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

Amrit Kashyap[a], Eloy Geenjaar[c], Patrik Bey[a], Kiret Dhindsa[a], Katharina Glomb[a], Sergey Plis[d], Shella Keilholz[b,c], and Petra Ritter[a]

[a]Charite University Hospital; [b]Emory University; [c]Georgia Institute of Technology; [d]Georgia State University

This manuscript was compiled on October 23, 2023

Ongoing cortical activity is shaped by the interactions between brain regions connected on a macroscopic level. While components of the activity reflect specific processes responding to a presented stimulus, the vast majority persists as a background activity that innately exists due to cortical loops present in the connectivity. Simulations of the background activity, commonly known as resting state, have been developed over recent years using sophisticated ordinary differential equations (ODE) taking into account the macroscale organization of the cortex. However, the relationship between the rest activity compared to task or stimulus driven activity remains unclear although several different models have been proposed relating them. In this study, we develop a novel method of testing the relationship between rest and task by utilizing the framework that cortical activity can be represented by an ODE. The methodology uses a well-tested data driven approach known as Sparse Identification of Nonlinear Dynamics (SINDy) to construct an ODE to represent both rest and task functional Magnetic Response Imaging (fMRI) data separately. Since this technique is still relatively novel in the context of neuroscience, we validate the algorithm in its ability to identify an ODE that correctly predicts the signal dynamics, has a structure similar to those models currently used to recapitulate rest activity, and its ability to produce an impulse response as measured in fMRI data. Then, we systematically test the relationship between the two ODE models representing rest and task in order to model the task independent network activity. The task independent component is then subsequently removed from the measured signal and we utilize behavioral measures such as reaction time measured on a trial by trial basis to test if we have successfully separated the signal, as behavioral variables would be more correlated to the stimulus dependent activity than the raw unseparated signal. Our results show evidence that the stimulus independent signal is equivalent to all the processes in Rest that are not present in Task and can be estimated by subtracting the Task model from the Rest model. This suggests that Task can be approximated as a subset of Rest. We believe that our results are pertinent to the scientific community as they represent one of the first steps on using an ODE model to understanding the relationship of stimulus response in the cortex and allow for better decoding in terms of predicting the behavioral metrics.

fMRI modeling | Brain to Behavior | Structure to Dynamics | Activity to Behavioral Variables

The baseline activity of the cortex 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 (2). 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 brain network background activity generated by cortical loops although the exact na-

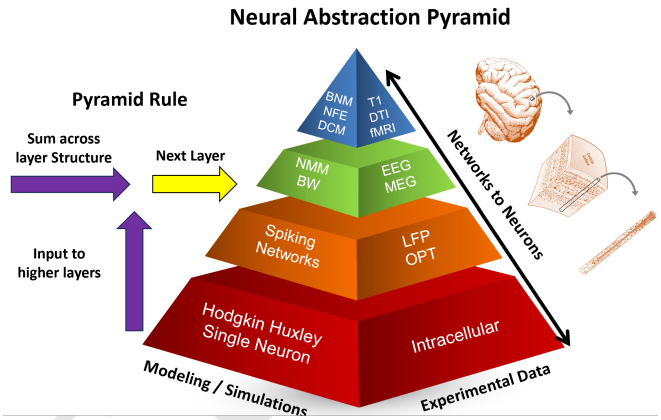

Fig. 1. The Neural Abstraction Pyramid representing different layers of neural activity from neurons to the whole network. Different measurement modalities have yielded insight to different components of the abstraction layers. In this study, we construct an fMRI specific model without considering other layers. 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)

Significance Statement

This article represents a pioneering effort in linking white matter structural connectivity, resulting in measured fMRI brain activity, and behavioral metrics within a single causal ODE model. Based solely on regression, a numerical approximation for intrinsic fMRI activity is constructed and then applied to separate aspects of the task fMRI activity based on background processes and stimulus-dependent processes. Therefore, this work represents a significant advancement towards modeling and interpreting fMRI data.

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

Corresponding Author: Amrit Kashyap E-mail: amrit.kashyap@charite.de

ture between them remains unclear. Therefore an important question that remains unanswered is the exact relationship between them and how to separate these two distinct processes from measured cortical activity such as in functional Magnetic Resonance Imaging (fMRI) (2) (3) (4). Fortunately, models of cortical activity based on macroscopic networks have progressed, to a point where we can extend them in order to answer 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. Input from the lower levels also plays a role in shaping the dynamics of the subsequent levels.

In the context of modeling fMRI, this has led to using models that utilize the white matter structural connectivity derived from Diffusion Tensor Imaging (DTI) or the shape of the surface derived from a T1 image as a structure to aggregate the activity of smaller neural elements (6) (7). This has led to Brain Network Models (BNMs) and Neural Field Equation (NFE) respectively, both of which have been able to recapitulate components of resting state activity (8) (9). Input from lower levels is usually represented by noise, as they represent unobservable variables and occur at random with the signal being modeled (10). This makes utilizing these previous frameworks to separate out the stimulus dependent and background activity difficult, as noise sources make modeling task input from lower levels challenging. Rather we utilize a technique more similar to Dynamical Causal Modeling (DCM), which uses a generative framework to derive equations directly from the data (11). This has been further developed as the Generalized Linear Model (GLM) for fMRI in order to model task input and relate specific brain voxels to task onset (12). However, unlike the DCM and GLM which do not produce timeseries simulations of whole brain cortical activity, we utilize the framework of ODE that are constructed to reproduce the timeseries dynamics like the BNM and NFE even during the task domain. Thus our approach represents a hybrid between the BNM and DCM communities which constructs equations from the data but still allows us to estimate the relationship between task and rest directly from the observational data on the timeseries domain. However, since these equations are estimated rather than derived, we need to make sure that they still represent valid biophysical models by comparing them to constructed network models.

