Using an ODE model to separate Rest and Task signals in fMRI

Amrit Kashyap[1*], Eloy Geenjaar[2,3], Patrik Bey[1], Kiret Dhindsa[1], Katharina Glomb[1], Sergey Plis[3], Shella Keilholz[2,4], Petra Ritter[1]

[1*] Charite University Medizin, Berlin, Germany.
[2] Georgia Institute of Technology, Atlanta, United States.
[4] Emory University, Atlanta , United States.
[3]Georgia State University, Atlanta, United States.

*Corresponding author(s). E-mail(s): amrit.kashyap@charite.de;

Abstract

Cortical activity results from the interplay between network-connected regions that integrate information and stimulus-driven processes originating from sensory motor networks responding to specific tasks. Separating the information due to each of these components has been challenging, and the relationship as measured by fMRI in each of these cases Rest (network) and Task (stimulus driven) remains a significant outstanding question in the study of large-scale brain dynamics. In this study, we developed a network ordinary differential equation (ODE) model using advanced system identification tools to analyze fMRI data from both rest and task conditions. We demonstrate that task-specific ODEs are essentially a subset of rest-specific ODEs across four different tasks from the Human Connectome Project. By assuming that task activity is a relative complement of rest activity, our model significantly improves predictions of reaction times on a trial-by-trial basis, leading to a 9 % increase in explanatory power ($\boldsymbol{R^2}$) across all 14 tested subtasks. Our findings establish the principle of the Active Cortex Model, which posits that the cortex is always active and that Rest State encompasses all processes, while certain subsets of processes get elevated to perform specific task computations. This study offers a crucial perspective on the nature of large-scale brain dynamics and introduces one of the first models to causally link large-scale brain activity, equations representing these dynamics, and behavioral variables within a single framework.

Keywords: fMRI, DTI, Brain Network Model, SINDy, Behavioral Correlates

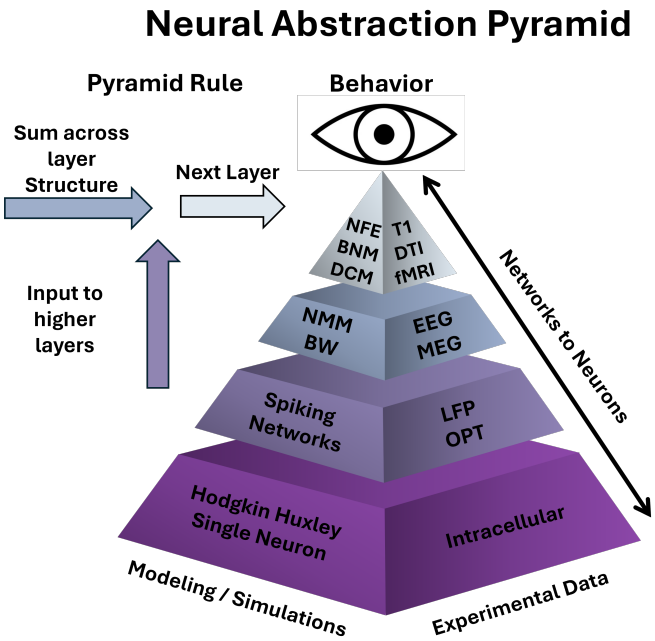

Fig. 1 The Neural Abstraction Pyramid representing different layers of neural activity from neurons to the whole network. Behavior is represented at the top most layer as it emerges from the activity of different components of the pyramid. Legend: BNM -Brain Network Model, NFE- Neural Field Equations, fMRI - functional Magnetic Resonance Imaging, DCM - Dynamical Causal Modeling, T1 - Structural Image, DTI - Diffusion Tensor Imaging, NMM -Neural Mass Models, EEG- Electroencephalogram, MEG - Magnetoencephalography, BW - Balloon Windkessel Model, LFP - Local Field Potential, OPT- Optogenetics Image Ref

1 Introduction

The baseline activity of the nervous system is surprisingly non-stationary and complex even when the organism is not performing any particular Task. Although this requires enormous metabolic cost, this activity known as Resting State, is hypothesized to exist in order to maintain the neural computational hardware rather than solving a particular problem [1]. Resting state activity is thought to originate due to cortical loops that exist in the white matter macroscale network of the cortex which lead to oscillatory style dynamics as brain areas cyclically activate each other. The presence of Resting State dynamics is ubiquitous even when a subject is requested to perform a particular Task, and exists in addition to the stimulus driven activity that is localized to brain regions responding to the particular behavioral assay. Thus, cortical activity is hypothesized to be a combination of stimulus dependent responses as well as the background activity generated by cortical loops although the exact relationship between them remains unclear. Therefore an important outstanding question is to understand the relationship between them and determine how to separate these two distinct processes from measured cortical activity such as in functional Magnetic Resonance Imaging (fMRI) [1] [2] [3]. Determining this relationship, would not only

further our fundamental understanding of how the nervous system operates, but would also have practical consequences in improving our ability to analyze and decode Task fMRI activity for which over 100,000 unique studies have been performed [4]. Computational models of cortical activity based on macroscopic networks have progressed to a point, where we can extend them in the following manner in order to answer the pertinent questions about the relationship between Rest and Task [5].

Current simulations of cortical activity can be understood by examining the abstraction layers of the neural pyramid as shown in Figure 1. Models at different abstraction layers of the nervous system are aimed at recapitulating different measurements that record the activity across various scales in the central nervous system. In Figure 1, the lower levels of the neural pyramid denote models of individual components of the nervous system such as neurons that can be fully modeled by differential equations. Higher layers represent the aggregate activity of individual lower-level elements across a structure e.g. a cortical column, or the macro level organization of white matter connections which eventually leads to the emergence of Behavior. Input from the lower levels also plays a role in shaping the dynamics of the subsequent upper levels.

In the context of modeling fMRI, this has led to using models that utilize macrostructes to aggregate the activity of smaller nerual elements to simulate fMRI activity [6] [7]. This has led to Brain Network Models (BNMs) which use estimates of white matter connectivity and Neural Field Equation (NFE) which uses the surface structure of the cortex to model large scale neural Resting State activity [8] [9] [10]. Input from lower levels is typically represented as noise, since these inputs involve unobservable variables and occur at random with respect to the measured signal [11]. However, simulations of these noise models, are ill equipped in answering any questions in transient dynamics that occurs during Task switching, as they integrate noise and have been largely relegated to recapitulating properties of steady state dynamics [8]. In order to model fMRI transient dynamics directly, it would require an ODE framework that for any measurable time point of fMRI activity can estimate the network component of the fMRI timeseries.

To this aim we utilize the Sparse Identification of Nonlinear Dynamics (SINDy), a framework of approximating the underlying Ordinary Diffrential Equation (ODE) from measured timeseries and can predict the network component of the activity [12]. Although we are the first to apply SINDy to determine an fMRI equation, the rationale to use a similar technique was one of the first tools applied on fMRI data, namely Dynamic Causal Modeling (DCM) [13]. However, since SINDy utilizes regression, the algorithm scales much better than DCM while providing stable consistent solutions unlike deep learning algorithms. We utilize SINDy to construct a network model, where each edge E between brain regions a and b is approximated using a Taylor polynomial of order p with corresponding coefficients Ξ_n: $E_{ab}(X_b) = \sum_{n=1}^{p} \Xi_n X_b^n$. The derivative of a single region is estimated as a sum across all edges $\dot{X}_a = \sum E_{ab}(X_b)$. Since these polynomials are fitted using L2-regression there exists a trade-off in either matching the dynamics or having sparse network solutions that are closer to white matter connectivity networks as estimated by DTI. In order to constrain the possible ODE solutions, we apply the following three criteria to ensure that they are biophysically

plausible: 1) The loss function namely how much the ODE matches the derivative of the fMRI signal 2) The degree of the Taylor series coefficients Ξ_n are related to the SC network as measured through Tractography such as in BNM 3) the Hemodynamic Response Function (HRF) of the system and comparing it to the canonical HRF response. By optimizing the metrics of these three criteria, it is possible to approximate network dynamics to some extent, enabling the examination of the relationship between Rest and Task fMRI. In this manner, separate ODEs for Rest and Task fMRI were trained on utilizing different datasets of the Human Connectome Project (HCP) data (Rest, Gambling, Working Memory, Emotion, Relational).

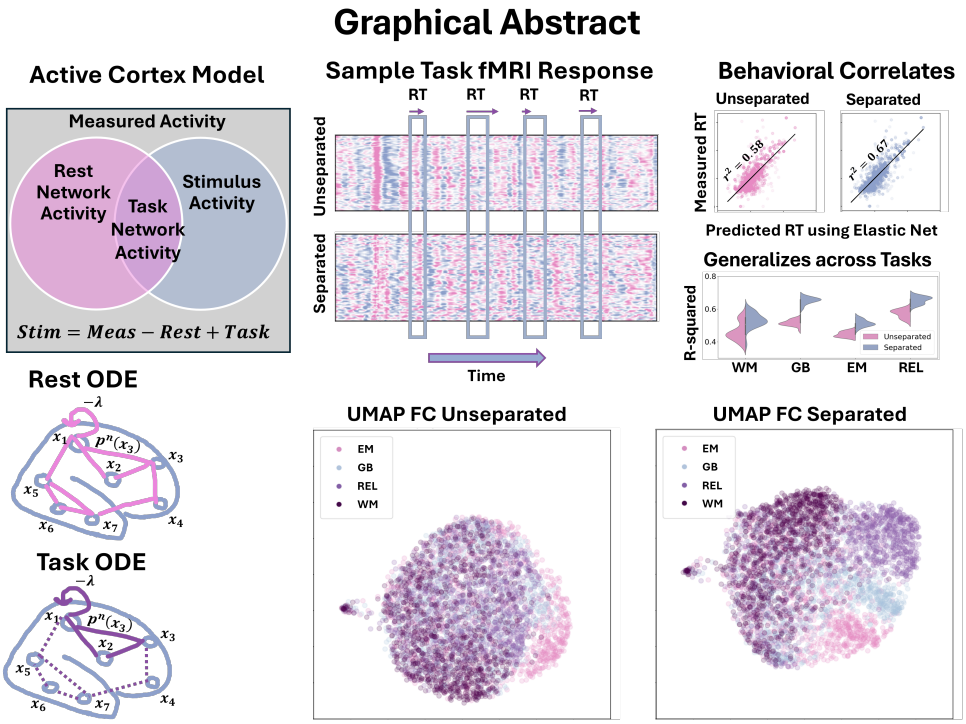

Fig. 2 Top Left: The Active Cortex Model - The stimulus activity couples with the Resting State Networks to produce Task Network activity. Bottom Left: The Rest ODE and specific Task ODE are estimated from Rest and specific Task fMRI using data driven approaches to estimate a network ODE by approximating Taylor series polynomials to describe each of the edge interactions. The Task ODE here is a subset of Rest ODE. Top Middle: The Task fMRI is separated by subtracting the Rest processes not present in that specific Task activation. Top Right: The reaction times for each trial have about 9 percent higher R^2 with the separated data vs the measured data on average across four HCP Tasks. Bottom Right: By removing the Rest processes not present in the specific Task, the functional connectivity separates in high dimensional space providing more interpretability between distinct Tasks. HCP Task abbreviations: EM- Emotional, GB- Gambling, REL- Relational, Working Memory

Interpreting the estimated ODE in terms of network and stimulus processes still remains challenging as stimulus processes could be a component of the identified network activity itself. Therefore, an experiment was designed using combinations of trained Rest and Task ODEs to estimate the stimulus-independent background activity. This background activity was then removed from the measured signal to improve the isolatation of the stimulus-dependent signal, as illustrated in 2. We verified the separation of the signal by comparing with the reaction time (RT) it took to complete each trial and applied this methodology across all HCP tasks. We show that the if and only if we assume the Active Cortex Model, where the Task specific network ODE are a subset of Rest network ODE and the background stimulus independent activity can be approximated by the difference between Rest and Task ODEs models, we are able to correctly separate the signal. We show that the Task network activity is specific to each type of Task, i.e Working Memory vs Gambling, but they are all subsets of the Rest ODE network. We quantify the total amount of increase in r-squared that our separation is able to explain, which is around 9 percent on average across all 14 subtasks. Moreover, we show that after the separation of background activity, the functional connectivity of separated signals naturally segregate by Task in a high dimensional space, unlike the functional connectivity of the measured unseparated signal as shown via Uniform Manifold Approximation and Projection (UMAP) in Figure 2 [14]. Thus, we conclude that the primary organization of the Rest and Task can be approximated by the Active Cortex Model, aptly named as all processes are active during Rest and Task, but a certain subset are elevated to perform the particular Task which confirms this hypothesis that has also been stated previously in literature [15] [1].

While more research needs to be performed to verify the relationship between Rest and Task, this work is a significant advancement in utilizing modeling in a practical manner to isolate the stimulus component of the signal. Thus this study represents one of the first successful constructions an ODE model that links structural connectivity, Rest-Task networks dynamics and behavioral variables in one coherent causal model. The stud, moreover, provides concrete evidence of the fundamental framework relating Rest and Task, namely the Active Cortex Model. The model can be potentially applied to improve all Task fMRI studies, influencing the already published 100,000 Task studies [4]. Therefore, we believe that in the future, this work will have a large impact on the field that studies the link between brain- behavior relationships.

