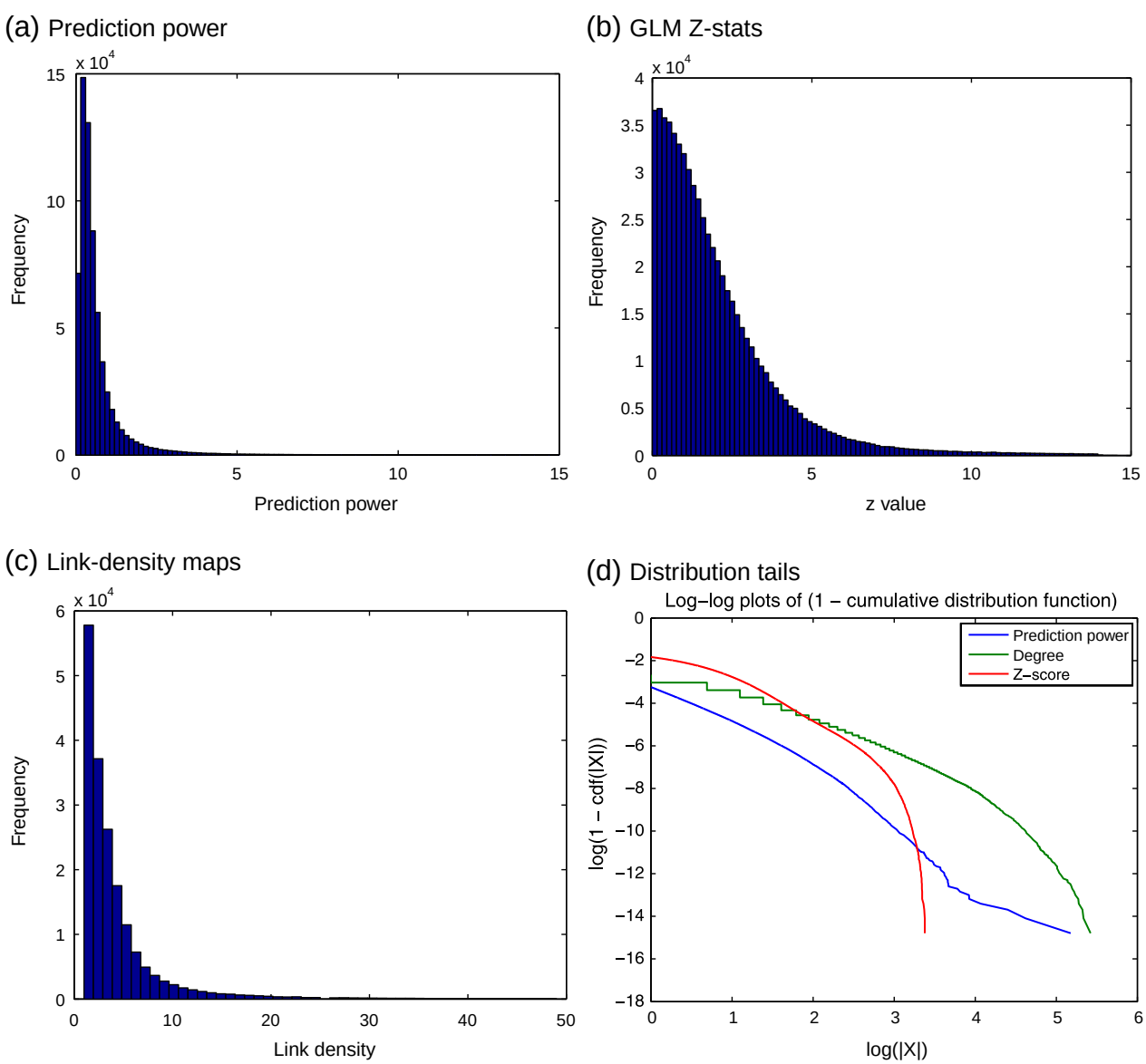



**Fig. 19.** Probability distributions for different types of maps.

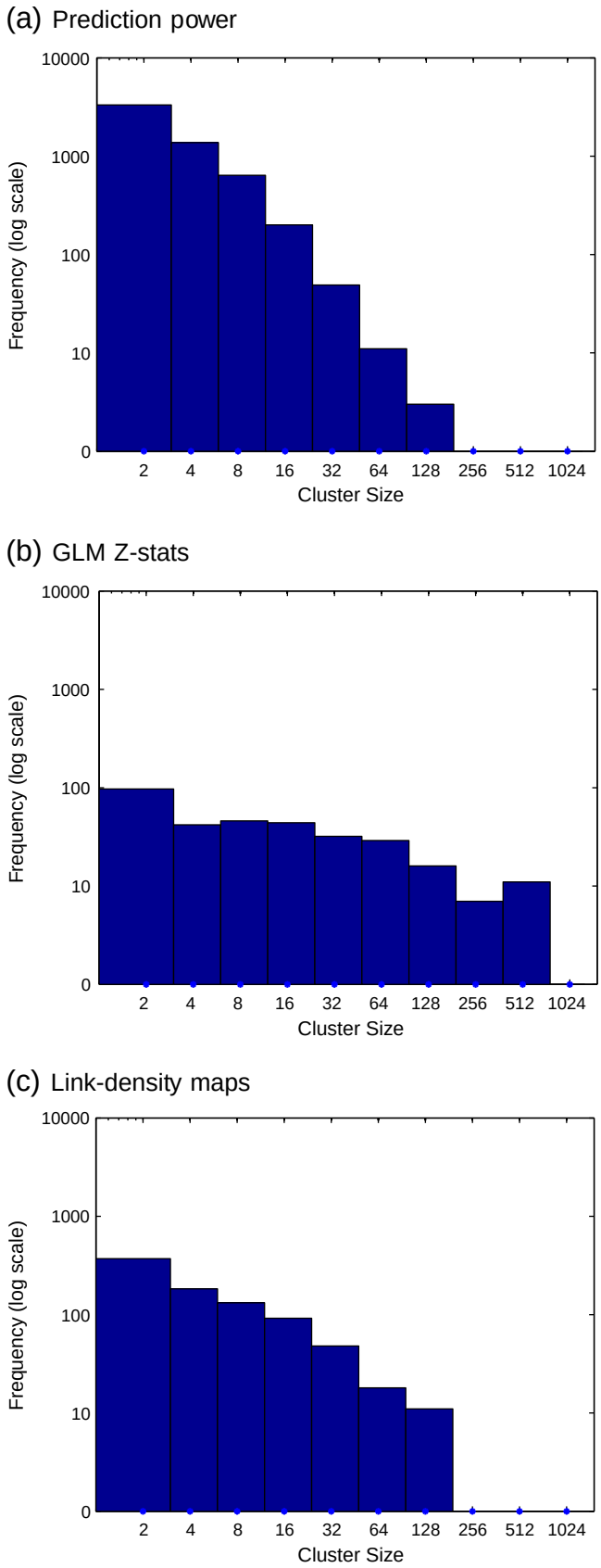



**Fig. 20.** This figure shows the distribution of cluster sizes generated by the following analysis techniques. The unit of measurement is a voxel, where the voxel size is $3.475 \times 3.475 \times 3$ mm. (a) Full-brain auto-regressive modeling. The mean cluster size is 4.59 and the standard deviation is 5.58. (b) GLM method. The mean cluster size is 42.95 and the standard deviation is 94.48. (c) Link-density analysis. The mean cluster size is 10.48 and the standard deviation is 18.92.

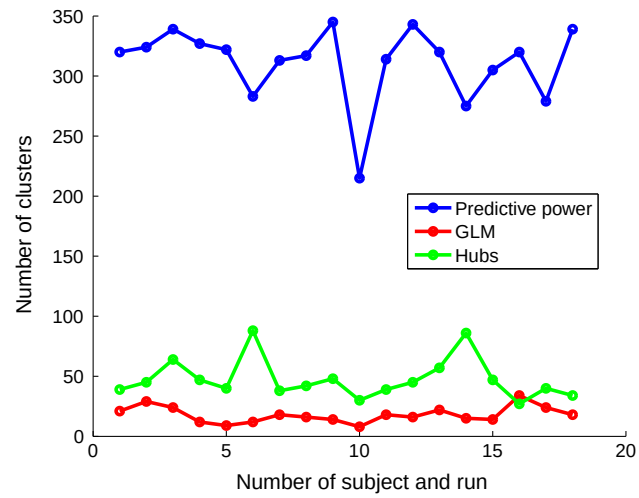



**Fig. 21.** Number of spatially contiguous clusters. The mean number of clusters and the standard deviation are 311.1 and 31.57 respectively for the prediction power method, 18 and 6.76 respectively for the GLM method, and 47.56 and 16.78 respectively for the hub analysis method.

The plots in Figs. 20 and 21 formally characterize the fundamental differences between the maps. The GLM activation maps tend to have small number of large clusters whereas the prediction power maps have large number of small clusters. This illustrates that the full-brain auto-regressive modeling identifies more spatially localized regions of activity than the GLM analysis, resulting in a spatially sparse interpretation of the data. In the GLM method, nearby voxels in a cluster are grouped together based on their shared response to the experimental protocol. However the full-brain auto-regressive modeling precisely identifies only those few voxels that are able to explain the activity of other voxels.