We utilize the Sparse Identification of Nonlinear Dynamics (SINDy), a framework of approximating the underlying ODE using a Taylor Series from the observed data using regression (13). Unlike previous approaches such as DCM, this regression approach scales with more variables, and is therefore able to solve for a more generalized connectivity matrix at higher orders for the entire cortical and subcortical regions. The algorithm solves for the ODE in the derivative space, by fitting the measured derivative $\dot{X}$ as a linear combination of functions raised to higher exponents of X representing the Taylor series, $\sum \Xi_n X^n$. Once the coefficients Ξ_N are estimated, the estimated ODE is validated by the following criteria to ensure it represents a biophysical model of the network component of the fMRI equation:

- The predictions of the ODE match the derivative of the fMRI signal, since this was what the SINDy algorithm was trained to implement.

- The Taylor series coefficients Ξ_N, referred to as the Effective Connectivity (EC), are related to a function of the SC network measured through Tractography. They represent a higher order polynomial relation between ROIs and should relate to the SC in some form especially the

DRAFT

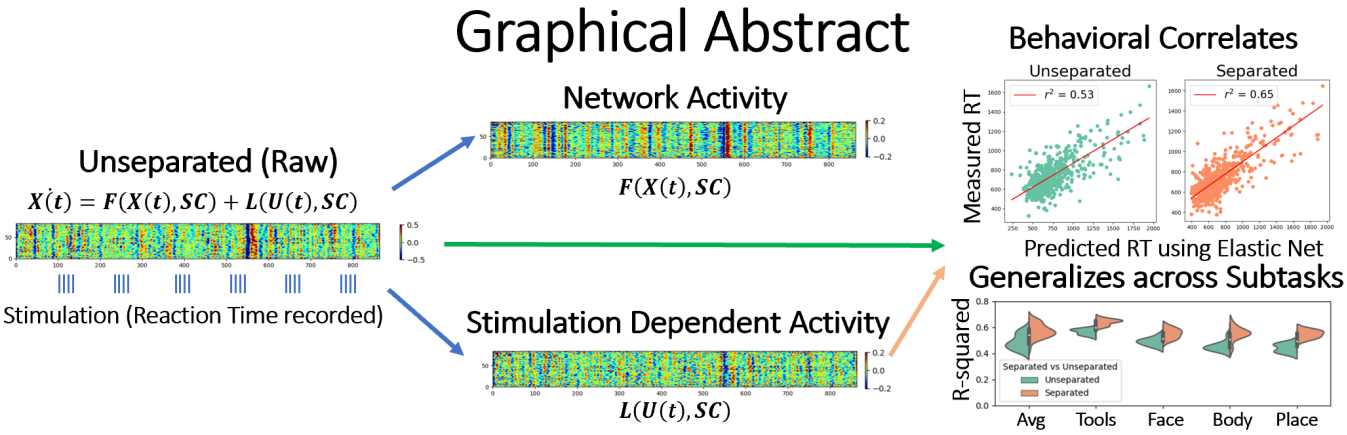

Fig. 2. Left: A scan of parcellated fMRI is recorded during the response to a set of stimulations (marked using blue lines) based on the working memory task. The dynamics of the ROIs over time X, is represented by a function of the white matter structural connectivity SC and the activity of other ROIs (Network Activity) as well as a function of the SC and the activity from external stimuli U (Stimulus Dependent Activity). Middle: The activity is separated into the Network Activity and the Stimulus Dependent Activity using our SINDy ODE framework. Right Top: The original as well as the stimulus dependent activity (first 12 sec after task onset across all ROIs) is used to construct an Elastic Net to predict the reaction time. The scatter plots show the fit between the predicted and measured reaction time response for each trial using a test dataset. Right Bottom: The 10-fold crossvalidation of the Elastic Net is displayed that is trained on subtasks for the separated and unseparated datasets. The separated signal is a better predictor for RT than the unseparated signal across all subtasks.

first order component.

- The impulse response of the ODE yields the Hemodynamic Response Function (HRF). It is known that response to a fMRI task paradigm is characterized by this function, and by ensuring we match it in some form might allow us to model the correct dynamics during Task.

By exploring the hyperparameterization space, a unique model can be identified that balances the tradeoff in all three of the criteria. The final identified resting state model contains a sigmoidal function with an argument being a product of the structural connectivity matrix, i.e. $\dot{X} = sigmoid(SC * X)$. The odd coefficients representing the Taylor series expansion have a ≥ 0.4 correlation with the measured SC from DTI while the even orders have small amplitudes and no correlation with the SC. Moreover, the estimated equation is able to make predictions on the derivative timeseries that have a correlate ≥ 0.45 with the measured signal. The resulting system also has impulse responses that have similar shapes to the HRF but also vary depending on the location of the ROI with respect to the network.

After validating that the approach is able to estimate the underlying ODE, we train two different models using rest and working memory (WM) task data. Using the assumption that these models can linearly separate task dependent and independent components, different linear combinations of the identified rest and task system were used to model stimulus independent activity during the task. This was then used to subtract from the measured signal during the task, and the subsequent separated timeseries signal was correlated with the reaction time needed to complete each WM task. In this manner, we test most simple configurations that would relate rest and task fMRI. Our results indicate that WM processes are a subset of resting state processes, and subtracting the relative complement, i.e. rest - task model from the measured signal, allows the separation of processes between stimulus dependent and independent processes. By removing these signals from the measured signal we show that we can separate the signal and show a significant increase in the correlation to the reaction time collected on a per trial basis against the original unseparated signal. This fits the notion of an Active Cortex Model, where the cortex is constantly active even during rest. These processes most likely exist to maintain the neural computational substrate while only a certain subset of those processes are coordinated to perform a single task at a given time. While more future work is necessary to explore both the relation between rest and task as well as the effectiveness of the SINDy approach in modeling fMRI, this study currently demonstrates that we are able to estimate an ODE system that is both able to approximate fMRI dynamics as well as interpret a subset of this system that gives rise to a particular task. Moreover, it naturally yields a single model combining brain structure, dynamics, and behavior and therefore we believe it has a promising future in neuroscience.