2 Results

The results section is organized into four sections. The first section introduces how we construct and validate network ODEs to represent network activity of Rest and Task fMRI data. The second section shows that during Task, the stimulus couples with a component of the Rest network. The Task activity can only be separated from background processes by removing only the processes present in the Rest network activity but not present in the Task network activity. The third section quantifies how much variance is explained by removing the background processes and shows that the separation results in increasing the R^2 by 9 percent on average across all 14 subtasks. The fourth section examines the relationship between different Tasks and shows that

by removing the shared background processes, the distinct Task FCs diverge from each other in hyper-dimensional space as the processes common to them are reduced.

2.1 Network Model for fMRI

2.1.1 Structural

To construct a network model for fMRI, we assume that the measured data is composed of network activity and unstructured residual activity as shown in Figure 3A. By contrast, the network activity is considered to be structured and can be modeled as a relationship between the distinct nodes. To construct this network, we fit a separate function for each edge in the network using a Taylor series polynomial p^n via SINDy and define the change in activity as the sum of all edges. A sparsity parameter determines the number of edges used to represent the ODE. To test whether these estimated edges are biophysically relevant, we examine the distribution of edge strengths with other measures of network connectivity namely using structural connectivity SC estimated from DTI. The coefficients are highly correlated (> 0.5 on

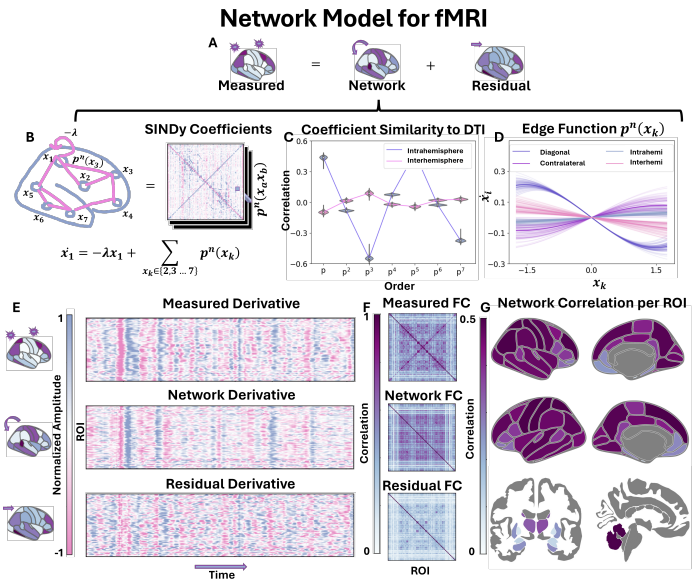

Fig. 3 A) A simplified diagram of separating the measured signal into network and unstructured residual components. B) The network component consists of estimating a Taylor series polynomial between each pair of brain regions. C) The correlation between the coefficients Ξ at each polynomial order to the tractography computed separately for inter and intrahemispheric edges. D) A derivative plot of all estimated edge functions between pairs of brain regions. E) An example of separating the measured signal derivative into network and stimulus components normalized to have the same amplitude. F) The resulting FCs of the measured signal, network and residual components. G) The correlation between the network and measured derivative timeseries computed separately for each brain region. All the results above are shown for Resting State fMRI ODE while the Task specific differences are shown in Methods 3.3.

average) for the odd orders of the polynomial for edges within a hemisphere (intra-hemisphere) while the even orders are small in magnitude and have low correlations (< 0.05 on average) with SC as shown in Figure 3C. We plotted the resulting odd edge functions using a derivative plot in Figure 3D. They are divided into four categories: 1) diagonal (self edges), 2) intrahemispheric, 3) contralateral (edges between homologous brain regions), interhemispheric (excluding contralateral edges) in Figure 3D. They all consist of sigmoid functions, where the relationship is linear around the origin and plateaus at the extrema. The intrahemispheric and contralateral edges are positive sigmoid functions, but the other interhemispheric connections (non-contralateral) have an opposite sign. The main difference between our identified network ODE and previously published BNMs is that our model has negative interhemispheric (non-contralateral) edges, while previously all positive functions except for the diagonal, were used to describe the dynamics (see Discussion 4.1.1). The diagonal represents the decay term and is a hyperparameter in our model and affects the HRF (see Methods 3.2.2). Setting values for this parameter as well as other hyperparameters that control the sparsity of the overall network are discussed in detail in the Methods 3.2.5. The identified network ODE is dependent on the data used for SINDy regression, and are unique for the different Tasks and Rest (see Methods **??**). Overall they exhibit similar properties as they all consist of sigmoidal edges with the relation shown above and are similar to previously used BNM constructed from SC with the exception of interhemispheric inhibition due to the negative interhemipheric edge functions.

2.1.2 Dynamics

The algorithm minimizes the unstructured residual dynamics while the rest of the activity is captured by the network model. The depiction of the signal being separated in the derivative space of the spatial temporal timeseries is shown in Figure 3E and has been normalized for illustration purposes. At each timepoint from the measured ROI activity, we can compute the derivative using our ODE model of network activity and subtract from the measured derivative to separate the signal. The model captures the large features of the derivative of the fMRI signal, which can be visually observed by the network FC as as shown in Figure 3F. The subtraction between them yields an unstructured timeseries that has low correlation (0.118 mean) between brain regions as seen in the residual FC compared to those seen in the measured (0.218 mean) or network (0.216 mean) FCs. We also quantified how similar the predicted network timeseries derivative and the measured timeseries derivative on a per ROI basis in Figure 3G. The median correlation across ROIs is > 0.38 and only a few brain regions that have known to have poor signal such as the small basal ganglia structures and temporal and frontal poles have low correlations (< 0.2) with our network timeseries predictions. The ODE model is also specific to the type of data it is trained on, as in model trained on Rest fMRI is more predictive on Rest dynamics than Task dynamics (see Methods section 3.3.5). The only requirement that is needed is the amount of data needed to train the dataset (see Supplementary Figure 11). In this manner we construct separate network models for the Rest dataset and each of the four respective Task datasets. The separation illustrated here serves as a simplified demonstration, because it is not clear if the residuals contain stimulus information, or if the stimulus

information is a component of the network or a mix of both. For that we turn to the following experiment.

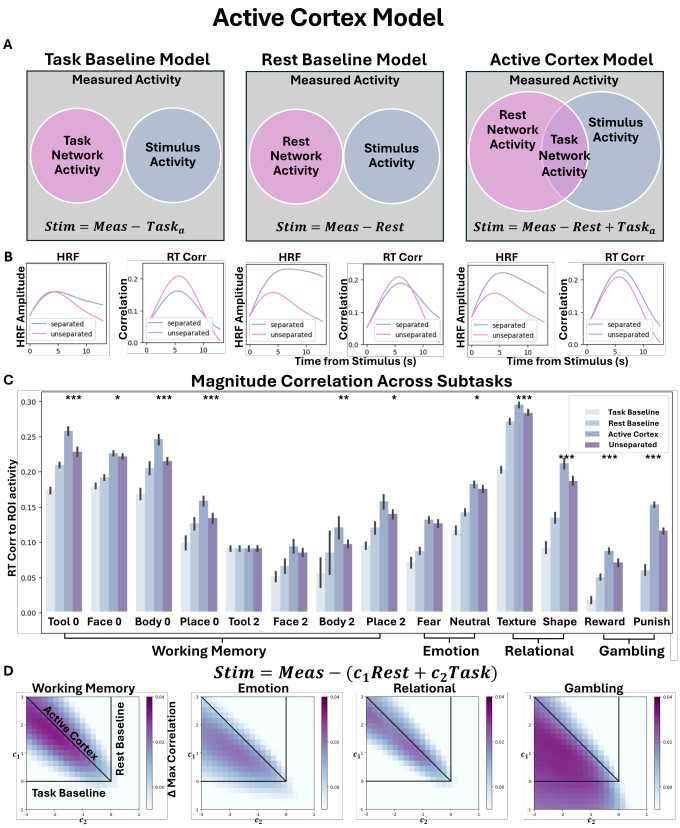

Fig. 4 A) Possible relationship frameworks between the Stimulus, the Measured signal and the fitted Rest and Task network models. The subscript a in Task, represents the Task specific model trained on that particular dataset. Left) The Task Baseline Model which assumes that the Stimulus activity is the Measured signal minus the Task network model. Middle) The Rest Baseline Model which assumes that the Stimulus activity is the Measured signal minus the Rest network model. Right) The Active Cortex Model that assumes that a certain subset of Rest, namely the Task network model is involved in the processing of the Stimulus Activity. B) The change of the average HRF of the max correlated ROI for the Working Memory and the corresponding change in correlation to RT across time for each of the three frameworks. C) The changes to RT correlates are quantified across fourteen substasks and are significant for 11 of the 14 subtasks. D) A generalized sweep of linear combination of Rest network and Task network models plotted against the change in max correlation to the signal. The diagonal represent scaled version of the ACM, while the vertical and horizontal lines corresponds to Rest and Task network models respectively. Stars represent different significance values: *** < 0.001, ** < 0.01, * < 0.05

2.2 Active Cortex Model

In order to separate the Task dependent stimulus component from the measured signal, we consider the following three plausible frameworks describing the relationship between Rest, Task and stimulus activity as shown in Figure 4A. All frameworks subtract from the measured signal the stimulus unrelated activity but estimate this activity in the following manners: 1) The Task Baseline Model, where the unrelated activity is the Task specific Network ODE, 2) The Rest Baseline Model, where the unrelated activity is the Rest Network ODE, 3) The Active Cortex Model (ACM), where the stimulus partially overlaps with the Rest Network ODE and results in the Task Network ODE. In the ACM, the stimulus unrelated activity is defined by the processes present in the Rest Network ODE but not in the Task specific Network ODE ,i.e $Rest - Task_a$. The term $Task_a$ here does not refer to a single model, but the distinct trained model using a particular Task fMRI dataset.

To determine which framework best fits the data, the three different signal separation strategies are applied for each trial and compared with the corresponding RT. Following the stimulus onset, the subsequent next 14.4 seconds are temporally aligned by resampling and separated based on the framework to obtain the stimulus dependent candidate activity for each trial (see Methods 3.3.2). In Figure 4B, the mean HRF activity for the ROI with the largest correlation to RT during the Working Memory Task, is plotted for the unseparated measured signal as well as for each of the separation frameworks. Only the ACM results in a larger HRF amplitude and a corresponding increase in correlation to RT compared to the unseparated case. In Figure 4C, the spatial temporal points that are within 90 percent of the max correlation with RT, are plotted for the unseparated and the three different strategies. ACM again is the only framework that has larger correlations after separation and shows significant increase compared to the unseparated signal for 11/14 of the subtasks. This increase in correlation is not due to a decrease in variance as the ACM actually increases the variance in the signal (see Supplemental Figure 16). While changes shown here are small as they correspond to changes in individual ROIs, in the following section we show that these changes are much larger when we aggregate all the small changes to each individual ROIs.

The association between the Rest and Task models is generalized in Figure 4D, where linear combinations between Rest and Task are considered to model the stimulus independent activity: $Stim = Meas - (c_1 Rest - c_2 Task)$ (see Methods 3.3.1. In Figure 4D, the changes to the maximum correlation value for each Task is plotted averaged across the subtasks. The horizontal line on the plot represents a scaled version of the Task Baseline Model, and the vertical line a scaled version of the Rest Baseline Model, and the diagonal a scaled version of the ACM. We clearly see in Working Memory, Emotion and Relational that the association between stimulus and the network activity is given by the ACM as seen by the diagonal streak where the correlation increases due to separation. In the case of Gambling the trend seems to be more washed out, but the largest changes to correlation is observed just off the main diagonal.

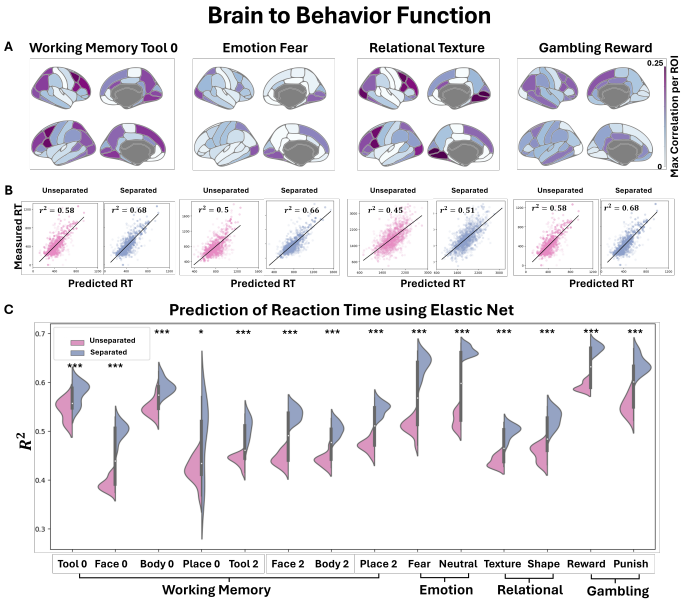

Fig. 5 A) Multiple brain regions are associated with the Task and each have moderate correlation with the RT. B) Results of utilizing a Elastic Net to aggregate the spatial temporal signal to predict the RT for both the unseparated and the separated using the ACM framework. Each point in B) represents a single trial and corresponds to the subtask above them. C) The distribution of running 10 fold cross validation using an Elastic Net on ACM separated and original unseparated data. The separated data performs significantly better in predicting RT on all subtasks, and results in an average increase of 0.09 R^2. Stars represent different significance values: *** < 0.001, ** < 0.01, * < 0.05

2.3 Brain to Behavior Function

The last section showed that by only applying the ACM, the stimulus component of the signal can be separated by showing increase in correlation to RT on a per ROI basis. In this section, the overall changes to RT are quantified using all of the spatial temporal information at once, since the correlation at each individual region increases separately. For example in Figure 5A, many ROIs are correlated to the Task and is unique to each Task being analyzed, and in Figure 8C shows that these changes also increase in time. To aggregate all of this spatial temporal information for each trial, we utilize an Elastic Net to map the spatial temporal brain activity to a single RT. Our results are shown in 5B and C using a 10 fold Cross Validated scheme on test data. The ACM improves the relationship to RT significantly after separating the signal into stimulus dependent and independent components for all fourteen subtasks, and showing an average increase of 0.09 R^2 across all subtasks. In Supplementary Figure 18, we show that these results also hold while accounting for twin relationships since HCP data, as it is known to contain large amounts of twin data. Thus the ACM, is able to explain nine percent of more variance by subtracting out the processes in the Rest ODE that are not contained in the specific Task ODE.