*Intra and inter-subject consistency*

Figs. 22 and 23 respectively show the intra-subject and inter-subject consistency of the maps produced by three methods, using the cluster cover metric. For intra-subject comparisons, the mean (symmetric) cluster cover was computed for each subject and for each pair of sessions

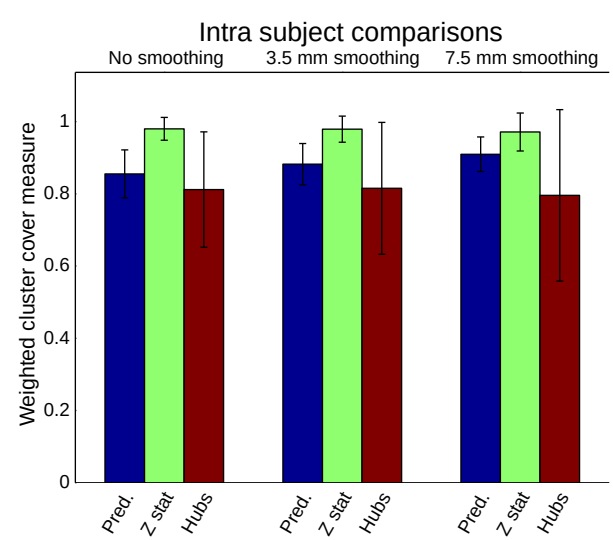



**Fig. 22.** Variability of the cluster cover measure as a function of analysis method and smoothing applied. Comparisons were performed within different sessions for the same subject (intra-subject).

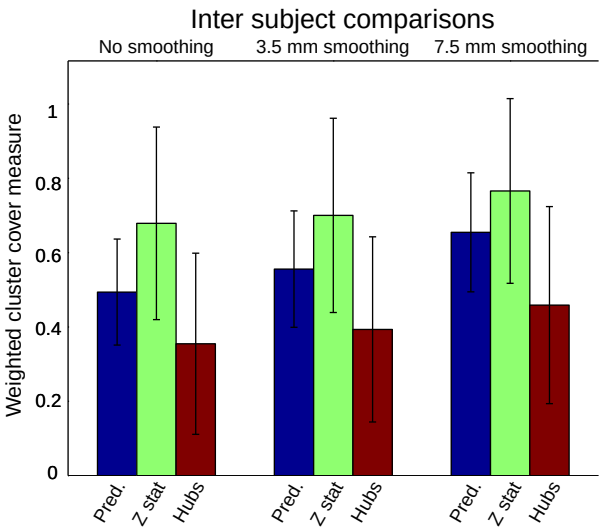



**Fig. 23.** Variability of the cluster cover measure as a function of analysis method and smoothing applied. Comparisons were performed between different subjects (inter-subject).

for the subject, using the top five percent of voxels. The average and standard deviations of the mean cluster cover are reported in Fig. 22 for the three types of maps considered. The mean intra-subject cluster cover for GLM z-statistics map was found to be 0.98 with a standard deviation of 0.03, as expected. The mean cluster cover for the prediction power maps was 0.86($\pm$0.07). Considering the fact that prediction power maps had an average of 311 clusters in each session (see Fig. 21), a cluster cover of 0.86 indicates that an average of 267 clusters was consistent between any two pair of sessions. Due to their small sizes (4.6 voxels on the average), a small amount of mis-registration can make otherwise intersecting clusters of prediction power maps to be disjoint. To account for this, cluster comparison was also done after applying spatial smoothing using spherical filters of different sizes. With spatial smoothing of 3.5 mm and 7.5 mm, the mean intra-subject cluster cover of prediction power maps increased to 0.88 and 0.90 respectively.

For inter-subject comparisons, the same session for every pair of subjects was compared using the cluster coverage metric. The mean inter-subject cover is smaller, as expected. The mean inter-subject cluster cover for z-statistic maps and prediction power maps was found to be 0.68($\pm$0.26) and 0.49($\pm$0.14) respectively. Spatial smoothing improves the mean cluster cover for all the maps. With 7.5 mm spherical spatial smoothing, the mean cluster cover for z-statistic maps and prediction power maps became 0.77($\pm$0.25) and 0.65($\pm$0.16) respectively.

*Cluster analysis*

We will next review the findings of our prediction power cluster analysis, and compare them with the patterns obtained with the general linear model to understand the extent of complementarity and redundancy between these two approaches.

The finger-tapping task has been widely studied in the imaging literature; the activation of a fronto-parietal-cerebellar network of sensory-motor areas involved in the execution, coordination and monitoring of movement has been consistently identified (Jäncke et al., 2000; Aoki et al., 2005). In addition, it might be possible to detect activity in further areas related to the particulars of the task design. Given that in our case subjects performed self-paced tapping with short auditory or visual start and stop cues, it is reasonable to expect the involvement of areas related to executive function, working memory, and auditory and visual processing, as well as sub-cortical regions, specifically the thalamus and basal ganglia (Riecker et al., 2003).

*GLM clusters*

The distribution of activation clusters identified by the GLM analysis confirms the above assumptions, as shown in Fig. 24. Positive activation includes clusters in the cerebellum, contra-lateral thalamus, putamen, primary somatosensory cortex (BA 2/3), primary motor cortex (BA 4), supplementary motor area (BA 6), as well as medial superior parietal cortex (BA 40), bilateral superior temporal gyrus (BA 22) and ipsi-lateral insula. The sensorimotor areas are precisely those expected for this task. The activation of BA 22 may also be expected, given that for the purpose of the present publication we lumped together trials with visual and auditory cues; although it is beyond the scope of the present study, it is interesting to note the specific lateralization of this activation.

The corresponding GLM de-activations, depicted in Fig. 24, are also consistent with previous findings, as they include, prominently, clusters in the ipsi-lateral somato-sensory cortex. An ipsi-lateral cluster of deactivation is observed in the temporal gyrus, which however appears to be centered in posterior part of the insula. The de-activations that are associated with the default mode system are represented by bilateral clusters in the posterior cingulate cortex (PCC, BA 31), and occipital cortex, specifically the cuneus (BA 18/19). It is important to note that the clusters in the PCC lie on a superior and anterior sliver of the clusters identified in previous studies (Greicius et al., 2003, 2004). We can consider two reasons to understand this: firstly, the finger-tapping task has a minimal attentional or cognitive load, and therefore the interference of the task with the ongoing default activity is expected to be proportionally small. Secondly, the interaction between the task and the ongoing activity may be locked temporally in a non-trivial manner; moreover, there might be no interaction at all, i.e. to the spatio-temporal resolution provided by fMRI, the task could appear for all practical purposes to be independent of any hidden variable determined by “other” events taking place. This second hypothesis is precisely the class of questions that we can ask with our analytic approach; we will return to it in the next sections.

*Prediction power clusters*

Following the procedure described in Section IV (Binomial testing methodology), supplementary material, we identified a number of clusters with sufficient group statistical significance. Shown in Fig. 24, these include clusters in the ipsi-lateral medial supplementary motor area (BA 6), contra-lateral superior temporal gyrus (BA 22), bilateral inferior frontal gyrus (BA 44/45), and the brain stem/red nucleus of the midbrain. While the first two clusters are congruent with GLM activations, the last two are not. The BA 44/45 in the inferior frontal gyrus is the site of Broca's area in the dominant hemisphere, but we observe a bilateral cluster here. Even though there is a stronger signal in the right hemisphere, the cluster in the left hemisphere is statistically significant. Interestingly, beyond the traditional association of Broca's area with language, increasing evidence points to a role in imagery of movements related to hand movements and speech-associated gestures (Binkofski et al., 2000; Skipper et al., 2007). On the other hand, the red nucleus has been identified as part of the network of activations during preparation and execution of finger-tapping tasks (Cunnington et al., 2001). Given that the report in (Cunnington et al., 2001) is event-related, we may conclude that the block-based approach we used for the GLM analysis may be a confounding factor in the detection of activity in this area. Be it as it may, it is interesting that the auto-regressive analysis can identify the brain stem/red nucleus as an area that consistently predicts the activity of other parts of the brain (Fig. 24).

The prediction power analysis also includes two clusters that capture noise sources: one is a susceptibility artifact located on the boundary between the corpus callosum and the ventricle near the frontal cingulate, the other on the boundary of the lower brain stem. This is similar to the type of results obtained with other data-driven approaches, such as ICA (McKeown et al., 1998).