Methods

Our methodology is organized into the following sections. In the first subsection, we briefly go through a background of ODEs used to fit fMRI data and identify constraints in the context of whole brain modeling. Subsequently, we illustrate how we set up SINDy to identify the ODE from the data as a regression problem with constraints. Most of the details found here are not focused on the SINDy algorithm, but rather on how to use it in the context of identifying ODEs for fMRI data. The subsequent subsection deals with how to validate our solutions in the context of fMRI research. These metrics assess its predictive capability for the fMRI signal, its correlation with measured structural connectivity, and its ability to generate an accurate impulse response in the shape of HRF. Following those, are subsections in which we establish our experimental paradigm in order to assess the relationship between rest and task and the metrics used to evaluate that relationship using reaction time behavioral measurements. The last section goes into the biological data used in this study and the preprocessing choices for the fMRI and DTI data.

Assumptions of the underlying fMRI ODE. The general framework of large-scale brain dynamics used to describe resting state fMRI activity expresses the change in brain region a given by the vector $\dot{X}_a$ as a function of the SC and the current activity across all other brain regions $X_{b \neq a}$, a decay to baseline $-\lambda X_a$, as well as local activity which looks like noise at the highest abstraction layer ηZ:

$$Eq1. \dot{X}_a = -\lambda X_a + F(SC, X_{b \neq a}) + \eta Z$$

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 (14) (15) (16). In the SINDy framework, we are modeling X directly which is related to the measurement obtained via fMRI from all regions of interest, as it represents the only measurable variable that we have access to in fMRI studies. 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 (16) (17). 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. These will later be used as constraints in the system which leads to more biologically plausible solutions.

Sparse Identification of Nonlinear Dynamics. To identify an underlying ODE system to model fMRI activity, we utilized the freely available open-source SINDy python package (13). 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:

$$Eq2. \dot{X} = f(X) = \Theta(X) * \Xi + \eta Z$$

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

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

The derivative of this matrix $\dot{X}$, can be estimated by numerical differentiation techniques. The library matrix $\Theta(X)$

DRAFT

contains functions of X, such as the constant, polynomial, or trigonometric terms that is assumed as a prior in the analysis. In this case, we study using a polynomial basis, which allows us to express the derivative function as a Taylor series approximation:

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

The polynomial expansion coefficients Ξ represent collectively a generalized form of network connectivity at higher orders. Here we refer to it as effective connectivity (EC), although this term has been used in slightly other contexts in the past, to clarify our methodology process. It represents the relationship between two neighboring regions as a higher-order polynomial and is a generalized form of the SC in higher dimensions. The values for X and its higher orders are computed for every single timepoint and can become quite large for many brain regions and timepoints. The EC matrix Ξ is thought to be sparse in nature as it combines different linear combinations of the library function to estimate the derivative. Therefore equation 1 reduces to the form of $Ax = B$, and can be solved using convex optimization tools that place penalties on the reconstruction as well as having sparse solutions (13). We expect that these coefficients are related to the measured DTI, as from Equation 1 and 2, we have the function $f(SC, X) = \Theta(X) * \Xi$. Moreover, the activity that is not related to the network, but a function of the local activity, is simulated as noise from Eq 1. The noise function namely ηZ in Eq 2., is minimized by the SINDy program. To further reduce the effects of noise, which affects the derivative even more than the regular signal, we use spectral differentiation provided by the SINDy package. This performs a low pass filter on the derivative of the signal and reduces noise while providing more robust solutions. SINDy offers a variety of different optimizers to solve the optimization problem. We chose an L2 optimizer that has a sparsity threshold parameter that controls the number of terms that are included in the Taylor expansion.

Obtaining Biophysically Plausible Solutions. When using SINDy, researchers typically select between various candidate libraries and constraints to enhance system identification. The polynomial library is a highly robust and offers versatile options, as it can represent a truncated Taylor series approximation of the function. However, there exists a tradeoff when including higher-order terms that allow for a better approximation of the non-linear function but can quickly lead to memory limitations. This is especially true for the interaction terms that increase exponentially with larger orders. In our implementation, we vary fitting the system from 1st to 7th orders. Moreover, due to memory issues, interaction terms are left out, which is a limitation of the approach. The Taylor series expansion representing the EC can also be separately compared at each order with the measured SC and thus a polynomial basis makes it easier to compare against other measurements of the network, unlike other basis functions.

The SINDy implementation also allows us to fix the dynamical system's coefficients, which improves more plausible solutions for the ODE. From Eq 1., we chose to fix the decay term $-\lambda$ as a constraint which resulted in solutions that were highly correlated with the measured SC. The λ parameter was varied from -0.06 to -0.4 and at each value the system is evaluated using the metrics below. The corresponding λ from the model that maximizes the metrics is then compared with the estimate of λ from the SPM HRF function which has an approximate decay rate of -0.22 (18). 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 local activity which would resemble noise. Therefore, fixing it as the decay term and minimizing it across noise allows the algorithm to solve for solutions that are biophysically plausible. Results without this constraint lead to more randomized connectivity matrices and are shown in Supplemental Figure 8. However, we can still estimate it from the data based on maximizing the metrics below.

Metrics to evaluate identified ODE. To evaluate the identified ODE, we used three main criteria in determining the plausibility of the system. Firstly, the coefficients Ξ for each polynomial order were compared to estimates of the SC to see if the resulting function matches our previous understanding of functions driving the network component of the underlying brain dynamics. Secondly, 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. We also used the first timepoint as an initial condition, and integrated over the predicted derivative to directly estimate the fMRI signal and compare against our measurements. Lastly, we tested the impulse response of the ODE across different ROIs, and compared the response against the canonical Hemodynamic Response Function. To reduce dependencies on initial conditions, 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.

In addition, we computed the functional connectivity by simulating the identified ODE with noise. This was primarily done to compare against the results from previous BNMs, but the metric was not very useful in our context in biasing the SINDy model as it was not as sensitive in detecting differences between ODEs unlike other metrics such as comparing the coefficients to the measured SC or the evaluation of the system dynamics as shown and explained in more detail in Supplementary Figure 9.