2.4 Relationship between Tasks

In the previous sections, one of the four meta Task models was used for the separation step, even though these meta Tasks each contained several sub-tasks. This leaves a few questions open regarding the specificity required for a Task ODE, how the different Task models compare to each other, and whether they are interchangeable in any manner. Unfortunately, there was not enough data to train a model for each of the fourteen subtasks, so instead four meta task models were trained to represent the Task activity ODE (see Methods 3.3.5).

We consider the following multitask relationship shown in Figure 6A where two overlapping tasks are shown with a partial overlap in network processes but also contain their own unique response to stimuli. By removing the unrelated network component activity from each of these specific Tasks, the unique Tasks should separate apart. This is demonstrated in Figure 6B, where the FC for each HCP participants task activity is calculated using the unseparated and ACM separated signals (see Methods 3.3.4). The differences are visualized in high dimensional space, using a UMAP 2D projection (see Methods 3.3.4). The ACM separated case yields much more localized Task localization in high dimensional space, with each of the individual Tasks separating to their own cluster. The mean Euclidean distances is shown in Figure 6 between the mean cluster centers of each Task FC to each instance of a specific Task FC. After separation, the distances within a single Task category decrease, while the distances between Task categories increase. Using a SVM classifier on the 2D UMAP data, we demonstrate that it is easier to distinguish the FCs of particular tasks after applying ACM separation than before. The confusion matrices shown in 6 D, show an accuracy of 0.42 for the unseparated case and 0.62 for the ACM separated case. The UMAP differences for the other two frameworks, Task Baseline and Rest Baseline is shown in Supplemental Figure 19 which show a decrease in separability compared to the unseparated measured signal. Although we apply a transformation that is different for each Task which could be trivial, since each Task Network Model is unique, the ACM is the only framework that separates the Task categories.

3 Methods

Our methodology is organized into the following sections. The first subsection examines the biological data as well as how they were preprocessed for this study 3.1. In the following subsection, examines the background of how previous BNMs were used to model fMRI and how they can be used to inform of biological constraints for data driven approaches 3.2.1. Subsequently, we formulate how to identify a network ODE from fMRI data by casting it as a regression problem with constraints and solve it using Sparse Identification of Nonlinear Dynamics (SINDy) 3.2.2. The subsequent subsection, examines how to validate the solutions and adjust the hyperparameters to find an optimal solution 3.2.4 3.2.5. Following those, are sections dedicated to the Task separation starting with how the signal is separated using the network models 3.3.1. Subsequently, we show how we preprocess the behavioral data and align it with fMRI signal 3.3.2. Following those are sections dedicated on how we relate the behavioral data with the brain activity 3.3.3 and on how the UMAP was applied to separate the

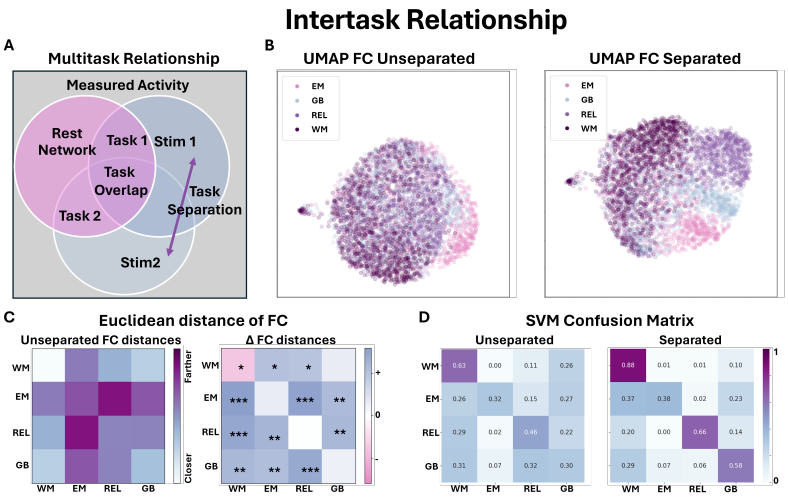

Fig. 6 A) A possible model of two distinct Tasks, where there exists an overlap in the processes being utilized but are also give rise to distinct responses for their respective different stimulus. B) A UMAP representation of the FC matrices in original unseparated space and the ACM separated space. Each point represents a single individuals task run. C) Left: The mean Euclidean distance of each individual Task FC to each cluster center, and Right: the change to these distances in the separated case. D) The classification rate using a SVM classifier on the 2D UMAP space for the unseparated and separated cases. Stars represent different significance values: *** < 0.001, ** < 0.01, * < 0.05

functional connectivity in hyperdimensional space 3.3.4. The last section examines the differences in identified Task ODE models and the identified Rest ODE model 3.3.5.

3.1 Biological Data

The data used in this study is from the Human Connectome Project [16]. HCP was chosen as it contains both Rest and Task datasets as well as the corresponding DTI for each subject. The Desikan Killiany atlas was used as the parcellation scheme which contains 68 cortical areas and 16 subcortical areas for a total of 84 regions to represent the central nervous system [17]. The Desikan Killiany atlas was utilized, since it is currently the most used atlas in the Brain Network Modeling community and allows for our results to be compared to previous results [18] [11] [11] [8] [19] [20] [21] [22] [5] [23]. All available ROIs were utilized and none were preselected. In theory, the methodology can be applied to other atlases with finer parcellations and probably would lead to overall improvements. We justify the inclusion of subcortical areas, even those brain regions that did not have great signal, as they seemed to improve the overall accuracy for the dynamic metrics in the cortex as shown in Supplemental Figure 10.

3.1.1 functional Imaging Data Processing

The HCP fMRI data was downloaded from the minimal preprocessing pipeline and had been registered to the Montreal Neurological Institute (MNI) space [24]. For the Rest data we utilized the smaller Q1 release of 447 participants using all four recordings. Since the Task data is much shorter and had only two recordings per individual, we

downloaded as much Task data that was available, and downloaded 1032 subjects for Working Memory, 1013 for Relational, 1017 for Emotion, and 1032 for Gambling. The atlas boundaries were computed with respective individual T1 space rather than using the projected data on the group MNI space. Thus we transformed the fMRI data from group space to T1 space using the process described by HCP frequently answered questions section. In T1 space we used the high resolution T1 DK atlas provided for each individual by HCP to compute the Region of Interest (ROI) boundaries. This approach of parcellating in the individual T1 space instead of the group space has been used in our previous publication and is useful for modeling fMRI for Brain Network Modeling for averaged ROI timeseries [22] [25]. We have also provided a contrast in performance in averaging the signal in group space vs individual space in Supplementary Figure 12. Our algorithm performed better in individual space than in group space in terms of reproducing the timeseries, although the coefficients estimated from using the group space were more correlated with the SC suggesting there exists a trade-off between the two approaches. The post processing was kept to a minimum as different preprocessing steps affected the identified ODE and would change the nature of the Taylor series polynomials identified. The only preprocessing steps that were applied was filtering from 0.008 to 0.1 Hz and normalizing (z-score) to the averaged ROI timeseries. The filtering step was used to minimize non-neural activity such as breathing and heart rate, instead of utilizing ICA denoising. This is due to fact that current ICA denoising has shown to reduce Task behavioral correlates and specifically cited as a reason by HCP for not including in their release for Task fMRI [26].

3.1.2 Behavioral Data

The behavioral data attached to the Task data were extracted from the behavioral tables. For the Working Memory Task these including the RT, the difficulty of the Task 0 vs. 2 back, the type of Task (Face, Place, Body, Tools), and the e-prime timings relative to the fMRI scan for each of the 80 trials in each WM scan. For the Gambling Task these included RT, Punish or Reward trial, and the e-prime timings relative to the fMRI scan for each of the 32 trials in each Gambling scan. For the Relational Task these included RT, Shape or Texture trial, and the e-prime timings relative to the fMRI scan for each of the 27 trials in each Relational scan. For the Emotional Task these included RT, Fear or Neutral trial, and the e-prime timings relative to the fMRI scan for each of the 30 trials in each Emotional scan. Since the Emotional script had a known bug during HCP recording, where only half of the last trial was recorded, we chose to truncate the last task block and thus only had 30 trials per scan.

3.1.3 Structural Diffusion Data Processing

We followed a modified version of the Hagmann procedure to extract structural connectivity matrices from DTI [27]. The diffusion data was downloaded from HCP for 442 subjects. We applied a modified form of the MRtrix HCP pipeline to estimate the tractography based on a previous publication that optimized the tractography for the HCP dataset [28] [29]. The tractography was aligned in individual T1 space to be consistent with the fMRI preprocessing. The maximum number of track lengths was set to 250, and the SC matrix was estimated by the number of tracks between

ROIs based on the DK atlas, the same T1 used for the fMRI parcellation. The final matrix connections were not normalized by volume area or any other method. Rather the raw number of tracts between ROIs was used to compare against the Taylor series coefficients.

3.2 Constructing a Network ODE using fMRI data

3.2.1 Assumptions of the underlying fMRI ODE

The general framework of network brain dynamics that has been used to describe resting state fMRI activity expresses the change in brain region a given by the vector $\dot{X}_a$ as a 1) function of the structural network SC and the current activity across all other brain regions $X_{b \neq a}$, 2) a decay to baseline $-\lambda X_a$, as well as 3) local activity which looks like noise compared to the structured network activity ηZ:

$$\dot{X}_a = -\lambda X_a + F(SC, X_{b \neq a}) + \eta Z \tag{1}$$

This framework has been extensively studied in BNM literature, by modeling F using complex functions as well as using multiple variables to express the neural mass activity X such as in the Wilson Cowan model, or the Jansen Rit Model that uses multiple excitatory and inhibitory populations to represent the dynamics of a single brain region [30] [31] [23] [8]. We cast this framework, slightly differently by approximating a function for each edge E, namely:

$$\dot{X}_a = -\lambda X_a + \sum E(X_{b \neq a}) + \eta Z \tag{2}$$

The activity X here represents the fMRI signal directly rather than underlying neural activity in BNM which is later transformed into fMRI signal via the Balloon Windkessel Model [11]. Thus we directly fit a specific network model for fMRI activity where the variables correspond directly to measurable activity. From previous literature, we know that the function F is usually modeled as a sigmoid to simulate biologically realistic responses, as it plateaus for extreme values of X while yielding a linear response that is tied to the SC for small values of X. We expect the estimated function E to be similar and also be a sigmoid response function [23] [32]. The decay rate $-\lambda$ determines how fast the system decays, which can be experimentally approximated by observing the decay during the tail end of the HRF. This constraint will later be used in the convex optimization problem which leads to more biologically plausible solutions and is not estimated directly from the data unlike the edge functions.

3.2.2 Sparse Identification of Nonlinear Dynamics

The unknown edge functions are approximated using an algorithm known as Sparse Identification of Nonlinear Dynamics (SINDy) [12]. SINDy has an freely available open-source python package that utilizes regression to solve for the coupled ODEs [12]. The algorithm identifies a sparse ODE system by writing the derivative of the state variables $\dot{X}$ as a product of a library of functions $\Theta(X)$ multiplied by the coefficients

Ξ plus some unstructured Gaussian noise ηZ:

$$\dot{X} = \Theta(X) * \Xi + \eta Z \tag{3}$$

The data matrix X contains the activity of n different brain regions over time:

$$X(t) = \begin{bmatrix} x_1(0), & x_1(1) & ...x_1(T) \\ x_2(0), & x_2(1) & ...x_2(T) \\ ... & ... & ... \\ x_n(0), & x_n(1) & ...x_n(T) \end{bmatrix} \tag{4}$$

The derivative of this matrix $\dot{X}$, can be estimated by numerical differentiation techniques for which there are many options. We utilized the spectral method but also have experimented with the smoothed finite difference and it gives relatively similar results (see Supplemental Figure 11). The library matrix $\Theta(X)$ contains functions of X, such as the constant, polynomial, or trigonometric terms that is assumed as a prior in the analysis depending on the problem. In this case, we wish to fit a Taylor Series Polynomial between each edge. Therefore we utilized a polynomial basis with no higher order interaction terms. This assumes that each pair of nodes interact with each other independently from other nodes and can be separately approximated by a function. This leads to solutions that have a first order network structure, where edges are between nodes, rather than edges polynomial between two or more nodes to a different node and is the framework that is widely used in the BNM community [8] [5] [23]. The network activity can then be represented as a sum across all edges. No symmetry condition is imposed, so the edge function between nodes A and B can be different than the function between B and A.