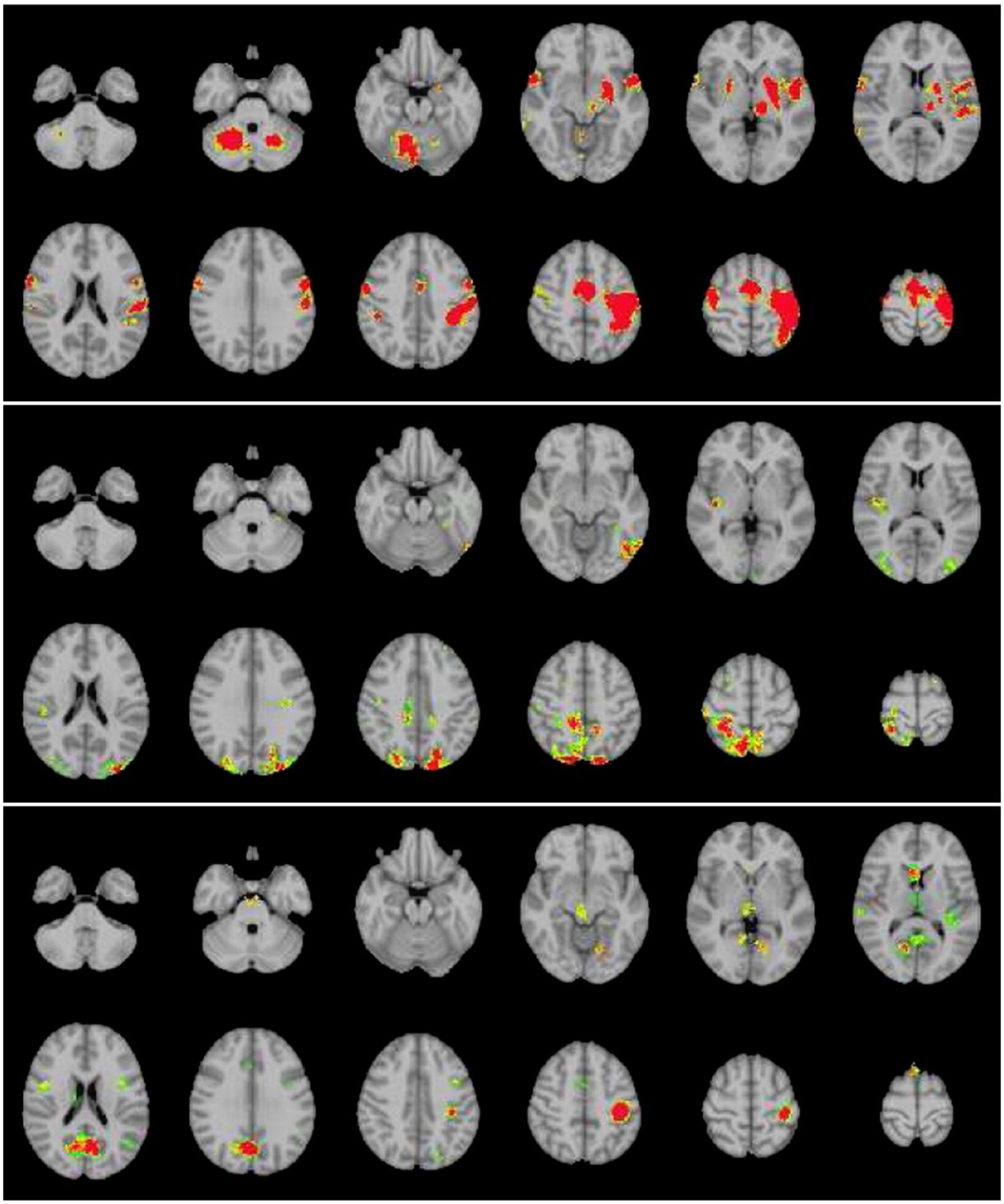


**Fig. 24.** GLM and prediction power statistics. The color maps are proportional to $log(1/p)$, obtained using the statistical model described in Methods, and corrected for multiple comparisons at 1%. The top and the middle panels correspond to GLM activation and deactivation statistics, respectively. The bottom panel represents the cluster of significant Prediction Power statistics. While there are some obvious overlaps between prediction power and GLM, prominently the parietal region, the prediction power clusters are sparser as they represent areas that can consistently predict the activity of other areas.

However, the top two identified regions (in terms of the number of voxels) correspond to a very large bilaterally extended cluster in the precuneus ([+12 −67 +23], 3000 voxels), and a cluster in contralateral superior parietal cortex, BA 40 ([−40 −23 +47], 1300 voxels). The latter is congruent, albeit spatially more constrained, with the activations observed with the GLM analysis. Although the identification of this area should not be surprising, it is still interesting that the analytic method, and the information it provides, are quite different from that of GLM: it is based on the ability of the voxels to consistently predict the future of other voxels. A simple interpretation of this observation could be provided by the presence of local correlations in the data, either stemming from function or artifacts; while this may

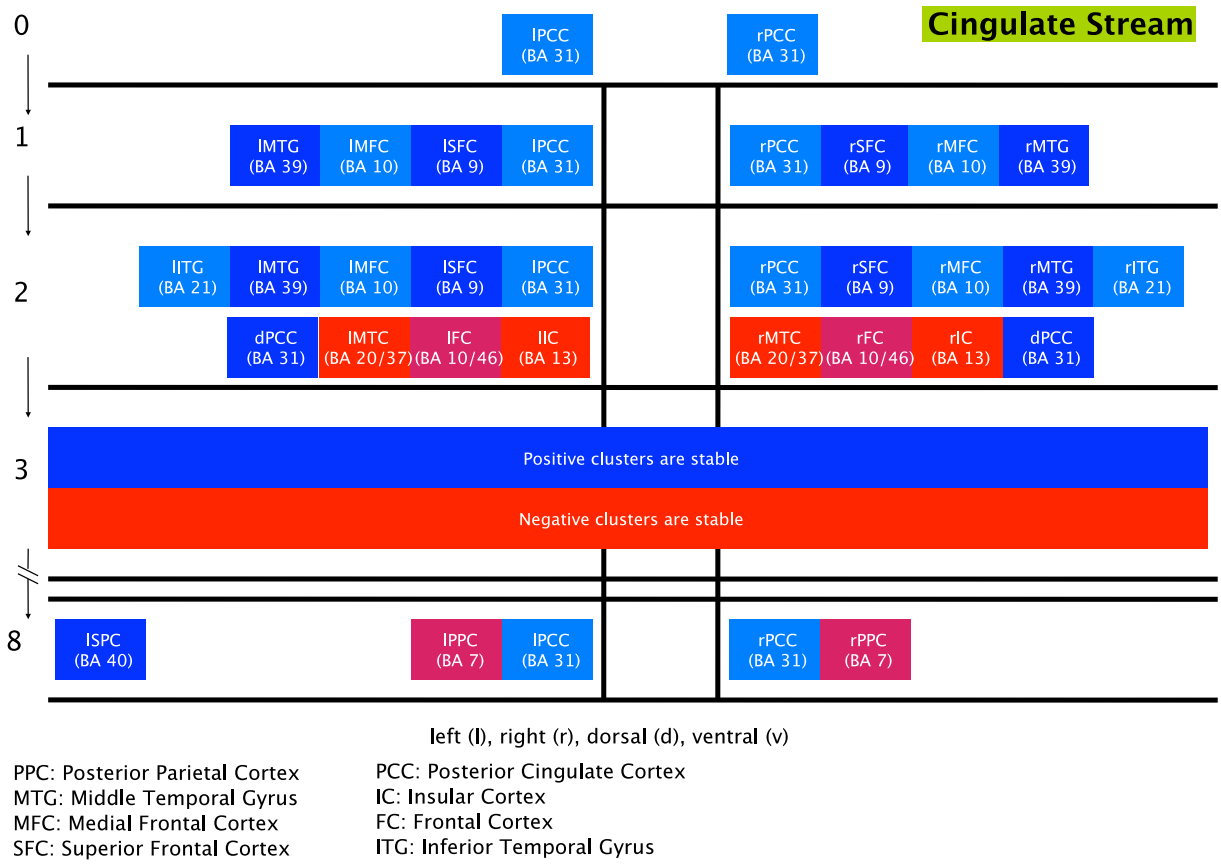



**Fig. 25.** Cingulate stream. Similar to the parietal stream, this chart summarizes the temporal structure of the activity originated in the posterior cingulate/precuneus cluster. This stream also has rich dynamics, but unlike the parietal stream it is to a large extent bilaterally symmetric, and it mostly spans cortical areas. It is also evident that there are areas of overlap between the two streams; in particular, the positive parietal stream overlaps with the negative cingulate stream in the insular cortex.

not be completely disproved, we will show evidence that our method is actually picking up non-trivial functional regularities.

The precuneus cluster, on the other hand, is significantly different from, and in fact non-overlapping with either GLM activations and de-activations. Moreover, it is indeed the largest observable cluster (by a factor of 3). Based on the default mode literature, this observation is not surprising; however, we will show that by fully exploiting the graphical structure of the auto-regressive model, it is possible to extract information about the organizing role of this area in terms of the spatio-temporal dynamics of the brain.

*Impulse response analysis*

As mentioned above, the identification of clusters with high predictive power only provides a partial perspective of the information captured by our analysis. The impulse response analysis proposed in the Methods section is intended to reveal part of this information, by tracing the propagation of activity through the graph defined by the auto-regressive model. To illustrate this approach, we computed the normalized impulse response function (IRF) generated by seeds in the first and second largest clusters of the prediction power maps, whose anatomical representation we will term cingulate/precuneus and parietal *streams*, respectively. The positive and negative evolutions of the IRF applied to both seeds are depicted for the first 9 time steps through coronal, sagittal and axial views in Fig. 27. Before diving into a detailed description of these streams, it is important to point out that the auto-regressive model is linear, which means that the solution to the superposition of two inputs is equivalent to the superposition of each of the solutions. Therefore, the four streams represented in the figure can be considered different components of the dynamics. Moreover, it would be possible to completely characterize the full dynamics (as captured in a linear model) by representing the streams corresponding to all of the prediction power clusters, i.e. the "hubs" of the graph.

*Precuneus/posterior cingulate and parietal streams*

The seed for the cingulate/precuneus stream corresponds to the largest cluster, and is situated, bilaterally and medially, in the posterior cingulate cortex (BA 31), an area that is considered to be part of the precuneus. The first time step of the IRF leads to positive activations growing in the precuneus (BA 31), medial frontal cortex (BA 9/10) and middle temporal gyrus (BA 39), all bilaterally. The next time step adds positive clusters in the inferior temporal gyrus (BA 21), and dorsal posterior cingulate cortex (BA 31), whereas negative clusters develop in the middle temporal gyrus (BA 20/37), frontal cortex (BA 10/46), and insular cortex (BA 13), bilaterally and right supplementary motor area (BA 6) (see Fig. 25).

The next time steps show stabilization of the positive and negative clusters, and later a reduction in most of them, with the exception of the initial posterior cingulate area, as well as negative activation of the posterior parietal area (BA 7) and middle frontal cortex (BA 9/10). Interestingly, by time step 7 positive activation is observed growing in the same left superior parietal cortical area that is the seed for the parietal-motor stream.