Optimization and Hyperparameters. We primarily varied three parameters 1) the polynomial order, 2) the decay value, and 3) the sparsity of the coefficient matrix. Each model was then evaluated on the above mentioned criteria of matching the coefficients to SC, reproductions of the dynamics, and the HRF function. The polynomial order represented how many terms were included in the library to model the derivative and was limited by our memory configuration to seven orders. The decay value λ was varied from -0.06 to -0.4 and the range includes the prior estimate of the decay exponent of -0.22. The sparsity of the coefficient matrix determines how many terms were included in the final model and resulted in a tradeoff

DRAFT

explained in 5. Each instantiation of the SINDy model is influenced by these hyperparameters as well as the data used to train the system. To test how much data is needed to converge we ran experiments to determine the length needed as shown in Supplementary Figure 9 which was around 100,000 time points.

For the WM section, we constructed linear combinations of the two models, task and rest to model the stimulus-independent dynamics. For this, we swept the two coefficients, c_1, c_2 needed in a linear combination from the range from -3 to 3. The linear combination of the two models was then subsequently used to subtract from the measured signal to determine the stimulus dependent dynamics.

Separating Task. We utilized WM fMRI data from the Human Connectome Project, to test if the methodology could separate fMRI activity into task dependent and independent activity (HCP)(19). The original WM experiment was set up, by repeatedly asking the participant to perform a WM n-back task (either 0 or 2 back) on several different visual objects (Places, Faces, Body, Tools). The reaction time (RT) in ms, defined as how long it took to complete each task, was used as a ground truth measure of the trial performance. The fMRI signal was aligned to each stimulus onset, and the subsequent eleven seconds were resampled to align across all trials. The spatial temporal signal was compared to the singular RT for each trial in the following two manners, by directly correlating the signal at each spatial temporal location across all trials with the RT measure across all trials as well as by constructing a linear model using an Elastic Net to aggregate the spatial temporal information to predict the RT. The Elastic Net was 10-fold cross-validated, by training on 80 percent and tested on the held-out 20 percent test data. For the unseparated model, the parcellated spatial temporal task signal was used, and for the separated model the following steps were applied.

In our formulation, the goal was to separate the processes during the task fMRI into activity due to task dependent and independent activity. We hypothesize that the task independent activity can be constructed by understanding the network equation in relationship to the stimulus, and subsequently removed to separate the task dependent activity. To isolate the function $F_{task \neq rest}(X)$ might not be as simply removing our estimate $F_{rest}(X)$ from the measured data as some processes during rest and task might be shared. Therefore, using SINDy we tested a general formulation that utilized both models F_{rest} and F_{task} that was estimated by training on rest and task data respectively to estimate the Task independent dynamics. Therefore, we relate the RT behavioral variables to the separated signal, namely the measured derivative minus the task independent signal, using c_1 and c_2 as hyperparameters for linear combinations of the rest and task SINDy models:

$$RT \propto \int \dot{X} - (c_1 * \Xi_{rest} + c_2 * \Xi_{task}) * \theta(X)$$

The right side of the equation was then integrated to estimate the fMRI response of the separated signal. The separated signal then subsequently correlated with the RT at each spatial temporal location as well as constructing an elastic net to aggregate the information to to predict the RT. By varying the c_1 and c_2, we test many different relationships between the measured signal and the ODE models. For example, when c_1 is 0, and c_2 is 1, we test whether the residuals of fitting a task model using SINDy correspond to the behavioral metrics, i.e $RT = G(\dot{X} - F_{task}(X))$. When c_2 is 0, and c_1 is 1, we test if the rest corresponds to the task independent activity and by subtracting it , i,e $RT = G(\dot{X} - F_{rest}(X))$. Finally, when we have $c_1 = 1$ and $c_2 =$ -1, we test if the task independent activity is a relative complement, and can be approximated by the difference namely, $RT = G(\dot{X} - (F_{rest}(X) - F_{task}(X)))$, where some processes in rest are involved in computing the WM task and related to the reaction time. Thus this framework, allows us to test the linear relationship between the rest and task model as well as the behavioral variables. Although it 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.

Biological Data. All of the data used in this study is from the Human Connectome Project Q1 release of 447 participants (20). HCP was chosen as it contains both rest as well as 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 (21). All available ROIs were utilized and none were preselected. The reason the DK atlas was chosen instead of an atlas that had a larger number of ROIs, is that inclusion of more ROIs result in a massive increase of coefficients that need to be solved for and leads to memory issues in our system. In theory, however, the methodology can be applied to finer parcellations and probably would lead to overall improvements. The inclusion of subcortical areas, even those brain regions that did not have great signal, seemed to improve the accuracy for the dynamic metrics in the cortex as shown in Supplemental Figure 8.

functional Imaging Data Processing. The volume HCP files were downloaded for the WM and rest files. They were from the minimal preprocessing pipeline and had been registered to Montreal Neurological Institute (MNI) space (22). The files were then subsequently transformed back into the native T1 space, using the process described by HCP frequently answered questions section which was used in our previous publication as well (23). In the native space the parcellations using Free Surfer Library for the Desikan Killiany were aligned to the functional scans and the average signal for each region was extracted (24). The SINDy algorithm performed better with parcellating in the volume space than in the surface space in terms of fitting to the dynamics, although the coefficients estimated using the surface timeseries were more correlated with the SC. The tradeoff between these two different alignment schemes across individuals is shown in Supplementary Figure 10. The post processing was kept to a minimum as different preprocessing steps affected the identified ODE and would change the coefficients as well as the determined equation. 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 behavioral data attached to the WM task data was also extracted from the behavioral tables, including the reaction time, 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. The corresponding WM fMRI files were also downloaded and preprocessed as described above. In our analysis of the WM data,

DRAFT

a single model was used to train for a WM equation. However, the behavioral correlates were independently calculated between the 4 types of tasks and 2 levels of difficulty.

Structural Diffusion Data Processing. We followed a modified version of the Hagmann procedure to extract structural connectivity matrices from DTI (25). 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 (26) (27). 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 final matrix connections were not normalized by volume area or any other method. Rather the raw number of tracts between ROIs was used to correlate against the Taylor series coefficients for comparison.