Thus the library function $\Theta(X)$ can be expressed as the following, and are computed for every single timepoint of X.

$$\Theta(X) = \begin{bmatrix} X(t) & X(t)^2 & X(t)^3 & ...X(t)^N \end{bmatrix} \tag{5}$$

Based on the derivative and the library terms, we can determine the coefficients Ξ from Eq.3. In essence Eq.3 reduces to regression. The problem can be reduced to the form of $AX = B$, where X represents the coefficients Ξ which we are solving for, and A is represented by the library polynomial function $\Theta(X)$ and B the observed derivative. The estimated coefficients Ξ represent a Taylor Polynomial between each edge and can be thought of an approximation of network connectivity using a polynomial basis. We expect that the edges that are determined by the algorithm are related to the physical wiring between brain regions and thus should be related to other measures of network connectivity such as through tractography using DTI. Here, we refer to the collection of edge polynomials as the Effective Connectivity (EC). Although, this term has been used in a slightly different context in the past, here it helps clarify our methodology. The EC is thought to be sparse in nature as there are few physical connections between brain regions as they are restricted by the amount volume available inside of the skull.

The regression problem can be solved using convex optimization tools that optimize the tradeoff between penalties on the reconstruction and the amount of coefficients used which is controlled by the sparsity parameter [12]. Again in the implementation

of SINDy, there are many different optimizers choices, but we utilized $L2$ regression, as $L1$ yielded very sparse solutions that did not fit the dynamics well. The residuals represents activity that is not related to the network ODE, and is assumed to be unstructured noise ηZ from Eq. 1 and is minimized by the SINDy program.

3.2.3 Setting the Diagonal Constraint

The SINDy implementation also allows us to incorporate prior knowledge of the dynamical system by fixing certain coefficients. This improves the overall accuracy of the algorithm and constraints the solutions to have realistic network connectivity. The use of constraints is also recommended from the original authors of SINDy [33]. From Eq.1, we chose to fix the decay term $-\lambda$ as a constraint which results in EC that has a network structure that is highly correlated with the measured SC obtained from tractography (see Supplementary Figure 10). In previous BNMs that model neural dynamics λ is set usually to 1 [11] [5] [23]. However, in the context of constructing a dynamical system for fMRI, the decay rate is related to how fast the HRF decays. From the Statistical Parametric Mapping (SPM) provided HRF, we can approximate the decay rate to be -0.22 [34]. Thus for our implementation, we varied the λ parameter from -0.06 to -0.4 and at each value the system is evaluated using the metrics below (see Methods **??**. The reason that λ is a constraint rather than a solvable variable, is that there are two processes that occur simultaneously from the perspective of a single ROI, the decay to baseline and the input from other regions. Therefore, fixing it as the decay term it improves the algorithm ability to solve for solutions that are biophysically plausible. Results without this constraint lead to more randomized connectivity matrices and are shown in Supplemental Figure 10. However, we can still estimate λ from the data by varying it as a hyperparameter and by maximizing the metrics below.

3.2.4 Metrics to evaluate identified ODE

To evaluate the identified ODE, we used three main criteria in determining the plausibility of the system. Firstly, we compared the predicted timeseries with the measured timeseries. As the constructed equation predicts the derivative $\dot{X}$ for every X using Equation 1 and 2, $\dot{X}_a = -\lambda X_a + \sum \Xi_n X^n_{b \neq a}$, we can at every time point directly compare the predictions against the measured derivative over time. We tested this using a hold out test dataset, where we predicted the derivative using coefficients determined using 392 subjects, and compared it against the measured derivative on unrelated 50 subjects using the Resting State dataset (see Figure 3G). Secondly, the coefficients Ξ for each polynomial order were compared to estimates of the SC to see if the resulting network of edges matches our estimates of network using different data modalities. For this we utilized the 443 DTI tractography matrices and correlated the tractography matrix with each polynomial order networks (see Figure 3C). The third metric we used, is by comparing the simulated HRF using the identified ODE to the SPM HRF function. To simulate the HRF, 100 random points from fMRI timeseries close to zero were chosen, and the average HRF trajectory was calculated across all points. The input location was varied across ROI, but only a single triangular input was used at singular ROI for each individual simulation.

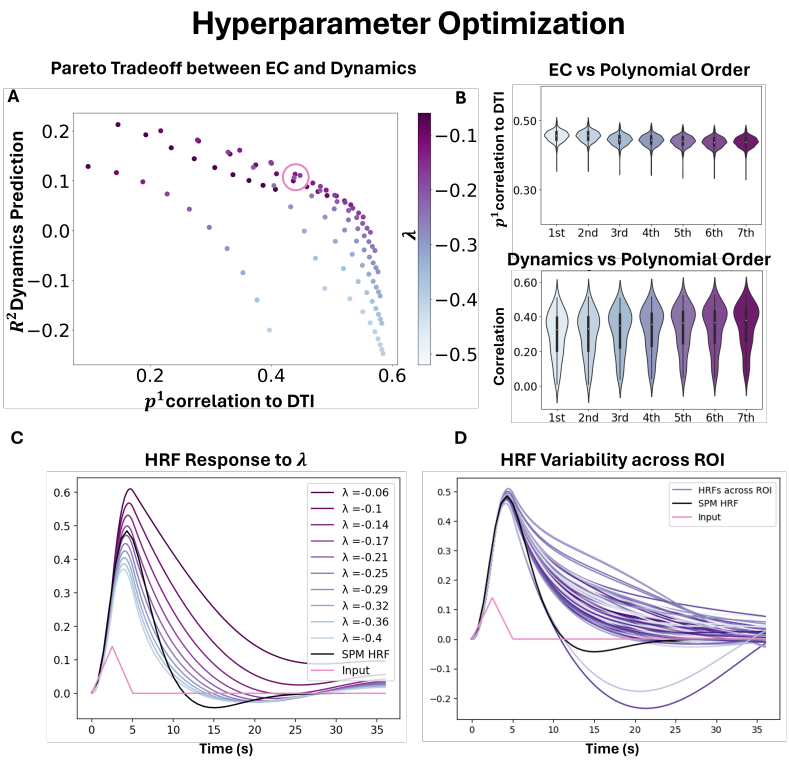

Fig. 7 A) Pareto curve illustrating the tradeoff between having a model that captures the dynamics (y-axis) vs. the solution containing more biologically plausible network similarity to measured SC (x-axis). The models are varied using different sparsity thresholds as well as varying the diagonal constraint λ which controls the decay rate of the system. B) The effect of using different orders in the Taylor expansion on SC and the dynamics metrics. The correlation to SC decrease slightly with the inclusion of higher order terms, but the reproduction of the dynamics improves with the inclusion of higher order terms. C) HRF function simulated from the Thalamus across different levels of λ, where more negative values result in a faster decay rate. The canonical SPM HRF is plotted in black and the input injected into the derivative space is plotted in pink. D) The HRF function for $\lambda = -0.22$ is plotted different ROIs showing that there is a considerable amount of HRF variability depending on the location within the network.

3.2.5 Optimization and Hyperparameters

The principle tradeoff in utilizing SINDy to characterize fMRI, is between reproducing dynamics and identifying a network that resembles the SC obtained from other data modalities. Figure 7A illustrates this tradeoff using a Pareto curve. Points on the lower right have higher sparsity with less network edges. These sparser solutions, have higher correlation to the SC, while including more edges results in a better fit in predicting the timeseries shown in the top left corner of the figure. A second hyperparameter λ controls the decay rate while also affecting the models performance and is plotted using different colors in Figure 7A. They correspond to the HRF trajectories shown in Figure 7C from a single ROI from the thalamus. The SPM HRF function plotted in black, decays at an approximate exponential decay of $\lambda = -0.22$ [34]. The corresponding

HRF for that constraint, decays however at a rate slower than λ due to the network connections providing input into the ROI. Depending on the ROI's location and its connections, it leads to variability in the HRF, as demonstrated in Figure 7C. The model that is presented in the majority of the study was chosen to have a decent fit for all three metrics and is circled in pink in 7A. The decay rate was set to -0.21, close to the SPM decay rate, and the corresponding sparsity was chosen such that the adjusted r-squared using the sklearn toolbox is larger than 0.1 and the first order EC correlation to the SC is over 0.4. To remain consistent, all SINDy models regardless of Rest and Task were all trained using the same decay rate and sparsity threshold. Additionally, multiple polynomial orders were varied, as shown in Figure 7B, up to the maximum order of seven before encountering memory limitations. The higher polynomial orders improved the fit to the dynamics without significantly altering the coefficients relation to the SC. However, it should be noted that most of the dynamics can be captured by the first-order linear approximation, which is in line with earlier works [32] [35].

3.3 Separating Task

3.3.1 Signal Separation

To test which framework best represents the Rest and Task data, we applied the following general procedure to test different linear associations between Rest and Task. We hypothesize that the stimulus independent activity can be constructed by relating the Rest and Task network equations here referred to by their EC coefficients Ξ_{Rest} and Ξ_{Task} and subsequently removed from the measured signal. Since we are operating in the derivative space, we integrate and take the initial measurement at Task onset as the initial condition to determine the separated spatial temporal signal. Therefore, we relate the RT behavioral variables to the separated signal, using c_1 and c_2 as hyperparameters for linear combinations of the Rest and Task SINDy models:

$$RT \propto \int \dot{X} - (c_1 * \Xi_{Rest} + c_2 * \Xi_{Task}) * \theta(X) \tag{6}$$

By varying c_1 and c_2, we test many different relationships between the measured signal and the ODE models. The Rest Baseline Model is given by $c_1 = 1$ and $c_2 = 0$. The Task Baseline Model is given by $c_1 = 0$ and $c_2 = 1$. The Active Cortex Model is given $c_1 =$ 1 and $c_2 =$ -1. We also swept the parameters c_1 from -1 to 3 and c_2 from -3 to 1 as shown in Figure 4D. Although this association doesn't cover all possible combinations and is limited in testing linear relationships, it allows for an initial estimation of how the cortex might be organized.

3.3.2 Behavior Processing

An overview of our Task to Behavior processing is shown in Figure 8. Each individual scan contains multiple trials that can be divided into categories of subtasks. For the Working Memory dataset, we utilized eight subtasks Place 0back, Tools 0back, Faces 0back, Body 0back, Place 2back, Tools 2back, Faces 2back, Body 2back. For the Emotion dataset, we utilized two subtasks, Fear and Neutral. For the Relational

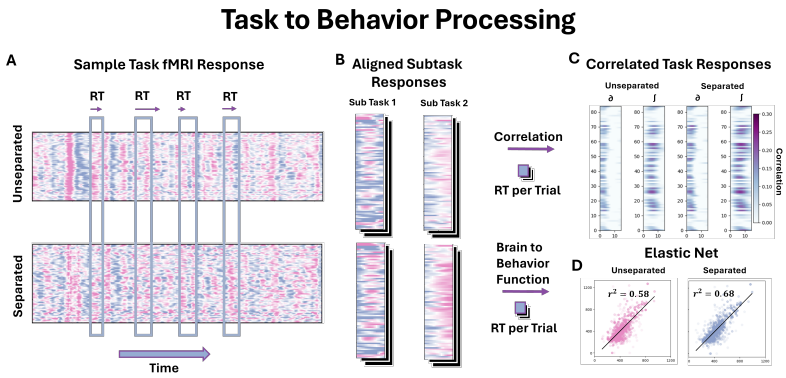

Fig. 8 A) An example of a Task fMRI spatial temporal signal showing both the unseparated and ACM separated signals. The windows represent 14.4 seconds after the task stimulus onset and for each stimulus the RT is recorded for how long the person took to respond and complete the task (displayed by differing arrow lengths). B) The windows are resampled starting from stimulus to align them on the same time frame and divided up by each sub task category. C) Each of the spatial temporal points are correlated with the RT for the task, illustrating when and where the RT variables have the highest correlation to the measured brain activity. This is performed in the derivative space as well as for the fMRI time series space for both the unseparated and ACM separated signals. D) In order to calculate the brain to behavior function, the spatial temporal signals (both the derivative and the original time series) are fed into an elastic net to predict the single RT from the resulting brain activity.

dataset, we used Shape and Texture. For the Gambling dataset, we used the Reward and Punishing and ignored the Neutral labels (single trial labels not block labels). For each trial the time of the stimulus onset was also extracted from the HCP behavioral tables as well as the RT. The RT is recorded in ms, defined as how long it took to complete each trial. Since RT is pretty generic to any Task, it was used as a measure to compare and quantify our Task separation across all Tasks. The fMRI signal was aligned to the start of the stimulus onset, and the subsequent 14.4 seconds (20 consecutive recordings) were resampled and divided by each subtask category as shown in Figure 8 B.