The seed cluster for the parietal stream is located in the left superior parietal cortex, corresponding to BA 40. After one time step of the IRF, the activation evolves to cover also the supplementary

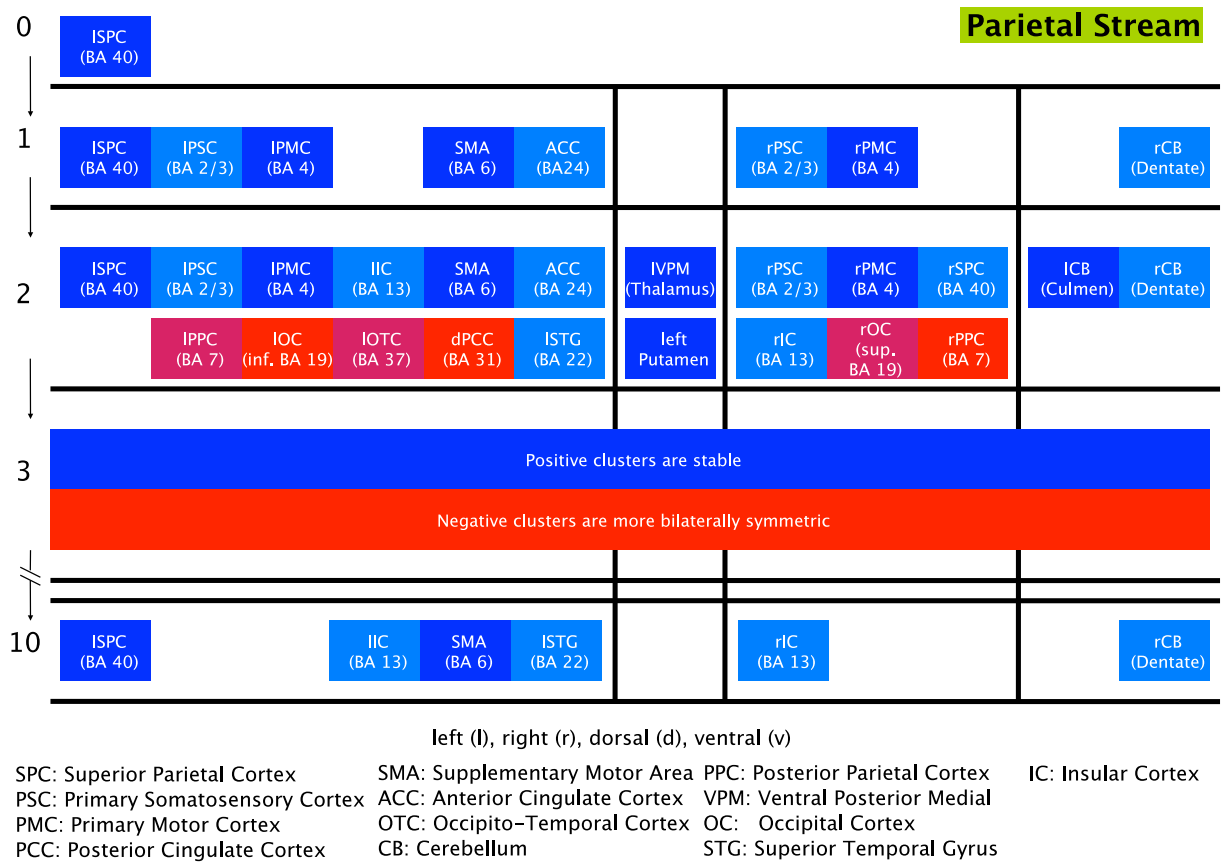



**Fig. 26.** Parietal stream. The chart summarizes the progression of the impulse response analysis, revealing how the activity flows from the seed area to other cortical and sub-cortical areas, which in their turn may also have high prediction power. One can identify, for instance, a path flowing from the posterior parietal, to the supplementary motor area and the cerebellum, to the thalamus and ipsi-lateral sensory-motor areas, highlighting the richness of the dynamics of this stream.

motor area (BA 6), and the anterior cingulate cortex (ventral BA 24). At the same time, bilateral activations of the primary somatosensory cortex (BA 2/3) and primary motor cortex (BA 4) are also observed, along with the right dentate of the cerebellum (see Fig. 26).

The next time step shows the growth of the previous clusters of activation, and the emergence of a bilateral cluster in the superior temporal gyrus (BA 22), the right superior parietal cortex (BA 40), the left culmen of the cerebellum, the left putamen, and the left ventral posterior medial nucleus (VPM) of the thalamus, as well as left insular cortex (BA 13). The negative activations also appear, in the left posterior parietal cortex (precuneus BA 31/7), left dorsal occipital cortex (BA 19), left occipito-temporal cortex (BA 37), right ventral occipital cortex (BA 19), right posterior parietal cortex (BA 7) and posterior parts of the right insular cortex (BA 13). Subsequent time steps show a stabilization of the positive clusters, and a clear symmetrization of the negative clusters; subsequent time steps show many of the initial clusters disappearing, leaving by time step 10 only the original left parietal area (BA 40), as well as clusters in BA 6, 13, 22 and right cerebellar dentate.

Moreover, as shown in Fig. 28, this stream recovers all of the activation clusters identified by GLM after just a few time steps. This finding is reassuring, as well as expected; however, it is pertinent to reiterate that our methods does so while uncovering temporal relationships between the same areas, and in a way that is consistent across subjects.

*GLM recovery and stream interactions*

As already mentioned, we find that the parietal stream recovers the GLM activation and deactivation clusters. This can be quantified using the measures of voxels cover and cluster cover described in the supplementary material, Section III (Comparison of maps). The voxel overlap analysis is presented in Fig. 28: the upper left panel shows, in blue, the percentage of voxels in the GLM activation map that overlap with the positive posterior parietal stream ($GLM+ \cap PPC+ \mid GLM+$), as a function of the IRF time steps (measured in seconds). Conversely, the red trace shows the percentage of the positive posterior parietal voxels overlapping with GLM activation voxels ($GLM+ \cap PPC+ \mid PPC+$). Superimposed is the evolution of the total number of (statistically significant) voxels in the positive posterior parietal stream ($PPC+$ vox). As the graph shows, after a few time steps, the stream map intersects with a very large percentage of the GLM voxels of activation, before slowly shrinking; on the other hand, there are relatively more voxels in the stream that do not overlap with GLM, reaching a maximum of nearly 40%. A different perspective on the relationship between the maps is presented in Fig. 29, which shows the result of our cluster cover analysis. The upper left panel shows, in the blue trace, that essentially all of the GLM activation clusters are overlapping with stream clusters, while as similarly observed with the voxels, the converse is not true, i.e. there are clusters in the stream that do not overlap with GLM clusters.

The GLM deactivation map is comparably less recovered by the negative posterior parietal stream, as shown in Fig. 28, lower right panel. This is in part due to the stream having a relative larger extent than the GLM map, as well as spanning other areas as already mentioned. In terms of clusters, however, the difference between the map and the stream is not so marked, as shown by the cluster cover measure in Fig. 29, two upper panels. This confirms that the auto-regressive analysis is, to a large extent, identifying the same spatial

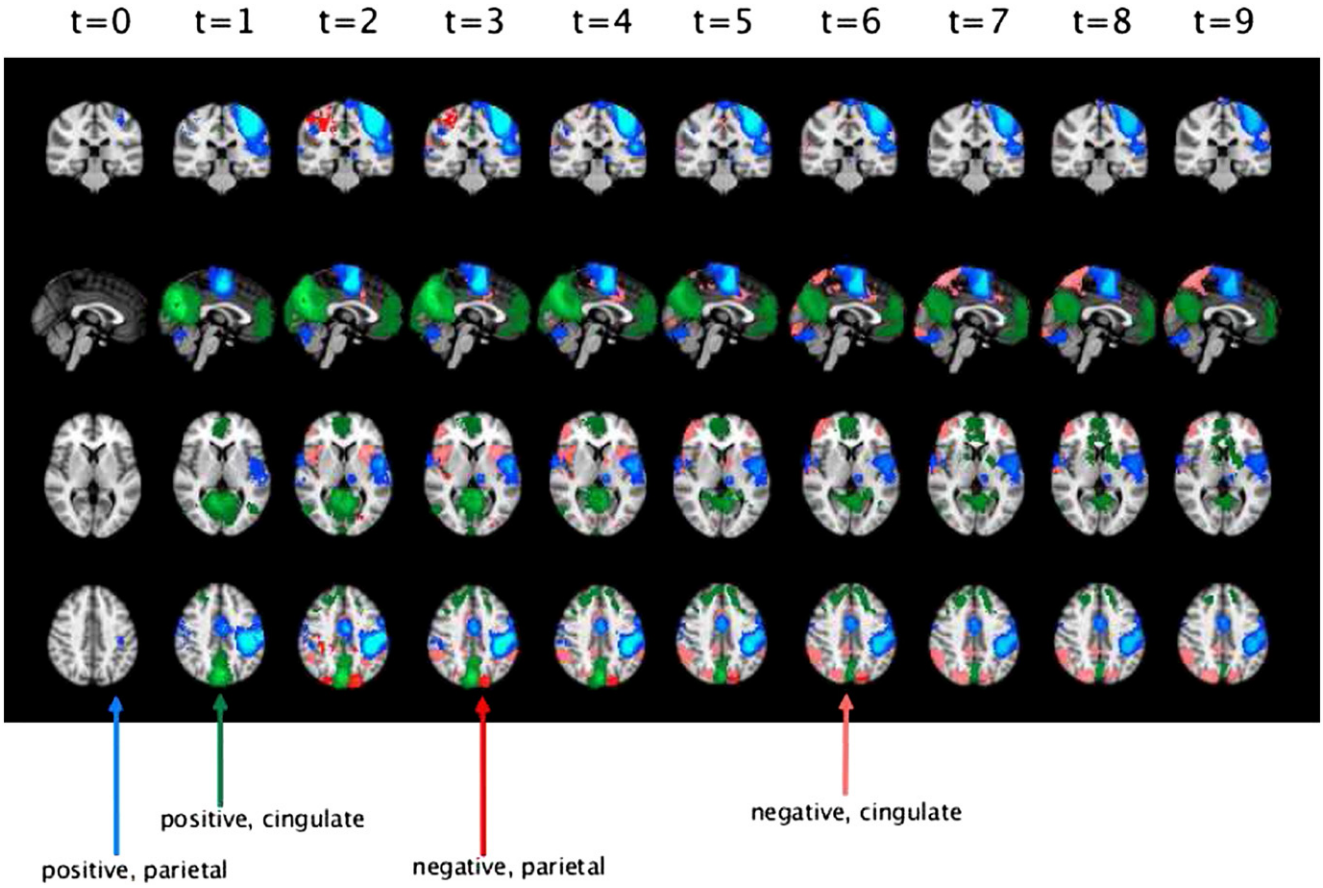



**Fig. 27.** Progression of the parietal and cingulate streams over the first 9 time steps of the impulse response function. For ease of presentation, we separated the streams into their positive and negative components, color-coded here by blue, red, green and magenta, respectively. The initial clusters of both streams grow very fast in the first time step, where evidently most of the information is present given the temporal resolution. However, there are significant effects of subsequent time steps across different areas, and a clear difference between the aggregate power of the parietal over the cingulate stream, with the latter decaying faster than the former.

patterns of activation and deactivation as GLM, providing validation to our approach.