Validation

To validate the candidate ODE obtained from the SINDy algorithm, we show that the estimated equation satisfies the following three criteria:

- The ODE is able to predict the derivative of the resting state ROI timeseries using the Taylor series expansion coefficients.
- The coefficients of the ODE roughly match those of a sigmoidal function based on the white matter estimates of Structural Connectivity.
- The impulse response of the ODE produces a Hemodynamic Response similar to those measured in fMRI.

While different SINDy models are able to predict some of these metrics better than others, the hyperparameter space has a smaller localized set of solutions that satisfies all three criteria. We select a model from this space and the corresponding parameters are subsequently used to construct models in order to separate WM fMRI data in the Results section.

Analysis of Dynamics. In this section, we assess how well the ODE predictions align with the measured ROI fMRI timeseries. Since the ODE was identified by applying regression on resting state data to fit the derivative, the predictions from the equation should match the measured derivative of the timeseries. An example of the comparison between measured and predicted derivatives is shown in Figure 3A over the span of a single resting state HCP scan, demonstrating that the rough alignment of peaks and troughs is accurately predicted by the equation. The correlation between these signals is quantified over different time windows in Figure 3B. By integrating the predicted derivative and using the first timepoint as the initial condition, an estimate of the original timeseries is also estimated and correlated with the measurements in Figure 3B (denoted by the $\int$ symbol). The correlation of the predicted timeseries and its derivative with the empirical observations is more similar to each other for shorter time windows. However,

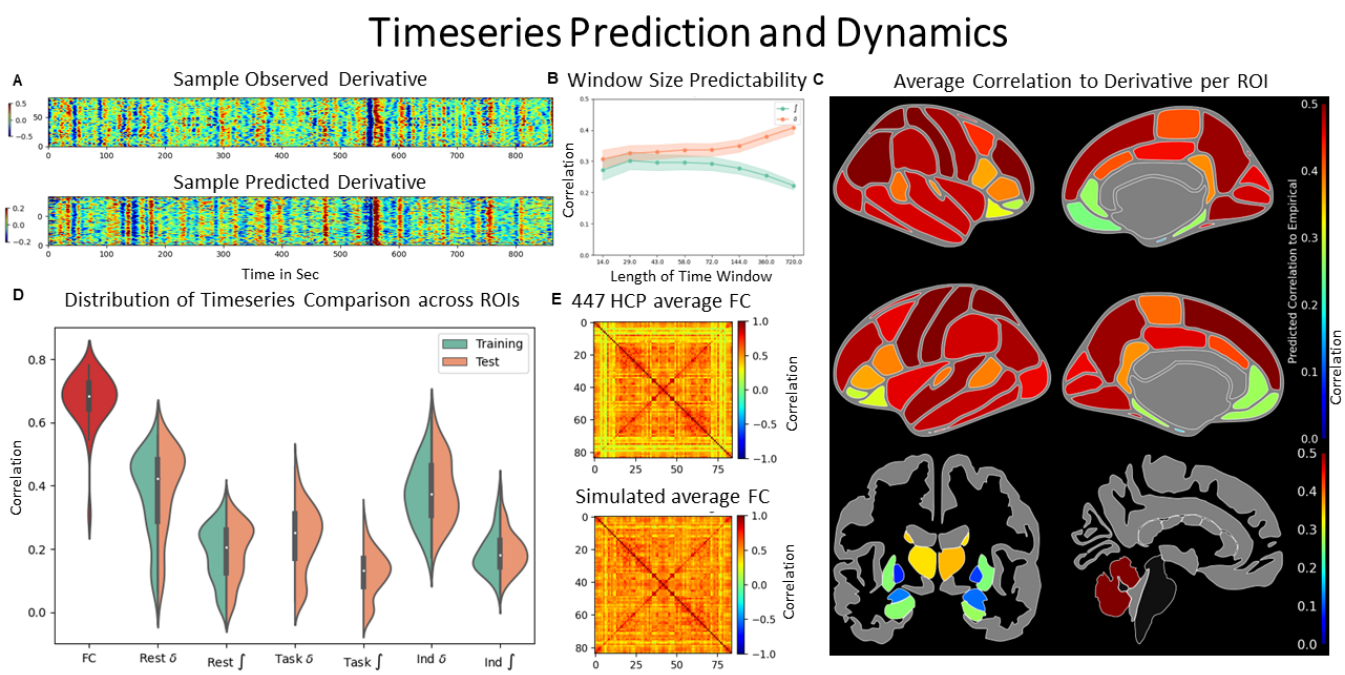

Fig. 3. A) A sample of the timeseries derivative over 20 min compared to that predicted from the fitted equation. B) The average correlation computed over different time intervals for both the derivative and the original signal (denoted using the integral symbol) predicted by SINDy against the respective measured signals (averaged across ROIs). C) The correlation of the predicted derivative timeseries to the measured timeseries computed for each ROI separately and plotted across the cortex and subcortical regions. D) 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. E) The average FC across all HCP subjects as well as the average FC computed by simulating the fitted equation with noise.

when computing over longer periods, these metrics diverge with the correlation to the predicted derivative increasing while the correlation to the predicted integral decreasing. Figure 3C illustrates the correlation to each ROI derivative computed separately. Most ROIs show very high correlations > 0.45, except for areas in the frontal and temporal lobes, as well as some deeper subcortical areas. In Figure 3D, we compare how well the dynamics are generalized across training and test datasets, and across individuals, and against unseen task data for both the derivative and the integral of the predictions. The results indicate that the function trained on several individuals generalizes to timeseries from unseen individuals. Moreover, the Gaussian distribution of correlations across the population suggests that the estimated function represents an average, where some individuals fit the estimated ODE better than others, but the majority are close to the average value. The predictions also illustrate that the model captures the dynamics for the rest fMRI dataset better than for the task fMRI, as the model was not trained on for the latter. This discrepancy is further explored in the WM separation section, suggesting that each task has is its own unique dynamical system. The leftmost panel in Figure 3D compares a simulated FC to the measured FC, by using the SINDy function and simulating resting state dynamics using noise integration, similar to older methods (8)(5). The average FC for the simulations and across HCP subjects is shown in Figure 3E. The correlation values of these FC matrices are comparable to or slightly lower than those reported in existing literature, which will be discussed in detail in subsequent sections. Overall, the ODE obtained from SINDy captures the dynamics of the resting state fMRI, as the predictions from the ODE generalize fairly well over time, different brain regions, and unseen timeseries.