3.3.3 Behavior to Behavior analysis

The spatial temporal signal was compared to the singular RT for each trial in the following two manners: 1) by directly correlating the signal at each spatial temporal location on a per ROI basis across all trials (Figure 8 C) and 2) by constructing a linear model using an Elastic Net to aggregate the spatial temporal information (Figure 8 D). For the ROI analysis, we calculated the max correlation across all ROIs and time and defined a subset of all spatial temporal points that had 90 percent value of the maximum correlations. These spatial temporal points were defined on the unseparated signal, and the changes to these correlations are plotted in Figure 4 C. The Elastic Net was trained using the spatial temporal patterns to match to the spatial temporal activity to a single RT. Since the separation is actually performed in the derivative space before integrating (see section below for more details), we utilized both the derivative as well as the integrated signal as a feature vector. The Elastic Net

is implemented using sklearn toolbox and uses both a L1 and L2 penalty. The alpha and the L1 ratio was varied and optimized to be alpha=1, L1 ratio=0.95. The same values were used for all instantiations of the Elastic Net across all Tasks. We trained the model using a 10-fold cross-validation scheme, namely by dividing the dataset into a 80/20 training test ratio. For the Working Memory dataset, we also applied a more rigorous test as shown in Supplemental Figure 18, where we repeated the analysis with only Non-Twin participants. The R^2 metric that was reported was by taking the square root of the correlation between the predicted and measured RT and represents the explained variance not the coefficient of determination.

3.3.4 Applying UMAP on separated Functional Connectivity Matrices

To compute the different Functional Connectivity (FC) matrices, we gathered all the trials (14.4 seconds after stimulus onset) to each corresponding Task (Working Memory, Emotion, Gambling, and Relational) regardless of their subtask labeling on an individual level. Then a FC was calculated for each data modality, unseparated vs separated and fed into a UMAP implemented using the sklearn toolbox [14]. The parameters for the UMAP were the following: nearest neighbors set to 100, minimum distance 0.1, with 2 components. The metric to measure the distance was chosen as correlation. The latter was the most important in separating the FCs into distinct categories as seen in 6. The SVM algorithm, also implemented using sklearn toolbox, was used to gauge their separability in the UMAP space. This was implemented using the two UMAP components as inputs and using a RBF kernel to separate it into the four categories. A separate test set consisting of 20 percent of the data were then used to assess the SVM accuracy.

3.3.5 Differences between Task Models

Once establishing our hyper-parameters for the fMRI data, we trained five network ODE models to represent Rest, Working Memory, Relational, Gambling and Emotion. The amount of fMRI data needed for the SINDy algorithm to converge is shown in Figure 11. It roughly equates to around half a million timepoints. Since the Task fMRI dataset is considerably smaller this represents about all of the Relational, Emotion, and Gambling datasets and about two thirds of the Working Memory dataset. This also makes it impossible to train a model for each subtask, rather just enough data to train one meta Task model containing all subtasks within. For the Working Memory, we also designed an experiment to test the results for robustness. Since HCP contains a high percentage of twin data, we repeated the analysis without containing the twin pairs. Therefore we trained a SINDy model using roughly two-thirds of the dataset containing no twins, and tested on the other third that had no twins. However the results in Supplemental Figure 18, shows that there is no significant difference by including and excluding twins, and demonstrates that our models generalize well across healthy HCP individuals.

A comparison of the different Task and Rest models is shown in Figure 9. Each of the Task models have EC coefficients that are relatively of similar correlation to the SC. The correlation of the first order EC to the SC is for Relational correlation is

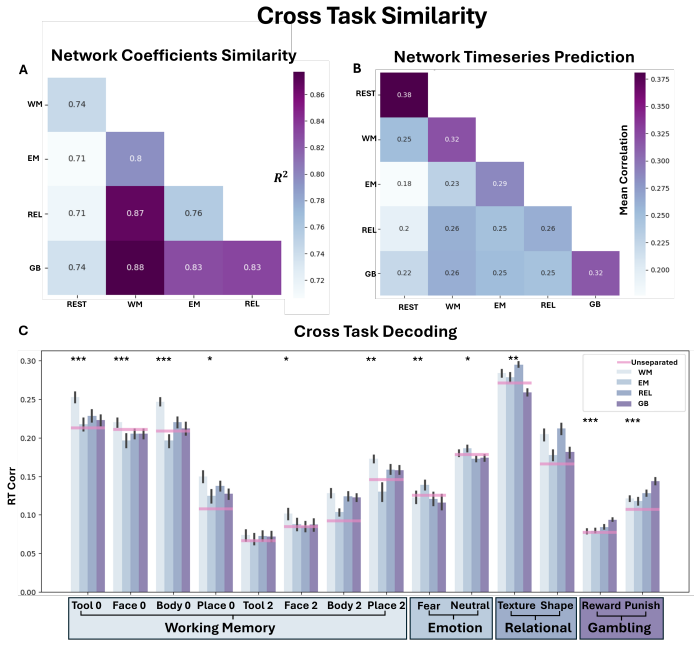

Fig. 9 A) The similarity of the EC coefficients between each of the respective network models. B) The quantification of the predicted network dynamics of the model trained using one data Rest/Task modality and used to predict the other on withheld unseen data timeseries of another Rest/Task modality. C) The quantification of using one Task model to decode the signal from another Task and the changes to the resulting behavioral correlates. For all of the subtasks, the largest increase in correlation corresponds to decoding using the same Task vs a different Task, although only 11 of the 14 these differences are significant compared to other models. Stars represent different significance values: *** < 0.001, ** < 0.01, * < 0.05

0.46, for Emotional its 0.41, for Gambling it is 0.44, for Working Memory 0.46, and for Rest it is 0.44. They are also fairly similar to each other as shown in Figure 9A where the R^2 is larger than 0.7. Interestingly the Task models EC are most similar to each other than the EC of Rest. In addition each of the models is specific to the dynamics they were trained on as the accuracy is decreased about $0.1 - 0.2$ for predicting on other dynamic processes as shown in Figure 9B. This is not a difference that arises because of testing on an unseen dataset. The accuracy is unchanged on the same Task/Rest dataset using a withheld testing dataset as can be seen in Supplementary Figure 17. This shows that the Rest model performs equivalent on unseen test dataset but not in predicting the Working Memory Task dataset. The differences in the Task models propagates in their ability to decode and separate the stimulus-dependent signal as shown in shown in Figure 9C. Applying each of the Task models to separate a different Task results in a decrease in correlation to behavioral variables. However, these differences are only significant in 11 of the 14 subtasks, suggesting an overlap between certain Task processes.

4 Discussion

In the previous sections, we demonstrated that by utilizing network models to capture fMRI dynamics, and by assuming the Active Cortex Model, we can interpret and separate the fMRI signal. The first discussion section delves into the network model and its implication. The second discussion section discusses the Active Cortex Model and its implications. Following those sections, we contextualize our findings in the current research framework of whole brain modeling and speculate on how this approach could develop in the future. Finally, we examine limitations of the approach and explore methods that might overcome these barriers to build better models in the future.

4.1 Analysis of Network Model

4.1.1 Analysis of Effective Connectivity

The Resting State network ODE identified by the SINDy algorithm exhibits both expected properties and surprising differences from previous studies. Earlier studies represented the fMRI equation as a sigmoidal function with the argument being a product of SC and the activity of each brain region. The SINDy solution affirms that the intrahemispheric connectivity does indeed have the prominent features of the SC network. The interhemispheric connections on the other hand are different than those present in the SC. The predicted interhemispheric connections are dominated by strong contralateral connections between homologous brain regions, and negative connections between non-contralateral regions. This suggests a segregation vs integration architecture where regions within a hemisphere are integrated while the two hemispheres are segregated and pushing each other apart and at the same time homologous regions between hemispheres are being coordinated together.

The intrahemispheric connections contribute to the majority of the connections of the measured SC. Their presence in the EC supports the main BNM theory that the derivative is directly proportional to the activity of its connected neighbors except in the extremes where due to biophysical constraints it is limited. Thus these processes can thereby be represented by a sigmoidal function with an argument being the product of the SC and the activity of brain regions[23] [8]. The formulation is also quite similar to connectome harmonics, where the graph Laplacian is a linear version of the identified EC [36]. Interestingly, the first order coefficients were enough to explain most of the variance in the dynamics as shown in Figure 7D, which is in line with conclusions of previous publications as well indicating that resting state dynamics is mostly linear [32] [20].

4.1.2 Analysis of predicted Dynamics

The temporal predictions of the Resting State network model in our work represents a novel metric in brain dynamics, and as there are only a few previous studies that have utilized ODEs to predict the timeseries of the fMRI spatial temporal signal. While its hard to evaluate predictions without sufficient precedence, the time series predictions seem remarkably accurate across the entire cortex. The brain regions where the algorithm performs poorly are areas that are known to have poor signals in fMRI,

namely the medial orbital frontal, frontal pole, and temporal pole. The model also had poor predictions for the smaller subcortical areas in the basal ganglia, while the predictions for the thalamus and cerebellum were fairly accurately. Moreover, from Supplementary Figure 10, the inclusion of subcortical regions in the model were key to increasing the predictability of cortical areas even if some of them had poor signal quality. The individual variance in these predictions indicates the possibility of using multiple models to represent the population, but remarkably the results do seem to generalize pretty well across the population as can be seen with no difference in accuracy between the predictions test and training sets (see Supplementary Figure 17).

There are at least two studies that have published similar metrics to these timeseries predictions. One is an combined EEG/fMRI paper, where the EEG signal is used as noise to simulate the fMRI signal using a BNM. However, this has yielded poor correlations at an average of around 0.2, which is much smaller than the model shown here with mean correlations above 0.38 [22]. A more recent paper uses a linear model to predict the fMRI signal from one ROI as a linear combination of other ROIs using the white matter SC, i.e. $X_a = SC * X_{b \neq a}$. This is not a differential equation but the current activity of one region is related to a linear combination of the current activity in other ROIs. This has yielded very accurate predictions with a large correlations mean between 0.6 and 0.7[35]. If we assume that our differential equation is in practice the $sigmoid(SC * X)$ then the integral is equivalent to $log(e^{SC*X} + 1)$ and if we assume the exponent is much larger than 1, then the equation reduces to the linear relationship which would explain both our results as well as the results of the other paper [35].

The simulation of FC via noise integration is easier to compare with previous work as it has been extensively simulated and studied [8]. Our results do not have the highest correlation in comparison to other methods, as some recent studies have correlations utilizing higher parcellations above 0.9 compared to our average correlation of about 0.7 [37]. However, as illustrated in Supplemental Figure 11, the FC metric is the least specific in analyzing dynamical systems and we believe that the predictions of the timeseries is much more informative of the performance of the model. Moreover timeseries predictions allow us to separate the fMRI signal which was crucial to test the relationship between Rest and Task.

4.1.3 Analysis of HRF response

Our predictions of the HRF are also quite novel. Although a lot of work has been done using the Balloon Windkessel model (BW), less work has been done on predictions of the HRF due to Task input from the BNM or the Neural Field Model community [38]. This is because in most of these models, a BW model is used to transform neuronal activity to a Hemodynamic response and hence by definition the HRF response of these models are designed to match the measured HRF. Here the HRF naturally arises while setting a proper value for λ. The model also predicts variability of the HRF based on its location in the network. In practice previous BNMs usually set λ to -1, but this should be revised in future models as it doesn't match the decay rate expected from experimental results which is namely -0.22 [23]. It is also interesting

to note that λ needs to be set as a constraint in the algorithm, rather than being estimated directly unlike all the other network connections. This might be explained due to the single nodal dynamics being a function of both the decay function and the network activity. Thus due to the two conflicting sources the SINDy algorithm might be unable to differentiate between them. By setting the decay value to a constant, we can estimate the rest of the EC as shown in Supplemental Figure 10. While we can predict variation in the simulated HRF, it is hard to compare with experimental predictions as it is difficult to design Task paradigms to elicit HRF responses unique to each ROI [39]. Furthermore, depending on the Task the HRF function can vary in a given brain region. Nevertheless, such comparisons should be attempted in future work as it is a good metric to narrow down the choice of models aimed at reproducing Resting State dynamics.

4.2 Active Cortex Model

Working memory and other Task fMRI paradigms have been studied extensively in literature, but whole brain models that attempt to separate the signal into components or model Task activation are rare. Even in modeling literature, the exact nature of Task input and its relationship with Rest is unclear as early DCM papers represent Task input as an additional cross coupling matrix, and later BNM models add noise in place of Task input [13] [20] [8]. However, in this study we found rather that Task would seem as a subset of Rest processes that can be to some degree linearly separated from one another. Hence, we describe whole brain dynamics as an Active Cortex Model, where during Rest all processes are present. These are probably used to maintain the neural substrate so that asynchronously the brain can respond to any given workload at any moment. During Task stimulus, a certain subset of processes are coordinated to perform a certain behavior. This switching is probably achieved using a process of stimulus expansion of Task networks or quenching of Task independent networks [1]. Moreover, the large overlap between Rest and Task would explain how Rest and Task networks have very similar FC, as many processes exist in both paradigms [15]. The above reasoning is consistent with how GLM models are currently run, in the sense that processes that are unique to the Task activate during the presentation of stimulus, and other Rest processes that always exist would not show up as statistically significant [40]. This would also explain why SINDy captures the Task only processes that are repeatedly activated when training with WM data, and not the Task independent processes that exist during Rest only and probably look like noise during Task processing.