More interestingly, however, we also observe that the parietal and precuneus streams *interact* as a function of the IRF time steps. Fig. 30 illustrates this interaction. The left panel shows that after 3 time steps, the positive parietal stream overlaps with the negative precuneus stream in a region that includes prominently the insular cortex. Fig. 30, right panel shows a similar overlap between the positive parietal and cingulate streams: after 9 time steps, the negative cingulate stream returns to the seed area for the parietal stream. It is important to emphasize what the latter finding means in the context of the IRF analysis: while the posterior parietal cluster centered in BA 40 is a strong locus of prediction power for the evolution of various sensory-motor areas, it is less strongly yet significantly predicted by the stream originating in the posterior cingulate/precuneus cluster. There is a more subtle effect: the positive parietal and negative precuneus streams grow simultaneously in the culmen of the cerebellum; that is, two contiguous regions of the cerebellum are independently predicted by the parietal and cingulate streams.

The interaction between the GLM map and the precuneus stream can also be quantified as in Fig. 28, using the voxel cover measure. Fig. 31 shows the four combinations of overlap between the positive and negative parietal and precuneus streams, superimposed for comparison on the normalized total count of voxels in the streams, as a function of the time steps. It is evident that the task imposes a global structure to the graph representing the auto-regressive model; however, the overlap between the streams is significant at different stages to warrant the claim that the interaction cannot be explained by the block design of the task. In particular, the late interaction between the positive posterior parietal and the positive posterior cingulate streams (Fig. 31, upper left) has been already pointed out in Fig. 30 (right), while the interaction with the negative cingulate stream (Fig. 31, upper right) is persistent after one time step, and includes among other areas, as mentioned, the insular cortex (Fig. 30. (left)).

## Discussion

The work of Raichle and colleagues (Fox et al., 2005) has shown the existence of two independent systems, the default mode or task-negative, and the task-positive networks respectively, displaying different task-related responses with respect to their behavior during resting conditions. The task-positive system exhibits increased activation during the performance of cognitively loaded tasks, associated with attentional and self-referential or introspective activities. This positive system spans the inter-parietal sulcus and the frontal eye fields, involved in attention, as well as a network that includes the dorsal-lateral and ventral pre-frontal areas, the insula and the supplementary motor area, more involved with cognitive processing. The task-negative or default mode network, on the other hand, significantly reduces its activity during task performance. This negative system includes lateral parietal, medial pre-frontal, and posterior cingulate cortical areas. In particular, it includes the precuneus of the posterior cingulate as one of the areas whose extensive connectivity makes it a potential fulcrum of the default mode network. Our method identifies precisely the precuneus as the

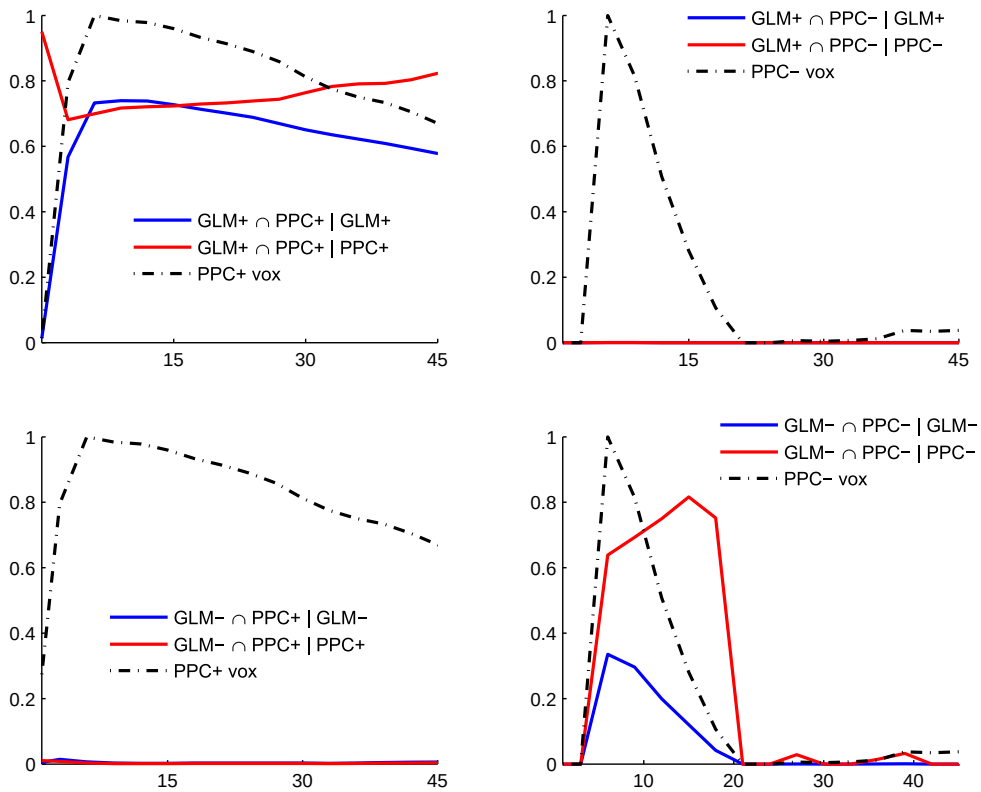



**Fig. 28.** GLM and stream overlap using the voxel cover metric. The four panels depict the overlap between the GLM activation and deactivation maps (GLM+/−), and the positive and negative posterior parietal streams (PPC+/−), in terms of the number of *voxels in common* as a fraction of the total number of voxels in each stream, for each time step. The hashed line represents the evolution of the total (normalized) number of voxels in the stream. There is a significant overlap between the GLM activation clusters and the positive posterior parietal stream (upper left), as well as between GLM deactivation and the negative posterior parietal stream (lower right).

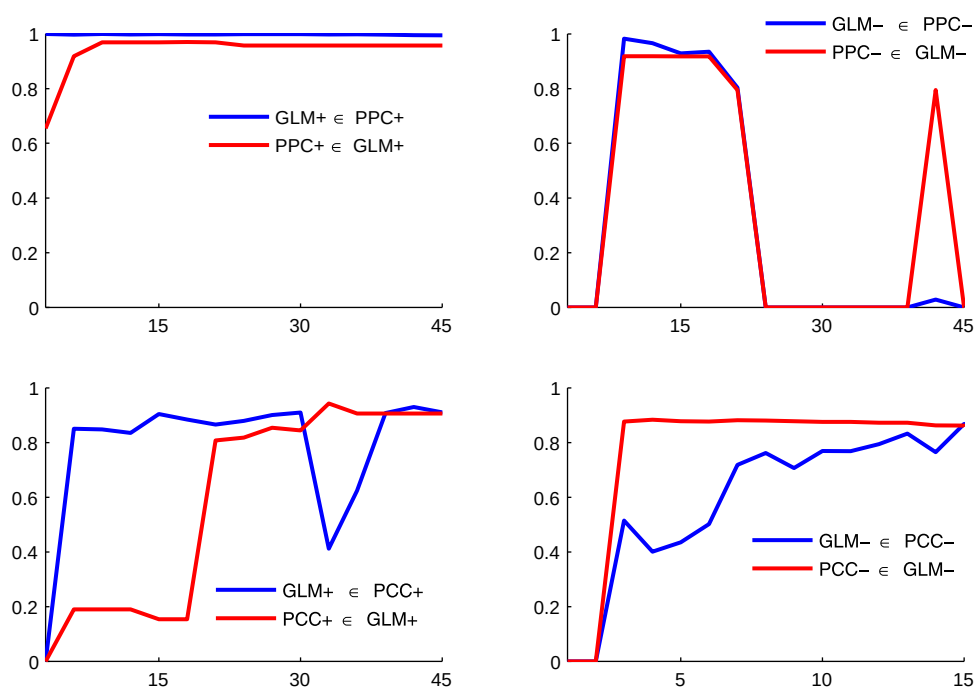



**Fig. 29.** GLM and stream overlap using the cluster cover metric: The four panels depict the overlap between the GLM activation and deactivation maps, and the posterior parietal and posterior cingulate/precuneus streams, in terms of the number of overlapping or *covered* clusters as a fraction of the total number of clusters in each stream, for each time step. The cover measure, more sensitive than the voxel overlap measure, reveals interactions between GLM de-activations and the positive posterior parietal stream, as well as GLM activations and the negative parietal stream.