Analysis of Effective Connectivity through SINDY coefficients. In this section, we focus on characterizing the coefficients of the ODE by analyzing each term of the polynomial expansion in relation to the white matter structural connectivity (SC) network. Figure 4A displays the first three terms of the EC, where each term of the coefficient matrix Ξ is multiplied by X raised to the corresponding order and summed to estimate the derivative at X(t). The full seven EC terms are shown in Supplementary Figure 11. The EC can be compared individually compared to the SC (Figure 4B) at each polynomial order as shown in Figure 4C. The correlation is calculated separately for interhemispheric and intrahemispheric connections excluding the diagonal. The intrahemispheric connections can be easily identified by the top left and bottom right corners of the structural matrices due to this particular ROI ordering that mirrors across the midline between the two hemispheres. The intrahemispheric connections exhibit a strong correlation with the measured SC for odd orders of coefficients, such as $X(t), X(t)^3, X(t)^5$, with signs alternating between consecutive odd orders. In contrast, the even orders lack a discernible network structure and have much smaller

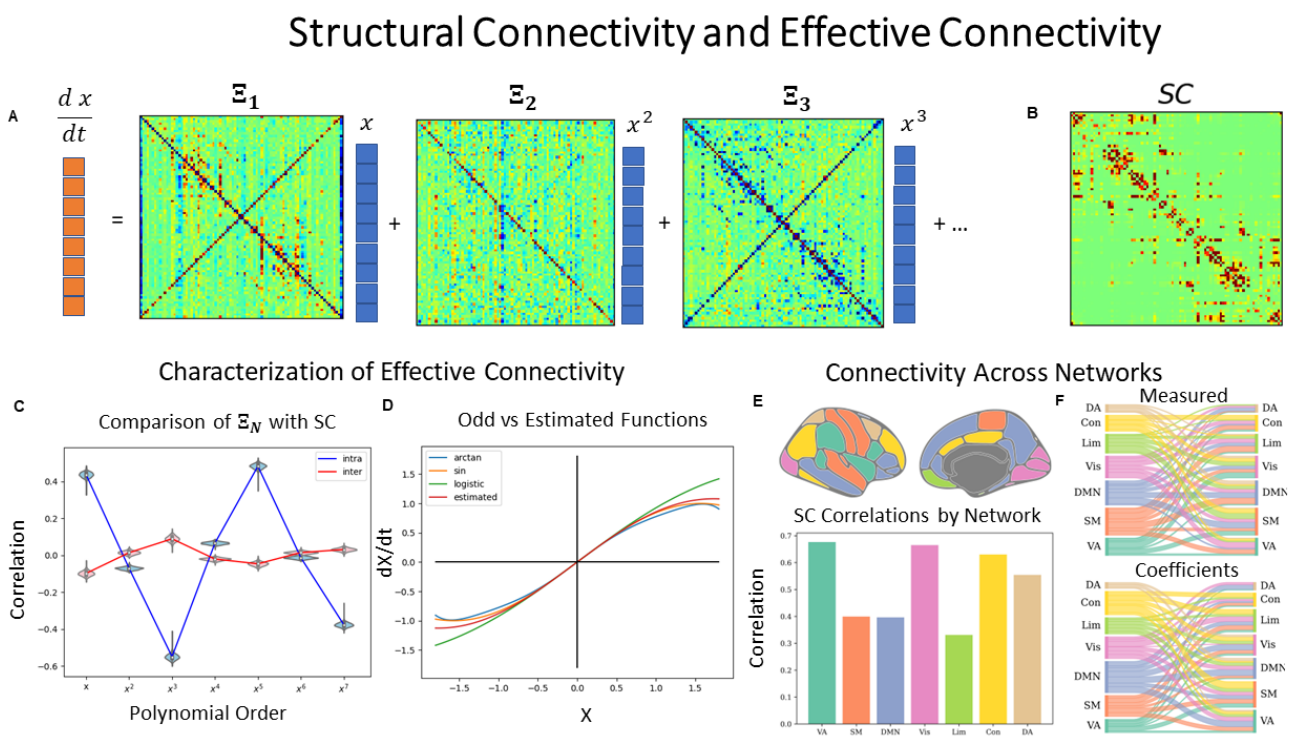

Fig. 4. A) The derivative is expressed using the first three estimated terms of the Taylor expansion. B) The average SC estimated from 443 subjects' tractography using the Desikan Killiany Atlas. C) The correlation between the interhemispheric and intrahemispheric connections of each polynomial coefficient with the SC matrix. D) An estimate of a single function of the Taylor Series expansion by taking the SC as the argument compared to other odd sigmoidal functions. E) The correlation computed over Yeo seven networks individually between the first order term and the averaged SC (DMN - Default Mode Network, VA -Ventral Attention, DA - Dorsal Attention, SM- Somato Motor, Con - Control, Lim - Limbic, Vis - Visual). E) Top part: The seven Yeo cortical networks are modified to match the Desikan Killiany parcellation. Bottom part: The correlation of each of the networks separately computed from the estimated first order and the SC matrix F) Sankey plots for the white matter SC and the coefficients from the first order Taylor expansion.