From this understanding, we can improve how Task paradigms are designed in the future and understand the relationship between brain activity and resulting behavioral measures. When comparing our results such as the behavioral correlates to other modeling results, a single paper that utilizes DCM is probably the closest in experimental paradigm to our methodology as they use modeling to match the RT on a per trial basis [41]. This study has also used the same HCP WM dataset but they achieve a lower degree of correlation using their DCM model for 2 back (r-squared of 0.2 vs r-squared of 0.5) [41]. This substantial difference might be in part explained due to the inclusion of many more brain regions in our model which improved the elastic net's

estimation of behavioral correlates even for the unseparated case. However, it is still hard to compare the two results since they do not provide a null model of their data, and have only a single model for the Task so it is still difficult to judge our results as they seem to have no direct precedence. Compared to other estimates of behavioral correlates using HCP WM data, our results are quite promising as getting r-squared beyond 0.4 seems challenging [42]. Another recent paper uses network modeling to make inferences about WM using HCP data, but on a different scale by averaging RT metrics across WM trials rather than predicting them on a per trial basis, which makes it hard to compare against [37]. However, we have a better fit to behavioral correlates, suggesting that fitting on a per trial basis might be more appropriate in modeling WM. While we find our results promising in separating Rest from Task, further work needs to be done in fully understanding Task activation and we hope this work will act as a catalyst and facilitate more research in this direction.

4.3 Connections to other whole brain models for fMRI and its future

Our approach to solving for the fMRI equation is most similar to DCM, which was one of the first models used to describe fMRI. In fact, our equations are very similar to those first proposed by Friston et al., except rather than using a single linear model, the SINDy framework solves it using a Taylor Expansion which includes higher order terms [13]. The main bottleneck of DCM is that the method does not scale well and has thus been hampered in characterizing the dynamics of large sets of nodes. It wasn't until SINDy was proposed in 2016 that a viable method using regression existed in estimating the whole ODE for the entire network. Meanwhile, the DCM equations, were modified by the generative modeling community by incorporating estimates of the SC in order to give rise to BNM [6] [8]. These network models starting from SC were popular in generating resting state like signals by integrating candidate network equations with noise over a long period of time. Another popular approach, uses NFE that utilize a wave equation to describe the dynamics, and solves for eigenmodes based on the cortical structure and simulates resting state dynamics as a superposition of these modes [9]. The eigenvalue decomposition of the Jacobian, namely the first order terms of the EC is given in 14 shows that similar to NFE, we have global eigenmodes with large eigenvalues, and smaller non bilateral modes with smaller values, suggesting that SINDy is able to extract a similar structure from fMRI data directly. A recent paper also showed that the simulations of the NFM are similar to those of BNM [10], which is in line with the idea that the SC is mostly organized according to the principle that areas that are geodesically closer are more connected with each other [43] [35]. Therefore, the BNM and NFE are roughly equivalent. Both are able to show that the simulated signal has dynamic properties, i.e FC, multiple states, which are similar to those extracted from resting state fMRI.

It should be noted, however, neither of these methods as they are currently being constructed allow for a direct comparison between the simulated and the measured time series signals. This limitation is likely due to the unknown local neural activity modeled as noise in the simulations and thus they have relied upon metrics that summarized the signal over time and space in order to compare the simulations with the

predictions of the BNM or NFE [21]. The SINDy algorithm circumvents this problem by working in the derivative space, where from the measured signal the derivative can be predicted and can be compared against the measurements directly. Therefore, we propose that this algorithm fills a niche in modeling fMRI equations and is much more direct in obtaining brain dynamics compared to other commonly used approaches such as FC, ICA, or various other techniques developed to categorize dynamics in fMRI[19] [44]. Moreover since it is regression based, being fairly computationally inexpensive it can be used to regenerate or predict dynamics, while providing an estimate of EC. Thus, we believe it is a strong contender for the future of modeling fMRI dynamics. Moreover, unlike any other approach it is also able to help interpret and decode between Task and Rest dynamics potentially opening up a completely new perspective on how information propagates in the cortex. Therefore, we can reasonably conclude that this approach shows much promise to revolutionize the study of brain dynamics.

4.4 Limitations

The current implementation of utilizing SINDy on fMRI data has a few limitations that should be addressed in future work. The basis functions used to fit the derivative are currently limited to only a few terms up to seven due to the memory constraints of our system (64 GB DRAM). This also limits the number of ROIs that can be effectively modeled to be under 100. There seems to be already development in using sparse matrices in SINDy, that would allow for larger regression matrices and thus allow for more ROIs but it has not been officially released as part of a SINDy package [45]. Another way by which one can circumvent this memory limitation, is by using deep SINDy where one uses gradient descent to estimate the underlying equation [46]. This would consume much less memory but the trade-off would result in longer training time. The regression based SINDy, when memory requirements are satisfied, can run under 10 minutes using the whole HCP resting state data on 447 subjects. Therefore in this publication, we utilized the regression based approach but in the future deep learning approaches should be explored.

Another limitation towards using this for clinical applications and building individual models is that longer time scans are required in order for the algorithm to converge. Larger amounts of data yielded more robust estimations of the underlying ODE and improved almost all metrics. In Supplementary Figure 11, we quantify how much data was needed to achieve 90 percent of the connections when using the whole training dataset and we showed it required at least the resting state data of 100 individuals (4 runs per individual) using high quality HCP data in order to train the model. That is equivalent to roughly half a million time-points or about 100 hrs at 0.72 sec sampling rate. Thus in principle using strong gradients and priors could also be used in the future in order to reduce the training size required.

The Task separation section also assumes certain linear separability using the identified network models. This is also a limitation in the approach as nonlinear separation schemes have not been tested in this manuscript. For future studies, nonlinear separation schemes should also be considered and might provide a substantial boost in accuracy.

5 Conclusion

From our analysis, we find that using the SINDy algorithm to estimate an ODE for the resting state fMRI shows great promise as a technique for future analysis of fMRI. It allows us to estimate a model for the dynamics that can be checked against measurements and even separate Task fMRI data in relation to behavioral metrics. Moreover, it allows for modeling the dynamics using an EC and thus establishing a direct relationship between structure and function. Thus, this work represents a first of its kind by relating brain structure, activity and behavioral outcomes within a single model and therefore holds significant potential for advancing our understanding of the central nervous system.

6 Data and Code Availability

The data is available on HCP. We are unable to provide the preprocessed data due to HCP privacy rights, but we have provided the code that reproduces the resulting SINDy analysis in a Jupyter Notebook.

7 Funding

This study acknowledges support by the Virtual Research Environment at the Charité Berlin – a node of EBRAINS Health Data Cloud; Digital Europe TEF-Health 101100700, EU H2020 Virtual Brain Cloud 826421; Human Brain Project SGA3 945539, ERC Consolidator 683049; German Research Foundation SFB 1436 (project ID 425899996); SFB 1315 (project ID 327654276); SFB 936 (project ID 178316478; SFB-TRR 295 (project ID 424778381); SPP Computational Connectomics RI 2073/6-1, RI 2073/10-2, RI 2073/9-1; PHRASE Horizon EIC 101058240; Horizon Europe AISN 101057655, Horizon Europe eBRAIN-Health 101058516, Horizon Europe EBRAINS-PREP 101079717, DFG BECAUSE-Y (project ID 504745852), ERAPerMed Pattern-Cog and German Ministry of Health 2522FSB904 Berlin Institute of Health & Foundation Charité, Johanna Quandt Excellence Initiative.

8 Author contributions

AK- Preprocessing, Analysis, Code, Figures, Writing of Manuscript, EG, PB, KD, KB, SP, SK, PR- Discussion, Revision, Comments on Manuscript SP, SK, PR- Advising

No competing interests.

References

[1] Laumann, T.O., Snyder, A.Z.: Brain activity is not only for thinking. Current Opinion in Behavioral Sciences **40**, 130–136 https://doi.org/10.1016/j.cobeha.2021.04.002 . Accessed 2023-10-17

[2] Wobst, P., Wenzel, R., Kohl, M., Obrig, H., Villringer, A.: Linear aspects of changes in deoxygenated hemoglobin concentration and cytochrome oxidase oxidation during brain activation. NeuroImage **13**(3), 520–530 https://doi.org/10.1006/nimg.2000.0706

[3] Berkovich-Ohana, A., Harel, M., Hahamy, A., Arieli, A., Malach, R.: Alterations in task-induced activity and resting-state fluctuations in visual and DMN areas revealed in long-term meditators. NeuroImage **135**, 125–134 https://doi.org/10.1016/j.neuroimage.2016.04.024 . Accessed 2023-10-17

[4] Acar, F., Maumet, C., Heuten, T., Vervoort, M., Bossier, H., Seurinck, R., Moerkerke, B.: Review Paper: Reporting Practices for Task fMRI Studies. Neuroinformatics **21**(1), 221–242 (2023) https://doi.org/10.1007/s12021-022-09606-2 . Accessed 2024-07-09

[5] Ritter, P., Schirner, M., McIntosh, A.R., Jirsa, V.K.: The virtual brain integrates computational modeling and multimodal neuroimaging. Brain Connectivity **3**(2), 121–145 (2013) https://doi.org/10.1089/brain.2012.0120

[6] Jirsa, V.K.: Neural field dynamics with local and global connectivity and time delay. Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences **367**(1891), 1131–1143 (2009) https://doi.org/10.1098/rsta.2008.0260

[7] Deco, G., Jirsa, V.K., Robinson, P.A., Breakspear, M., Friston, K.: The dynamic brain: From spiking neurons to neural masses and cortical fields. PLOS Computational Biology **4**(8), 1000092 (2008) https://doi.org/10.1371/journal.pcbi.1000092 . Publisher: Public Library of Science. Accessed 2023-06-20

[8] Breakspear, M.: Dynamic models of large-scale brain activity. Nature Neuroscience **20**(3), 340–352 (2017) https://doi.org/10.1038/nn.4497

[9] Robinson, P.A., Zhao, X., Aquino, K.M., Griffiths, J.D., Sarkar, S., Mehta-Pandejee, G.: Eigenmodes of brain activity: Neural field theory predictions and comparison with experiment. NeuroImage **142**, 79–98 https://doi.org/10.1016/j.neuroimage.2016.04.050 . Accessed 2023-10-17

[10] Pang, J.C., Aquino, K.M., Oldehinkel, M., Robinson, P.A., Fulcher, B.D., Breakspear, M., Fornito, A.: Geometric constraints on human brain function. Nature **618**(7965), 566–574 (2023) https://doi.org/10.1038/s41586-023-06098-1 . Accessed 2023-06-20

[11] Cabral, J., Hugues, E., Sporns, O., Deco, G.: Role of local network oscillations in resting-state functional connectivity. NeuroImage **57**(1), 130–139 (2011) https://doi.org/10.1016/j.neuroimage.2011.04.010

[12] Brunton, S.L., Proctor, J.L., Kutz, J.N.: Discovering governing equations from

data by sparse identification of nonlinear dynamical systems. Proceedings of the National Academy of Sciences **113**(15), 3932–3937 (2016) https://doi.org/10.1073/pnas.1517384113 . Publisher: Proceedings of the National Academy of Sciences. Accessed 2023-06-20

[13] Friston, K.J., Harrison, L., Penny, W.: Dynamic causal modelling. NeuroImage **19**(4), 1273–1302 (2003) https://doi.org/10.1016/s1053-8119(03)00202-7

[14] McInnes, L., Healy, J., Saul, N., Großberger, L.: UMAP: Uniform Manifold Approximation and Projection. Journal of Open Source Software **3**(29), 861 (2018) https://doi.org/10.21105/joss.00861 . Accessed 2024-07-11

[15] Smith, S.M., Fox, P.T., Miller, K.L., Glahn, D.C., Fox, P.M., Mackay, C.E., Filippini, N., Watkins, K.E., Toro, R., Laird, A.R., Beckmann, C.F.: Correspondence of the brain's functional architecture during activation and rest. Proceedings of the National Academy of Sciences **106**(31), 13040–13045 (2009) https://doi.org/10.1073/pnas.0905267106 . Accessed 2023-06-20

[16] Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E.J., Yacoub, E., Ugurbil, K., WU-Minn HCP Consortium: The WU-minn human connectome project: an overview. NeuroImage **80**, 62–79 (2013) https://doi.org/10.1016/j.neuroimage.2013.05.041

[17] Desikan, R.S., Ségonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., Albert, M.S., Killiany, R.J.: An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. NeuroImage **31**(3), 968–980 (2006) https://doi.org/10.1016/j.neuroimage.2006.01.021

[18] Cabral, J., Hugues, E., Kringelbach, M.L., Deco, G.: Modeling the outcome of structural disconnection on resting-state functional connectivity. NeuroImage **62**(3), 1342–1353 (2012) https://doi.org/10.1016/j.neuroimage.2012.06.007

[19] Hutchison, R.M., Womelsdorf, T., Allen, E.A., Bandettini, P.A., Calhoun, V.D., Corbetta, M., Della Penna, S., Duyn, J.H., Glover, G.H., Gonzalez-Castillo, J., Handwerker, D.A., Keilholz, S., Kiviniemi, V., Leopold, D.A., Pasquale, F., Sporns, O., Walter, M., Chang, C.: Dynamic functional connectivity: Promise, issues, and interpretations. NeuroImage **80**, 360–378 https://doi.org/10.1016/j.neuroimage.2013.05.079 . Accessed 2023-10-17