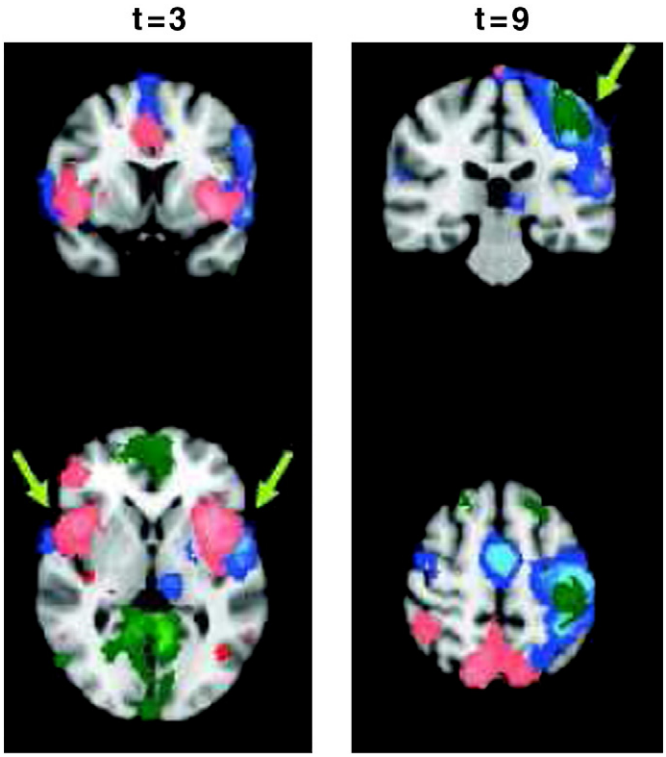



**Fig. 30.** Stream interaction: Left panel: interaction between the positive parietal stream (blue) and negative cingulate stream (pink) at time step 3; both streams converge on the insular cortex (although the positive parietal stream on the right hemisphere is mostly centered in BA 22). Right panel: interaction between the positive parietal stream and the positive cingulate stream (green), at time step 9; the parietal stream remains "active" in its own same seed area, while the cingulate stream converges on it.

largest prediction power cluster, whose corresponding stream encompasses all the regions known to participate in the task-negative network, as well as most of those in the task-positive one. This implies a consistent, predictable temporal precedence between the precuneus and the other areas, giving a specific meaning to the idea of its centrality (Frasson and Marrelec, 2008; Margulies et al., 2009).

The posterior parietal stream originates in the second-largest prediction power cluster. The dorsal inferior parietal cortex has been shown to participate in upper limb movement preparation and execution tasks, and in particular, to be differentially activated during movement imagination (Stephan et al., 1995). We interpret these facts as supporting the consistency of our results, given that besides the functional pertinence of the seed area, the evolution of the stream also includes many of the regions that are expected to be involved in the sensory-motor components of the task, such as motor, pre-motor and somatosensory areas, thalamus and cerebellum. Interestingly, the posterior parietal stream, as it evolves, spans a region inside the precuneus (see Fig. 25), which besides its stated role in the default mode network, has been reported to participate in motor imagery (Cavanna and Trimble, 2006). Our method, at least within the current resolution provided by fMRI, indicates that this area, the largest *predictor* (or driver) of future activity, is at the same time *predicted* (driven) by a task-related area, albeit with less strength and a longer time scale. This is not, however, the only region where the streams interact; in particular, we observe that they converge also on the insular cortex, which has been recently reported to share a functional connection with the precuneus in the resting state condition (Margulies et al., 2009). We conclude that the interactions between what seems to be, to some extent at least, unrelated spatio-temporal modes of activity, may provide a formal basis to understand the modulation of task-related responses by ongoing processes, such as self-monitoring or attention.

## Conclusions

In summary, we have shown that: (a) the solution of a full-scale, voxel-based auto-regressive model for functional MRI is computationally feasible, (b) the model can be visualized using the *prediction power maps* and the *impulse response dynamics*, (c) the voxels with high prediction power form a large number of small spatially contiguous clusters (mean size 4.6 voxels) which exhibit a high intra-subject as well as inter-subject consistency, (d) the results of the prediction power maps are consistent with previous results related to networks activated by the specific task, as well as networks related to the default mode activity, (e) we recover a global dynamical state dominated by two prominent (although not unique) *streams*, which (f) originate in the posterior parietal cortex (as expected in a sensory-motor task) and the posterior cingulate/precuneus cortex, respectively; moreover, these two streams (g) span default mode and task-specific networks, (h) interact in several regions, notably the insula, and imply that (i) the posterior cingulate is a central node of the default mode network, in terms of its ability to predict the future evolution of the rest of the nodes.

From a methodological point of view, our findings show that, at the very least, the proposed method can be considered as complementary to other functional connectivity methods. A direct comparison between the prediction and zero-lag correlation link-density maps (Figs. 15, 16 and 18) highlights this complementarity. Besides the removal of spurious correlations we expect to obtain through the auto-regressive modeling, the assumptions regarding the noise model and the imposition of the sparsity constraint imply that the models inferred by FARM should to be relatively sparser, and perhaps disregard valuable information. We have seen, however, how the "unfolding" of the dynamics through the impulse response function recovers a great percentage of the areas that are identified by GLM and the zero-lag correlation maps. This inclusiveness may be a result of the particular task we have analyzed, and therefore further experimentation is required to validate it.

Since the mechanisms involved in the fMRI BOLD response are still not very well understood, an element of doubt lingers on the origin of the spatio-temporal interactions discovered by full-brain auto-regressive modeling. Our results on fMRI data lead us to postulate the *information flow hypothesis* according to which, a part of the spatio-temporal patterns of activity discovered by the full-brain autoregressive modeling is due to the information flow in the neural networks. If proved wrong, this would imply that most of the patterns of activity discovered by FARM stem from the regional variability in the neuro-vascular coupling alone. The high levels of inter-subject and intra-subject consistency in prediction power maps and the consistency of the impulse response dynamics do indicate a systematic spatio-temporal pattern in the observed BOLD response. If the information flow hypothesis is incorrect, this would imply existence of a systematic pattern of temporal variability in the HRFs of different brain regions. This may shed some light on the long standing basic question in the field of fMRI-mechanics of BOLD response.

Our results on the consistency of the impulse response (which is computed without any information about the experimental protocol) with the GLM activations and de-activations and evidences available from other experiments (such as the centrality of precuneus in default mode networks (Frasson and Marrelec, 2008) and causal influence of precuneus in pathological EEG discharges during absence seizures (Vaudano et al., 2009) seem to support the information flow hypothesis. More experiments are needed to ascertain this conclusively.

If the information flow hypothesis is found to be correct, our results would mean that it is possible to investigate not just the network structure underlying a large class of brain states, but to also

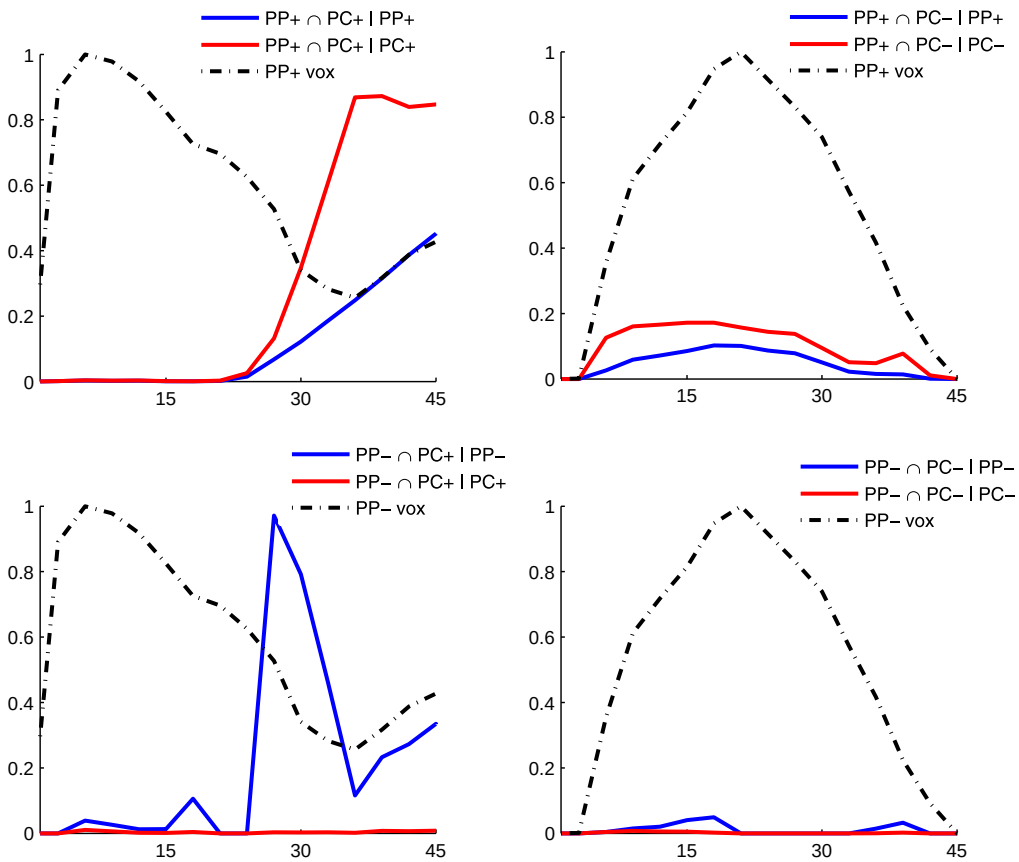



**Fig. 31.** Stream overlap using voxel cover. The four panels depict the overlap between the positive and negative projections of the posterior parietal and posterior cingulate/precuneus streams, in terms of the number of *voxels in common* as a fraction of the total number of voxels in each stream, for each time step. The hashed line represents the evolution of the total (normalized) number of voxels. The interaction between the posterior parietal and posterior cingulate is significant in specific regions at many time steps, but is extensive after time step 7.

describe them, analytically, as brain processes that progress with consistent spatio-temporal rules.

## Acknowledgments

We would like to thank Dr. A.V. Apkarian for providing us with the functional data, and his entire Pain and Passions lab at the Northwestern University for useful discussions. We also thank Dr. Irina Rish for discussions regarding sparse regression.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at doi:10.1016/j.neuroimage.2011.02.074.

## References

Achard, S., Salvador, R., Whitcher, B., Suckling, J., Bullmore, E., 2006. A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. J. Neurosci. 26 (1), 63–72 January.