magnitudes than their odd counterparts (Supplementary Figure 11). On the other hand, interhemispheric connections are characterized by strong connections between contralateral brain regions and are represented by the antidiagonal elements in the coefficient matrices. Very few off-antidiagonal terms exist for the interhemispheric connections, and the overall pattern of connections is less similar to the SC compared to intrahemispheric connections. These differences in predicted vs. DTI measured SC are commented later on in the discussion. To gain some insight into what the overall odd function (as the even terms are small) might represent, we further reduce the function to a one-dimensional simplified function with the SC as an argument, i.e. $\dot{X} = F(SC \cdot X)$. For this, we have extracted the magnitude of coefficients at each order to see how well the SC fits into each polynomial order of the EC. This results in a Taylor series expansion of a simplified single function for F with the SC in the argument. The resulting odd function is compared to several commonly known odd functions in Figure 4D and is shown to have a similar shape to sigmoidal functions, i.e., $F(SC \cdot X) = \text{Sigmoid}(SC \cdot X)$, which is similar to our understanding of brain dynamics from previous literature and discussed later on in the discussion. Furthermore, to gain a better understanding of which networks the ODE coefficients best represent, each of the seven Yeo networks was also characterized separately, as depicted in Figure 4E (28). As seen in the previous dynamics section, the coefficients representing the Limbic network, which includes the ROIs of the orbital frontal cortex, temporal, and frontal poles, are the most distinct predictions compared to the white matter SC network. This distinction is likely due to the poor fMRI signal in these regions, as mentioned earlier. The other two networks that are notably distinct from the SC are the default mode network (DMN) and the sensorimotor (SM) network, where the fitted coefficients exhibit fewer connections compared to the measured networks, as seen in the Sankey plots in Figure 4F). In addition to these analysis, we also estimated the eigenvalues and eigenvectors using the first order coefficients of Ξ in Supplementary Figure 12. These represent the Jacobian of the system and show that the system is stable as all real components of eigenvalues are less than zero. In addition the eigenvectors with largest eigenvalues show global eigenmodes while, smaller eigenvalues correspond to network or opposing hemispherical processes.

In general, the coefficients for the other networks show higher correlations with the measured SC network. Overall, the analysis of the effective coefficients obtained from the algorithm reveals that the intrahemispheric connections exhibit a resemblance to a sigmoidal function constructed from the SC.

Tradeoff and Hemodynamic Response Function. The previous sections demonstrate that the SINDy algorithm can utilize Taylor Series expansion to estimate an ODE solution for fMRI. The resulting solution is able to correctly predict the derivative over long periods while incorporating a network resembling elements of the measured white matter structural connectivity (SC). However, this is a simplified view as there exists a range of equations that roughly resemble the dynamics of resting state fMRI, and different solutions can be obtained based on hyperparameterization. Figure 5A illustrates the primary tradeoff between reconstructing the dynamics and having more biologically plausible solutions based on the SC. Points on the lower right have higher sparsity parameters that is used in the L2- regression and have fewer terms than those on the top left. These sparser solutions in their first order, Ξ_1, have higher correlation to the white matter SC, while including more terms in the expansion results in a better fit in the derivative space. Furthermore, the parameter λ which is used as a constraint controls the rate of decay for the hemodynamic response function (HRF), as shown in Figure 5A

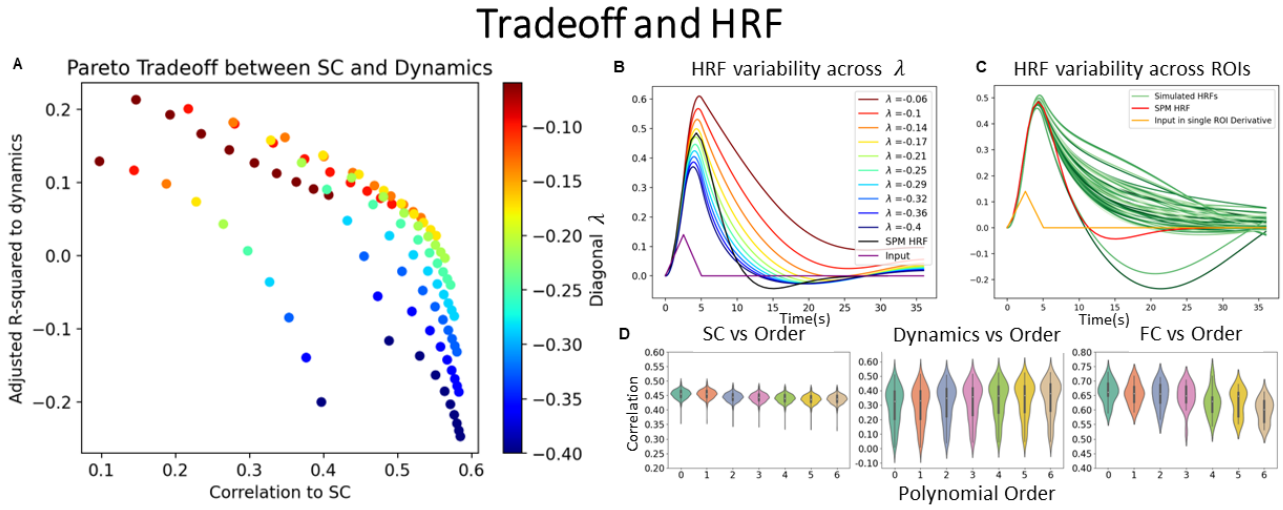

Fig. 5. A) Pareto curve illustrating the tradeoff between having a model that captures the dynamics (y-axis) vs. the solution containing more biologically plausible SC (x-axis). The models are varied using different sparsity thresholds as well as varying the diagonal constraint λ which controls the decay rate of the system. B) 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 as well as the input injected into the derivative space of the signal to elicit the simulated HRF response. C) The HRF function plotted different ROIs showing that there is a considerable amount of variability depending on the location within the network. D) SC, dynamics, and FC metrics are computed across models with different numbers of Taylor expansion terms. Both FC and SC decrease slightly when higher orders of Taylor expansion terms are included, but the dynamics improves with more coefficients included for the fitting.

using different colors. They correspond to the resulting HRF trajectories generated from the ODE solutions as shown in 5B for a single ROI in the thalamus. The SPM HRF function plotted in black, decays at an approximate exponential decay of $\lambda = 0.22$ (18). The corresponding HRF at 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 5C. The model that is presented in the majority of the study was chosen to have a decent fit for all three metrics. The decay rate was set to -0.21, close to the SPM decay rate, and the corresponding sparsity was chosen such that the r-squared to dynamics is over 0.1 and the first order correlation to the SC is over 0.4. The previous two sections show results for this hyperparameter coordinate that is also used in the subsequent sections to separate WM. Additionally, multiple polynomial orders were varied, as shown in Figure 5D, 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 (17) (29). It is important to exercise caution when applying SINDy to fMRI data, as multiple candidate models can be generated to represent whole brain dynamics. The final model was selected to optimize several metrics, including the prediction of timeseries, the relation of coefficients to the SC, and the HRF produced by the ODE.