[20] Kim, J.Z., Bassett, D.S.: Linear Dynamics & Control of Brain Networks (2019). https://doi.org/10.48550/arXiv.1902.03309 . http://arxiv.org/abs/1902.03309 Accessed 2023-06-20

[21] Kashyap, A., Keilholz, S.: Dynamic properties of simulated brain network models and empirical resting-state data. Network Neuroscience (Cambridge, Mass.) **3**(2), 405–426 (2019) https://doi.org/10.1162/netn_a_00070

[22] Schirner, M., McIntosh, A.R., Jirsa, V., Deco, G., Ritter, P.: Inferring multi-scale neural mechanisms with brain network modelling. eLife **7**, 28927 (2018) https://doi.org/10.7554/eLife.28927 . Publisher: eLife Sciences Publications, Ltd. Accessed 2023-06-20

[23] Sanz-Leon, P., Knock, S.A., Spiegler, A., Jirsa, V.K.: Mathematical framework for large-scale brain network modeling in the virtual brain. NeuroImage **111**, 385–430 (2015) https://doi.org/10.1016/j.neuroimage.2015.01.002

[24] Holmes, C.J., Hoge, R., Collins, L., Woods, R., Toga, A.W., Evans, A.C.: Enhancement of MR images using registration for signal averaging. Journal of Computer Assisted Tomography **22**(2), 324 (1998). Accessed 2023-06-20

[25] Dale, A.M., Fischl, B., Sereno, M.I.: Cortical surface-based analysis. i. segmentation and surface reconstruction. NeuroImage **9**(2), 179–194 (1999) https://doi.org/10.1006/nimg.1998.0395

[26] Salimi-Khorshidi, G., Douaud, G., Beckmann, C.F., Glasser, M.F., Griffanti, L., Smith, S.M.: Automatic denoising of functional MRI data: combining independent component analysis and hierarchical fusion of classifiers. NeuroImage **90**, 449–468 (2014) https://doi.org/10.1016/j.neuroimage.2013.11.046

[27] Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C.J., Wedeen, V.J., Sporns, O.: Mapping the structural core of human cerebral cortex. PLOS Biology **6**(7), 159 (2008) https://doi.org/10.1371/journal.pbio.0060159 . Publisher: Public Library of Science. Accessed 2023-06-20

[28] Tournier, J.-D., Calamante, F., Connelly, A.: MRtrix: Diffusion tractography in crossing fiber regions. International Journal of Imaging Systems and Technology **22**(1), 53–66 (2012) https://doi.org/10.1002/ima.22005 . _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/ima.22005. Accessed 2023-06-20

[29] Civier, O., Smith, R.E., Yeh, C.-H., Connelly, A., Calamante, F.: Is removal of weak connections necessary for graph-theoretical analysis of dense weighted structural connectomes from diffusion MRI? NeuroImage **194**, 68–81 (2019) https://doi.org/10.1016/j.neuroimage.2019.02.039 . Accessed 2023-06-20

[30] Wilson, H.R., Cowan, J.D.: A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. Kybernetik **13**(2), 55–80 (1973) https://doi.org/10.1007/BF00288786 . Accessed 2023-06-20

[31] Jansen, B.H., Rit, V.G.: Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. Biological Cybernetics **73**(4), 357–366 (1995) https://doi.org/10.1007/BF00199471

[32] Ashourvan, A., Pequito, S., Bertolero, M., Kim, J.Z., Bassett, D.S., Litt, B.: A

dynamical systems framework to uncover the drivers of large-scale cortical activity (2019). https://doi.org/10.1101/638718 . https://www.biorxiv.org/content/10.1101/638718v1 Accessed 2023-06-20

[33] Champion, K., Zheng, P., Aravkin, A., Brunton, S., Kutz, J.: A Unified Sparse Optimization Framework to Learn Parsimonious Physics-Informed Models From Data. IEEE Access **8**, 169259–169271 (2020) https://doi.org/10.1109/ACCESS.2020.3023625

[34] Ashburner, J., Barnes, G., Chen, C.-C., Daunizeau, J., Flandin, G., Friston, K., Gitelman, D., Glauche, V., Henson, R., Hutton, C., Jafarian, A., Kiebel, S., Kilner, J., Litvak, V., Mattout, J., Moran, R., Penny, W., Phillips, C., Razi, A., Zeidman, P.: SPM12 manual. SPM (2021)

[35] Ponce-Alvarez, A., Kringelbach, M.L., Deco, G.: Critical scaling of whole-brain resting-state dynamics. Communications Biology **6**(1), 1–11 (2023) https://doi.org/10.1038/s42003-023-05001-y . Number: 1 Publisher: Nature Publishing Group. Accessed 2023-08-16

[36] Atasoy, S., Donnelly, I., Pearson, J.: Human brain networks function in connectome-specific harmonic waves. Nature Communications **7**(1), 10340 (2016) https://doi.org/10.1038/ncomms10340 . Number: 1 Publisher: Nature Publishing Group. Accessed 2023-06-20

[37] Schirner, M., Deco, G., Ritter, P.: Learning how network structure shapes decision-making for bio-inspired computing. Nature Communications **14**(1), 2963 https://doi.org/10.1038/s41467-023-38626-y . Number: 1 Publisher: Nature Publishing Group. Accessed 2023-08-16

[38] Friston, K.J., Mechelli, A., Turner, R., Price, C.J.: Nonlinear responses in fMRI: The balloon model, volterra kernels, and other hemodynamics. NeuroImage **12**(4), 466–477 (2000) https://doi.org/10.1006/nimg.2000.0630 . Accessed 2023-06-20

[39] Aguirre, G.K., Zarahn, E., D'Esposito, M.: The variability of human, BOLD hemodynamic responses. NeuroImage **8**(4), 360–369 https://doi.org/10.1006/nimg.1998.0369 . Accessed 2023-10-17

[40] Cox, R.W.: AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. Computers and Biomedical Research **29**(3), 162–173 (1996) https://doi.org/10.1006/cbmr.1996.0014 . Accessed 2023-06-20

[41] Cai, W., Ryali, S., Pasumarthy, R., Talasila, V., Menon, V.: Dynamic causal brain circuits during working memory and their functional controllability. Nature Communications **12**(1), 3314 (2021) https://doi.org/10.1038/s41467-021-23509-x . Number: 1 Publisher: Nature Publishing Group. Accessed 2023-06-20

[42] Emch, M., Bastian, C.C., Koch, K.: Neural correlates of verbal working memory:

An fMRI meta-analysis. Frontiers in Human Neuroscience **13**, 180 (2019) https://doi.org/10.3389/fnhum.2019.00180

[43] Margulies, D.S., Ghosh, S.S., Goulas, A., Falkiewicz, M., Huntenburg, J.M., Langs, G., Bezgin, G., Eickhoff, S.B., Castellanos, F.X., Petrides, M., Jefferies, E., Smallwood, J.: Situating the default-mode network along a principal gradient of macroscale cortical organization. Proceedings of the National Academy of Sciences of the United States of America **113**(44), 12574–12579 (2016) https://doi.org/10.1073/pnas.1608282113

[44] Allen, E.A., Damaraju, E., Plis, S.M., Erhardt, E.B., Eichele, T., Calhoun, V.D.: Tracking whole-brain connectivity dynamics in the resting state. Cerebral Cortex (New York, N.Y.: 1991) **24**(3), 663–676 (2014) https://doi.org/10.1093/cercor/bhs352

[45] Sparse Matrices (scipy.sparse) — SciPy V1.11.3 Manual. https://docs.scipy.org/doc/scipy/reference/sparse.html Accessed 2023-10-17

[46] Champion, K., Lusch, B., Kutz, J.N., Brunton, S.L.: Data-driven discovery of coordinates and governing equations. Proceedings of the National Academy of Sciences **116**(45), 22445–22451 https://doi.org/10.1073/pnas.1906995116 . Publisher: Proceedings of the National Academy of Sciences. Accessed 2023-06-20

9 Supplementary information

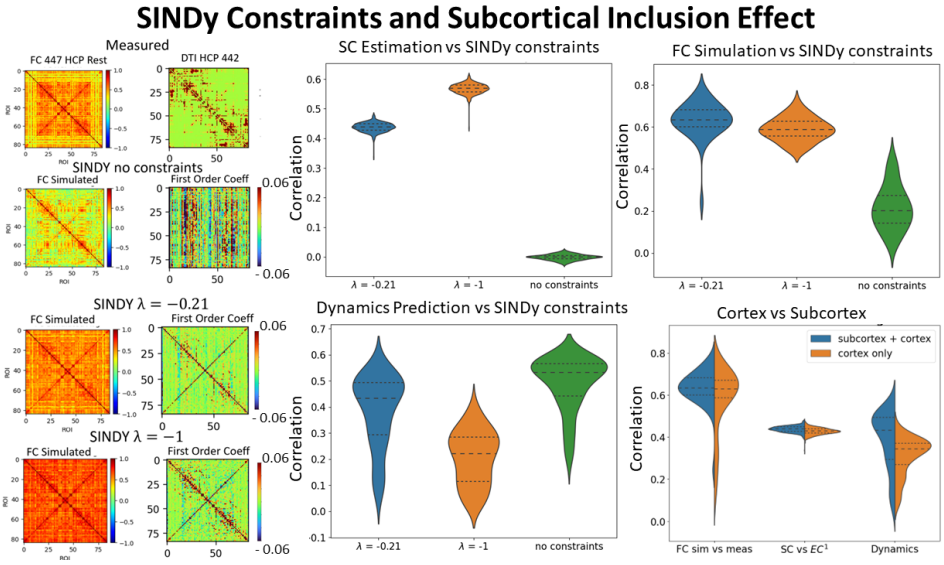

Fig. 10 Comparison between different constraint options using the SINDy algorithm. Left panel, the simulated FC and 1st Order Coefficient are provided for no constraints, constraint close to the decay rate for the measured HRF, amd a large constraint of -1. No constraint has the best fit for the dynamics but the coeffcients have no similarity to the SC, whereas the larger constraints tend to lead to solutions where the coeffcients are more similar to the SC. Close to the decay rate of the measured HRF we have a good fit for both the dynamical and SC metrics. On the bottom right, the simulated FC, estimated SC and the predicted dynamics are compared for inclusion of subcortical regions. The inclusion of the subcortex has a profound effect of improving the prediction of dynamics of other regions while the other metrics do not change too much.

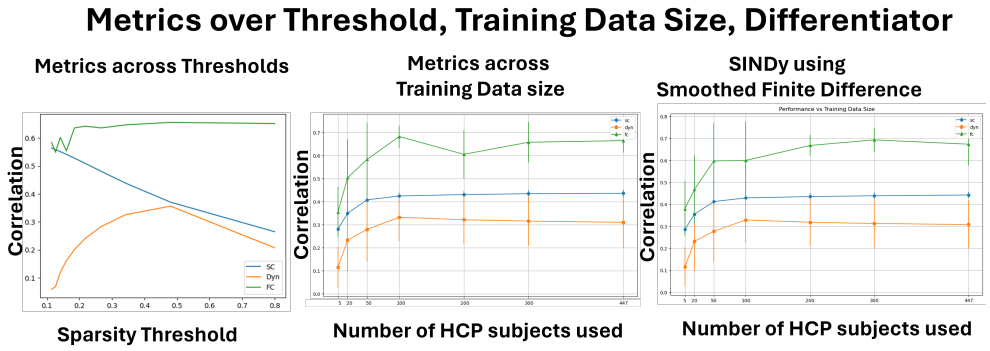

Fig. 11 Left: Looking at the performance of multiple metrics over the value of sparsity. While the correlation of the first order coefficients to the SC, and the prediction of the derivative vastly change over the range of sparsity threshold, the simulated functional connectivity barely changes across the whole range. Therefore we justified using SC and dynamics as well as the HRF as the main metrics to distinguish between the models. Center: The amount of training data needed for things to converge. Around 100 HCP subjects with 4 sessions of Rest (1200 timepoints * 4 * 100) was the minimum needed for things to converge. Right: Sample plot as center but using the smoothed finite difference differentiator instead of the spectral differentiator option in SINDy. The difference between them is not noticeable.