Aguirre, G.K., Zarahn, E., D'Esposito, M., 1998. The variability of human, BOLD hemodynamic responses. NeuroImage 8 (4), 360–369.

Anderson, C.W., Stolz, E.A., Shamsunder, S., 1998. Multivariate autoregressive models for classification of spontaneous electroencephalographic signals during mental task. IEEE. Trans. Biomed. Eng. 45 (3), 277–286 March.

Aoki, T., Tsuda, H., Takasawa, M., Osaki, Y., Oku, N., Hatazawa, J., Kinoshita, H., 2005. The effect of tapping finger and mode differences on cortical and subcortical activities: a PET study. Exp. Brain Res. 160, 375–383.

Arieli, A., Sterkin, A., Grinvald, A., Aertsen, A., 1996. Dynamics of ongoing activity: explanation of the large variability in evoked cortical responses. Science 273 (5283), 1868–1871.

Binkofski, F., Amunts, K., Stephan, K.M., Posse, S., Schormann, T., Freund, H.-J., Zilles, K., Seitz, R.J., 2000. Broca's region subserves imagery of motion: a combined cytoarchitectonic and fMRI study. Hum. Brain Mapp. 11 (4), 273–285.

Box, G.E.P., Jenkins, G.M., Reinsel, G.C., 2008. Time Series Analysis: Forecasting and Control (Wiley Series in Probability and Statistics), 4 edition. Wiley. June.

Brockwell, P.J., Davis, R.A., 1986. Time Series: Theory and Methods. Springer-Verlag New York, Inc., New York, NY, USA.

Brovelli, A., Ding, M., Ledberg, A., Chen, Y., Nakamura, R., Bressler, S.L., 2004. Beta oscillations in a large-scale sensorimotor cortical network: Directional influences revealed by Granger causality. Proc. Natl. Acad. Sci. U. S. A. 101 (26), 9849–9854 June.

Buchel, C., Friston, K.J., 1997. Modulation of connectivity in visual pathways by attention: cortical interactions evaluated with structural equation modelling and fMRI. Cereb. Cortex 7 (8), 768–778.

Bullmore, E., Sporns, O., 2009. Complex brain networks: graph theoretical analysis of structural and functional systems. Nat. Rev. Neurosci. 10 (3), 186–198 March.

Buxton, R.B., Uludag, K., Dubowitz, D.J., Liu, T.T., 2004. Modeling the hemodynamic response to brain activation. NeuroImage 23 (Supplement 1), S220–S233 Mathematics in Brain Imaging.

Candès, E.J., 2008. The restricted isometry property and its implications for compressed sensing. C. R. Acad. Sci. Paris Ser. I 346, 589–592.

Candès, E.J., Romberg, J., 2004. Practical signal recovery from random projections. SPIN Conference on Wavelet Applications in Signal and Image Processing.

Carroll, M.K., Cecchi, G.A., Rish, I., Garg, R., Rao, A.R., 2009. Prediction and interpretation of distributed neural activity with sparse models. NeuroImage 44, 112–122.

Cavanna, A.E., Trimble, M.R., 2006. The precuneus: a review of its functional anatomy and behavioural correlates. Brain 129, 564–583.

Cecchi, G.A., Rao, A.R., Centeno, M.V., Baliki, M., Apkarian, A.V., Chialvo, D.R., 2007. Identifying directed links in large scale functional networks: application to brain fMRI. BMC Cell Biol. 8 (1), S5.

Cecchi, G.A., Garg, R., Rao, A.R., 2008a. Inferring brain dynamics using Granger causality on fMRI data. Proceedings of the 2008 IEEE International Symposium on Biomedical Imaging (ISBI'08), pp. 604–607.

Cecchi, G.A., Ma'ayan, A., Rao, A.R., Wagner, J., Iyengar, R., Stolovitzky, G., 2008b. Ordered cyclic motifs contributes to dynamic stability in biological and engineered networks. Proc. Natl. Acad. Sci. U. S. A. 105, 19235–19240.