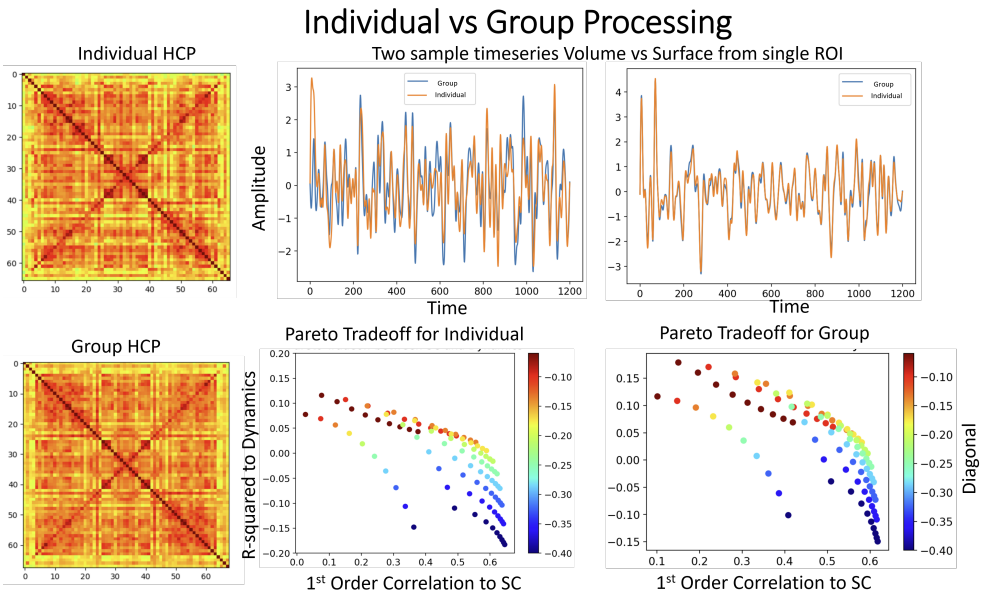

Fig. 12 Comparison of different types of preprocessing in the HCP pipeline. The Individual HCP refers to files where the ROI boundaries were determined in individual T1 space. The Group HCP processing to files that were segmented in the shared MNI group space. An example of two ROI timeseries is provided, where some ROI have few differences while others have substantial differences between the two processing, while yielding very similar functional connectivity matrices. The Group HCP processing had higher correlations of their first order coefficients to the *SC* while the Individual HCP processing had much better fit to the dynamics. Since our primary intention is to separate the signal dynamics the HCP Individual preprocessing was used. This is also more in line with previous processing used for BNM of fMDI data [22]

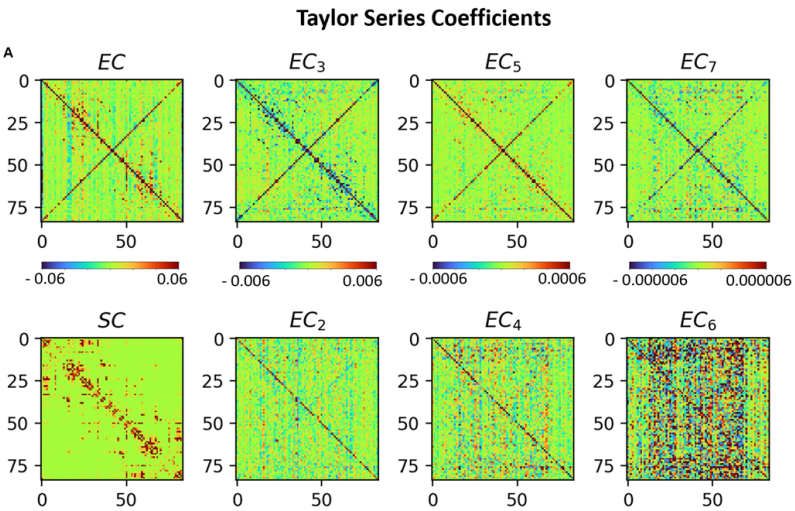

Fig. 13 A) The first seven Taylor series coefficients for $\lambda = -0.21$. Top and bottom have the same colorbar. This are the full coefficients displayed in Figure **??** and at the working biased point of the SINDy algorithm for fMRI.

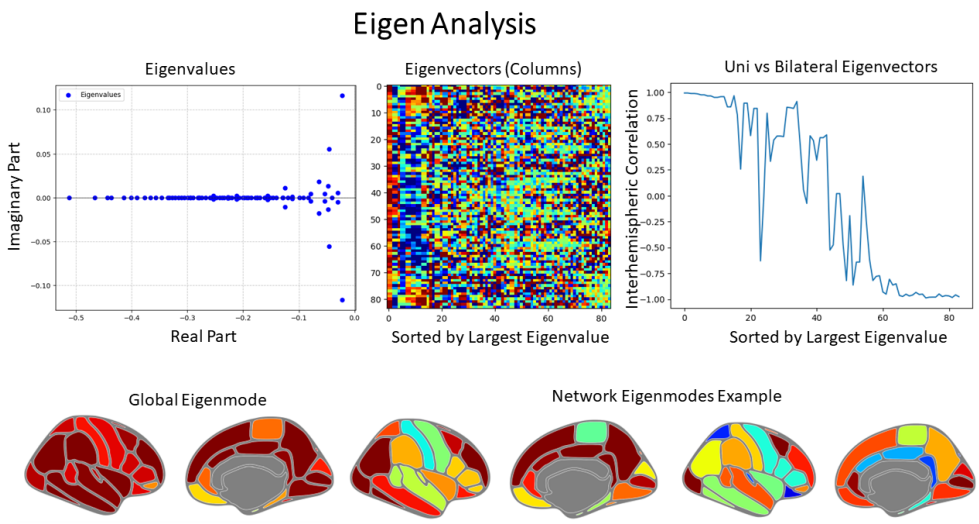

Fig. 14 Eigen Analysis of the Jacobian (first order taylor coefficients). The Eigenvalues shown in top left are plotted on the complex plane. All eigenvalues have a real value less than one, suggesting the system is stable, although some modes have real components fairly close to 0 suggesting the system is close to criticality. The corresponding real component of the eigenvectors plotted columnwise in top middle, sorted by largest eigenvalue (closest to 0). The first eigenvectors show a global mode and are highly symmetrical across the hemispheres. The smaller eigenvectors shown in top right, are mode unimodal and are even anticorrelated with the contralateral component. Bottom, the global eigenmode and some of the bilateral network eigenmodes are displayed across the cortex.

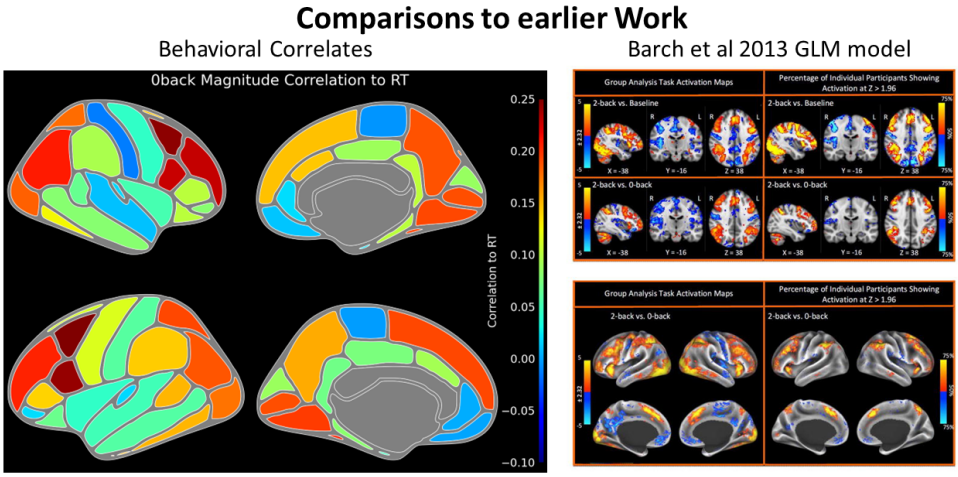

Fig. 15 Comparisons between our results spatially vs earlier GLM models on HCP WM data. Left the correlates for each brain region calculated based on RT for 0back averaged across all subtasks. Top Right: GLM model for accessing the HRF function and the significant voxels highlighted. Bottom Right: The voxels that activate for one particular Task over another.

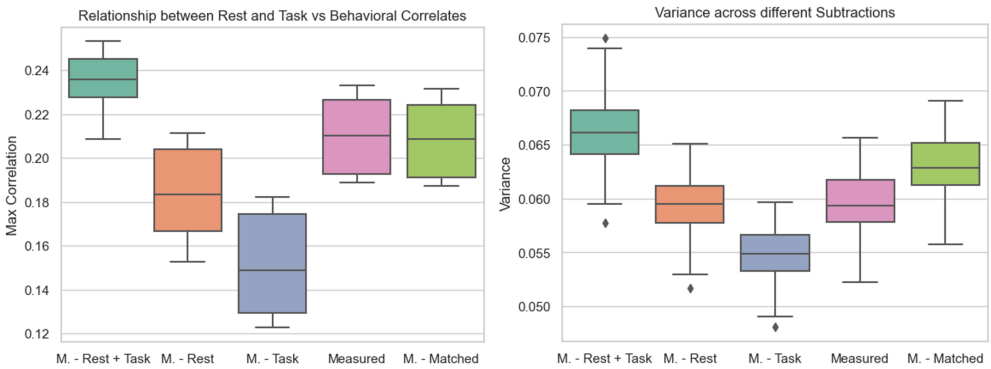

Fig. 16 The Separation of the signal using various models. M. stands for Measured. The signal separation using both task and rest ODE, the Active Cortex Model (Rest - Task), is the only separation that yields in significantly larger behavioral correlates (p ¡ 0.05) than the original Measured Signal. The matched signal represents subtracting the measured signal using the same mean and variance as the Active Cortex Model. The left Figure shows the variance significantly (p¡ 0.05) increases for the Active Cortex Model after the subtraction compared to the original and the matched model

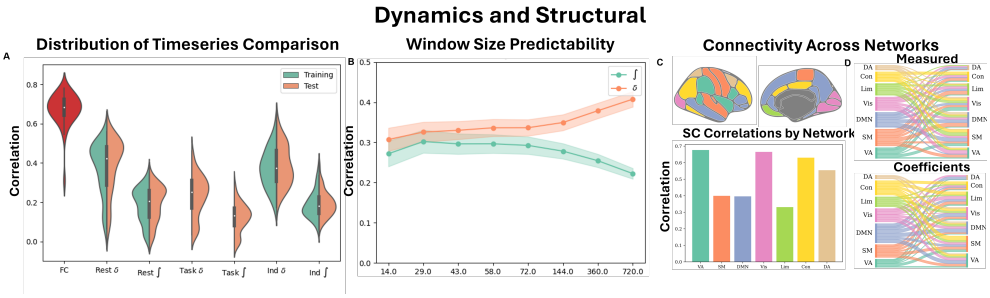

Fig. 17 A) The comparisons between many different varieties of predicted timeseries and the respective measured ones from the fitted equation. Rest and Task refer to the dataset types averaged across individuals, while Ind refers to predictions to Rest averaged across ROIs. The correlation between simulated FC and measured FC on far left. B) The average correlation computed over different time intervals for both the derivative and the original signal (denoted using the integral symbol) predicted by SINDy against the respective measured signals (averaged across ROIs). C) Top part: The seven Yeo cortical networks are modified to match the Desikan Killiany parcellation (DMN - Default Mode Network, VA -Ventral Attention, DA - Dorsal Attention, SM- Somato Motor, Con - Control, Lim - Limbic, Vis - Visual). Bottom part: The correlation of each of the networks separately computed from the estimated first order and the *SC* matrix. D) Sankey plots for the white matter *SC* and the coefficients from the first order Taylor expansion.

Fig. 18 A)Repeating the analysis of separating via ACM on a dataset that included everyone, as shown in 4 or redoing the analysis on a set of non twin HCP data. The yaxis is the correlation to RT and the xaxis the various subtasks. while there are a few significant changes they go both ways where subtasks show better performance in the non twin group and some in the Everyone group. The overall changes across the subtasks is not significant B) Observing the change in variance according to the ACM model for the two datasets. Again while some subtasks show significant differences, there exists no overall trend. The Everyone group dataset showed a mean variance increase of 0.069 and the Non-Twin dataset 0.066. Stars are different significance values: *** < 0.001, ** < 0.01, * < 0.05

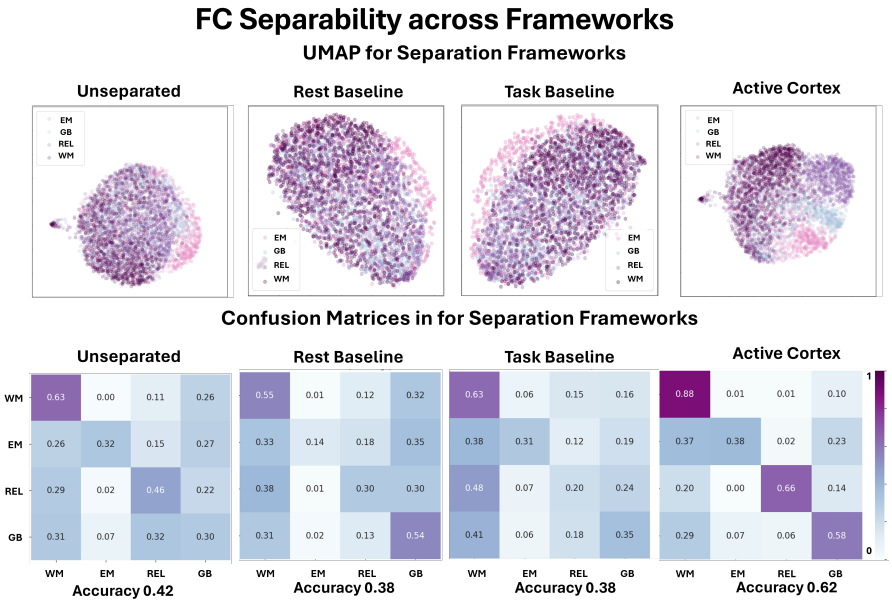

Fig. 19 A) The FCs for the different frameworks of separability. Only the Active Cortex model shows differentiation between the different Task FCs. B) The confusion matrices in attempting to separate out the data in the UMAP embedding space using a SVM model. The other baseline models perform worse than the unseparated measured signal. The active cortex model shows a twenty percent increase in accuracy of the resulting SVM model over the unseparated casees.