Cecchi, G.A., Rish, I., Thyreau, B., Thirion, B., Plaze, M., Palliere-Martinot, M.-L., Martelli, C., Martinot, J.-L., Poline, J.-B., 2009. Discriminative network models of schizophrenia. In proceedings of Advances in Neural Information Processing Systems (NIPS).
Chen, S.S., Donoho, D.L., Saunders, M.A., 2001. Atomic decomposition by basis pursuit. SIAM Rev. 43 (1), 129–159.
Chen, G., et al., 2009. Granger causality via vector auto-regression tuned for fMRI data analysis. ISMRM 17th Scientific Meeting.
Cunnington, R., Windischberger, C., Deecke, L., Moser, E., 2001. The preparation and execution of self-initiated and externally-triggered movement: a study of event-related fMRI. NeuroImage 15, 373–385.
D'Esposito, M., Deouell, L.Y., Gazzaley, A., 2003. Alterations in the BOLD fMRI signal with ageing and disease: a challenge for neuroimaging. Nat. Rev. Neurosci. 4 (11), 863–872 November.
Dahlhaus, R., Eichler, M., 2003. Causality and graphical models in time series analysis. In: Green, P.J., Hjort, N.L., Richardson, S. (Eds.), Highly Structured Stochastic Systems. Oxford University Press, London, pp. 115–137.
Darvas, F., Leahy, R.M., 2007. Functional imaging of brain activity and connectivity with MEG. Handbook of Brain Connectivity. Springer, Berlin / Heidelberg, pp. 201–220.
Deshpande, G., LaConte, S., James, G.A.A., Peltier, S., Hu, X., 2009. Multivariate Granger causality analysis of fMRI data. Hum. Brain Mapp. 30 (4), 1361–1373 April.
Dodel, S., Herrmann, J.M., Geisel, T., 2002. Functional connectivity by cross-correlation clustering. Neurocomputing 44–46, 1065–1070.
Duann, J.-R., Jung, T.-P., Kuo, W.-J., Yeh, T.-C., Makeig, S., Hsieh, J.-C., Sejnowski, T.J., 2002. Single-trial variability in event-related BOLD signals. NeuroImage 15 (4), 823–835.
Edelman, G., Tononi, G. (Eds.), 2001. A Universe Of Consciousness How Matter Becomes Imagination. Basic Books.
Efron, B., Hastie, T., Johnstone, I., Tibshirani, R., 2004. Least angle regression. Ann. Stat. 32 (1), 407–499.
Eguiluz, V.M., Chialvo, D.R., Cecchi, G.A., Baliki, M., Apkarian, V.A., 2005. Scale-free brain functional networks. Phys. Rev. Lett. 94, 018–102 Jan.
Eichler, M., 2005. A graphical approach for evaluating effective connectivity in neural systems. Phil. Trans. R. Soc. B 360 (1457), 953–967 May.
Fox, M.D., Snyder, A.Z., Vincent, J.L., Corbetta, M., Van Essen, D.C., Raichle, M.E., 2005. The human brain is intrinsically organized into dynamic, anti-correlated functional networks. Proc. Natl. Acad. Sci. U. S. A. 102, 9673–9678.
Frasson, P., Marrelec, G., 2008. The precuneus/posterior cingulate cortex plays a pivotal role in the default mode network: evidence from a partial correlation analysis. NeuroImage 42, 1178–1184.
Friston, K.J., 1994. Functional and effective connectivity in neuroimaging: a synthesis. Hum. Brain Mapp. 2 (1–2), 56–78.
Friston, K.J., Holmes, A.P., Poline, J.-B., Grasby, P.J., Williams, S.C.R., Frackowiak, R.S.J., Turner, R., 1995. Analysis of fMRI time-series revisited. NeuroImage 2 (1), 45–53.
Friston, K.J., Mechelli, A., Turner, R., Price, C.J., 2000. Nonlinear responses in fMRI: the balloon model, volterra kernels, and other hemodynamics. NeuroImage 12 (4), 466–477.
Friston, K.J., Harrison, L., Penny, W., 2003. Dynamic causal modelling. NeuroImage 19 (4), 1273–1302.
Friston, K., et al., 2007. Statistical parametric mapping, the analysis of functional brain images.
FSL Release 3.3. http://www.fmrib.ox.ac.uk/fsl 2006.
Gajic, Z., Qureshi, M.T.J., 1995. Lyapunov Matrix Equation in Systems Stability and Control. Academic Press, San Diego.
Garg, R., Khandekar, R., 2009. Gradient descent with sparsification: an iterative algorithm for sparse recovery with restricted isometry property. In Proceedings of the 26th International Conference on Machine Learning. Omnipress, Montreal, pp. 337–344. June.
Garg, R., Cecchi, G.A., Rao, A.R., 2009. Applications of high-performance computing to functional magnetic resonance imaging (fMRI) data. In: Rao, A. Ravishankar, Cecchi, Guillermo A. (Eds.), High-throughput Image Reconstruction and Analysis. Artech House Publishers, City, State of Publication, pp. 263–282. chapter 12.
Gitelman, D.R., Penny, W.D., Ashburner, J., Friston, K.J., 2003. Modeling regional and psychophysiologic interactions in fMRI: the importance of hemodynamic deconvolution. NeuroImage 19 (1), 200–207.
Glover, G.H., 1999. Deconvolution of impulse response in event-related BOLD fMRI. NeuroImage 9 (4), 416–429.
Goebel, R., Roebroeck, A., Kim, D.-S., Formisano, E., 2003. Investigating directed cortical interactions in time-resolved fMRI data using vector autoregressive modeling and Granger causality mapping. Magn. Reson. Imaging 21 (10), 1251–1261.
Granger, C.W.J., 1969. Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37 (3), 424–438.
Greicius, M.D., Krasnow, B., Reiss, A.L., Menon, V., 2003. Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. Proc. Natl. Acad. Sci. U. S. A. 100, 253–258.
Greicius, M.D., Srivastava, G., Reiss, A.L., Menon, V., 2004. Default-mode network activity distinguishes Alzheimers disease from healthy aging: evidence from functional MRI. Proc. Natl. Acad. Sci. U. S. A. 101, 4637–4642.
Grinband, J., Wager, T.D., Lindquist, M., Ferrera, V.P., Hirsch, J., 2008. Detection of time-varying signals in event-related fMRI designs. NeuroImage 43 (3), 509–520.
Handwerker, D.A., Ollinger, J.M., D'Esposito, M., 2004. Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. NeuroImage 21 (4), 1639–1651.
Harrison, L., Penny, W.D., Friston, K.J., 2003. Multivariate autoregressive modelling of fMRI time series. NeuroImage 19 (4), 1477–1491.
IBM Blue Gene Team, 2008. Overview of the IBM Blue Gene/P project. IBM J. Res. Dev. 52 (1/2), 199–220.
Jäncke, L., Loose, R., Lutz, K., Specht, K., Shah, N.J., 2000. Cortical activations during paced finger-tapping applying visual and auditory pacing stimuli. Cogn. Brain Res. 10, 51–66.
Kamiński, M., Ding, M., Truccolo, W.A., Bressler, S.L., 2001. Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance. Biol. Cybern. 85 (2), 145–157 August.
Koch, C., Davis, J. (Eds.), 1995. Large-scale neuronal theories of the brain, chapter 6: Perception as an oneiric-like state modulated by the senses. MIT Press.
Li, W., Crist, R.E., Gilbert, C.D., 2001. Learning to see: experience and attention in primary visual cortex. Nat. Neurosci. 4 (5), 519–525.
Li, K., Guo, L., Nie, J., Li, G., Liu, T., 2009. Review of methods for functional brain connectivity detection using fMRI. Comput. Med. Imaging Graph. 33 (2), 131–139.
Lippert, M.T., Steudel, T., Ohl, F., Logothetis, N.K., Kayser, C., 2010. Coupling of neural activity and fMRI-BOLD in the motion area MT. Magn. Reson. Imaging 28 (8), 1087–1094.
Logothetis, N.K., 2008. What we can do and what we cannot do with fMRI. Nature 453 (7197), 869–878 June.
Lu, Y., Bagshaw, A.P., Grova, C., Kobayashi, E., Dubeau, F., Gotman, J., 2006. Using voxel-specific hemodynamic response function in EEG-fMRI data analysis. NeuroImage 32 (1), 238–247.
Machamer, P., Wolters, G., 2007. Thinking about Causes: From Greek Philosophy to Modern Physics. University of Pittsburgh Press. May.
Margulies, D.S., Vincent, J.L., Kellye, C., Lohmann, G., Uddin, L.Q., Biswal, B.B., Villringer, A., Castellanos, F.X., Milham, M.P., Petrides, M., 2009. Precuneus shares intrinsic functional architecture in humans and monkeys. Proc. Natl. Acad. Sci. U. S. A. 106, 20069–20074.
Marrelec, G., Krainik, A., Duffau, H., Plgrini-Issac, M., Lehricy, S., Doyon, J., Benali, H., 2006. Partial correlation for functional brain interactivity investigation in functional MRI. NeuroImage 32 (1), 228–237.
Mcintosh, A.R., Gonzalez-Lima, F., 1994. Structural equation modeling and its application to network analysis in functional brain imaging. Hum. Brain Mapp. 2 (1–2), 2–22.
McKeown, M.J., Makeig, S., Brown, G.G., Jung, T.-P., Kindermann, S.S., Bell, A.J., Sejnowski, T.J., 1998. Analysis of fMRI data by blind separation into independent spatial components. Hum. Brain Mapp. 6, 160–188.
Menon, R.S., Kim, S.-G., 1999. Spatial and temporal limits in cognitive neuroimaging with fMRI. Trends Cogn. Sci. 3 (6), 207–216.
Menon, R.S., Luknowsky, D.C., Gati, J.S., 1998. Mental chronometry using latency-resolved functional MRI. Proc. Natl. Acad. Sci. U. S. A. 95 (18), 10902–10907.
Mitchell, T.M., Shinkareva, S.V., Carlson, A., Chang, K.M., Malave, V.L., Mason, R.A., Just, M.A., 2008. Predicting human brain activity associated with the meanings of nouns. Science 320 (5880), 1191–1195 May.
Natarajan, B.K., 1995. Sparse approximate solutions to linear systems. SIAM J. Comput. 24 (2), 227–234.
Neumann, J., Lohmann, G., Zysset, S., von Cramon, D.Y., 2003. Within-subject variability of BOLD response dynamics. NeuroImage 19 (3), 784–796.
Neylon, T., 2006. Sparse solutions for linear prediction problems. PhD thesis, Courant Institute, New York University, April 2006.
Norman, K.A., Polyn, S.M., Detre, G.J., Haxby, J.V., 2006. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends Cogn. Sci. 10 (9), 424–430 September.
Pearl, J., 1998. Graphs, causality, and structural equation models. Sociol. Methods Res. 27, 226–284.
Penny, W.D., Stephan, K.E., Mechelli, A., Friston, K.J., 2004. Comparing dynamic causal models. NeuroImage 22 (3), 1157–1172.
Pereda, E., Quiroga, R.Q., Bhattacharya, J., 2005. Nonlinear multivariate analysis of neurophysiological signals. Prog. Neurobiol. 77 (1–2), 1–37 September.
Raichle, M.E., MacLeod, A.M., Snyder, A.Z., Powers, W.J., Gusnard, D.A., Shulman, G.L., 2001. A default mode of brain function. Proc. Natl. Acad. Sci. U. S. A. 98 (2), 676–682.
Riecker, A., Wildgruber, D., Mathiak, K., Grodd, W., Ackermann, H., 2003. Parametric analysis of rate-dependent hemodynamic response functions of cortical and subcortical brain structures during auditorily cued finger tapping: a fMRI study. NeuroImage 18, 731–739.
Roebroeck, A., Formisano, E., Goebel, R., 2005. Mapping directed influence over the brain using Granger causality and fMRI. NeuroImage 25 (1), 230–242.
Salvador, R., Suckling, J., Schwarzbauer, C., Bullmore, E., 2005. Undirected graphs of frequency-dependent functional connectivity in whole brain networks. Philos. Trans. R. Soc. Lond. B Biol. Sci. 360 (1457) May.
Salvador, R., Martnez, A., Pomarol-Clotet, E., Sarr, S., Suckling, J., Bullmore, E., 2007. Frequency based mutual information measures between clusters of brain regions in functional magnetic resonance imaging. NeuroImage 35 (1), 83–88.
Sato, J.R., Morettin, P.A., Arantes, P.R. Amaro E., Jr., 2007. Wavelet based time-varying vector autoregressive modelling. Comput. Stat. Data Anal. 51 (12), 5847–5866.
Schneidman, E., Berry II, M.J., Segev, R., Bialek, W., 2006. Weak pairwise correlations imply strongly correlated network states in a neural population. Nature 440, 1007–1012.
Seth, A.K., 2005. Causal connectivity of evolved neural networks during behavior. Netw. Comput. Neural Syst. 16 (1), 35–54.
Skipper, J.I., Goldin-Meadow, S., Nusbaum, H.C., Small, S.L., 2007. Speech-associated gestures, Broca's area, and the human mirror system. Brain Lang. 101 (3), 260–277.
Small, M., 2005. Applied Nonlinear Time Series Analysis: Applications in Physics, Physiology and Finance. World Scientific Publishing Company.
Smith, B.D., Bullmore, E., 2006. Small-world brain networks. Neuroscientist 12, 512–523.
Stam, C.J., Jones, B.F., Nolte, G., Breakspear, M., Scheltens, Ph., 2007. Small-world networks and functional connectivity in Alzheimer's disease. Cereb. Cortex 17 (1), 92–99.

Stephan, K.M., Fink, G.R., Passingham, R.E., Silbersweig, D., Ceballos-Baumann, A.O., Frith, C.D., Frackowiak, R.S., 1995. Functional anatomy of the mental representation of upper extremity movements in healthy subjects. J. Neurophysiol. 73, 373–386.

Strang, G., 1988. Linear Algebra and Its Applications, third ed. Harcourt Brace Jovanovich College Publishers.

Sun, F.T., Miller, L.M., D'Esposito, M., 2004. Measuring interregional functional connectivity using coherence and partial coherence analyses of fMRI data. NeuroImage 21 (2), 647–658.

Thomason, M.E., Burrows, B.E., Gabrieli, J.D.E., Glover, G.H., 2005. Breath holding reveals differences in fMRI BOLD signal in children and adults. NeuroImage 25 (3), 824–837.

Tibshirani, R., 1996. Regression shrinkage and selection via the Lasso. J. R. Stat. Soc. B 58 (1), 267–288.

Valdes-Sosa, P.A., Sanchez-Bornot, J.M., Lage-Castellanos, A., Vega-Hernandez, M., Bosch-Bayard, J., Melie-Garcia, L., Canales-Rodriguez, E., 2005. Estimating brain functional connectivity with sparse multivariate autoregression. Philos. Trans. R. Soc. Lond. B Biol. Sci. 360 (1457), 969–981.

Vaudano, A.E., Laufs, H., Kiebel, S.J., Carmichael, D.W., Hamandi, K., Guye, M., Thornton, R., Rodionov, R., Friston, K.J., Duncan, J.S., Lemieux, L., 2009. Causal hierarchy within the thalamo-cortical network in spike and wave discharges. PLoS One 4 (8), e6475 08.

Vazquez, A.L., Noll, D.C., 1998. Nonlinear aspects of the BOLD response in functional MRI. NeuroImage 7 (2), 108–118.

Zivot, E., Wang, J., 2006. Modeling Financial Time Series with S-PLUS®. Springer-Verlag New York, Inc., Secaucus, NJ, USA.