Results

In the last few sections we validated our model for resting state fMRI, based on the trade-off on fitting three key characteristics of whole brain dynamics namely: 1) predicting the derivative 2) coefficient similarity to measured SC 3) similarity of impulse response to the HRF function. By balancing all three of these metrics, we construct an ODE that can plausibly estimate resting state dynamics. However, this is still rather quite theoretical, and in this section we show that by using this estimate of the resting state fMRI equation we can separate the signal during a known task fMRI paradigm into stimulus dependent and independent activity. Although the relationship between rest and task, has been explored by numerous studies, it remains nevertheless rather elusive and Rest and Task networks are known to be rather similar (30) (2). Most studies use a GLM based study to deduce which networks activate based on particular stimulus using statistics rather than a dynamics interpretation (12). Our goal here is to model the signal independent of the stimulus dynamics, which should be based on intrinsic processes based on cortical loops, and use it to separate the signal to task independent and dependent processes. Since the exact relationship between rest and task is not known, and some processes in rest could be used for task as well, we test a generalized relationship connecting task and rest to obtain the stimulus independent component and subtract from the measured signal:

$$RT \propto \dot{X}_{meas} - [c_1 * \Xi_{rest} + c_2 * \Xi_{task}] * \Theta(X)$$

Here the coefficients represent separate rest and task models trained on rest and task data separately. More information on the task model which has the same hyperparameters as the rest model is given in Supplementary Figure 15. The candidate separated signal is then correlated with reaction time measures at each task instance and compared against original unseparated signal to test whether the task independent signal was modeled correctly.

Separating Working Memory. The following subplots in Figure 6A-F evaluates the equation above at $c_1 = 1, c_2 = -1$, which assumes that the stimulus independent processes during are a relative complement of rest minus task which yielded the best results in improving the correlates compared to the unseparated model. Then we sweep the c_1 and c_2 parameters to show that the only solutions to the problem are when $c_1 = -c_2$.

A sample of an ROI timeseries from the Caudal Middle Frontal (CMF), is shown with the stimuli for each trial in Figure 6A with the unseparated and separated signals. For each trial, the next 15 seconds from the task onset are recorded, aligned and averaged as shown in Figure 6B. The separated signal shows an increase in the amplitude of the derivative as well as the integrated magnitude of the signal compared to the original unseparated signal. These increases in the separated signal correspond to increases in the correlations to RT for both derivative and magnitude shown in Figure 6C across the cortex. The correlation to the stimulus independent component, namely the signal that was removed is given in Supplemental Figure 15, and shows very little correlation to RT exists in the signal removed from the dynamics. To illustrate this better we plotted the CMF region for both signals, in Figure 6D to highlight how the derivative correlates to RT peak first, followed by a peak in the magnitude correlation and how the separation increases the correlation to both derivative and magnitude compared to the unseparated signal. Moreover, the regions that show high correlation in Supplemental Figure 13 match well with task activation maps generated from GLM analysis from previous literature suggesting that they are both tracking similar neural processes during task (19). While the correlations of either model are not much larger than 0.3, by fitting a linear regression model, the Elastic Net, which averages across the spatial temporal signal using both derivative and magnitude we can improve the predictions to RT. This is shown in Figure 6E, for unseen test set and shows a significant difference between ($p<0.02$) the separated and unseparated signals across all subtasks. The scatter plots in Figure 6F illustrates the relationship between the predicted reaction time from the Elastic Net and the recorded reaction time, as this is a function of all brain regions across time. These are calculated using a 10-fold cross validation on an unseen test dataset and are better than previous studies that achieved r-squared to RT using the HCP WM dataset around 0.2 for 2back compared to our r-squared over 0.5 (31). The 2back results repeat the procedure for 0back and are shown in Supplementary Figure 14.

In Figure 6G we compare different relationships between rest and task used to model the task independent signal. We try a few simple ways of modeling the task independent signal by just using the Rest or Task models. These vastly underperform compared to using the relative complement model (Rest- Task) and do significantly worse than doing nothing in the unseparated case. Lastly, in Figure 6H we compute the difference in correlation due to the model separation for a generalized form using the equation above by sweeping across coefficients c_1 and c_2 in order to separate the signal.

DRAFT

The diagonal where $c_1 = -c_2$ yields the highest changes in correlation with a maximum occurring when $c_1 = 2$. For all other configurations, the unseparated signal outperforms our candidate separated signal. Thus we conclude that up to a factor, by taking the relative complement from the Rest ODE from the Task ODE, we can model the task independent signal and separate the processes in Task fMRI into stimulus independent and stimulus dependent signals.

Discussion

In the previous sections, we demonstrated that by using SINDy, we can estimate the underlying ODE for resting state fMRI. This formulation helps us better understand the fMRI signal, which seems to be a combination of stimulus dependent and intrinsic network processes. Since both aspects are original contributions, in the discussion section, we delve into the resting state model and explore the connection between rest and task using these models. 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 the limitations of the approach and examine methods that might overcome these barriers to build better models in the future.

Analysis of Resting state Model.

Analysis of predicted Dynamics. The temporal predictions of the resting state model in our work represents a novel metric in brain dynamics, and as there are only a few previous studies that utilized BNM or NFE in order to predict the timeseries of the fMRI spatial temporal signal over time. As a result, it is hard to interpret the timeseries predictions without sufficient context. However, the predictions seem remarkably accurate across the entire cortex and using a single equation to represent fMRI whole brain dynamics. The brain regions where the algorithm performs can be also explained that these areas are known to have poor signal in fMRI, namely the medial orbital frontal, frontal pole, and temporal pole. The smaller subcortical areas in the basal ganglia also had poor predictions by the model, but on the other hand, the predictions in the thalamus and the cerebellum were fairly accurate. Moreover from Supplementary Figure 8, 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.

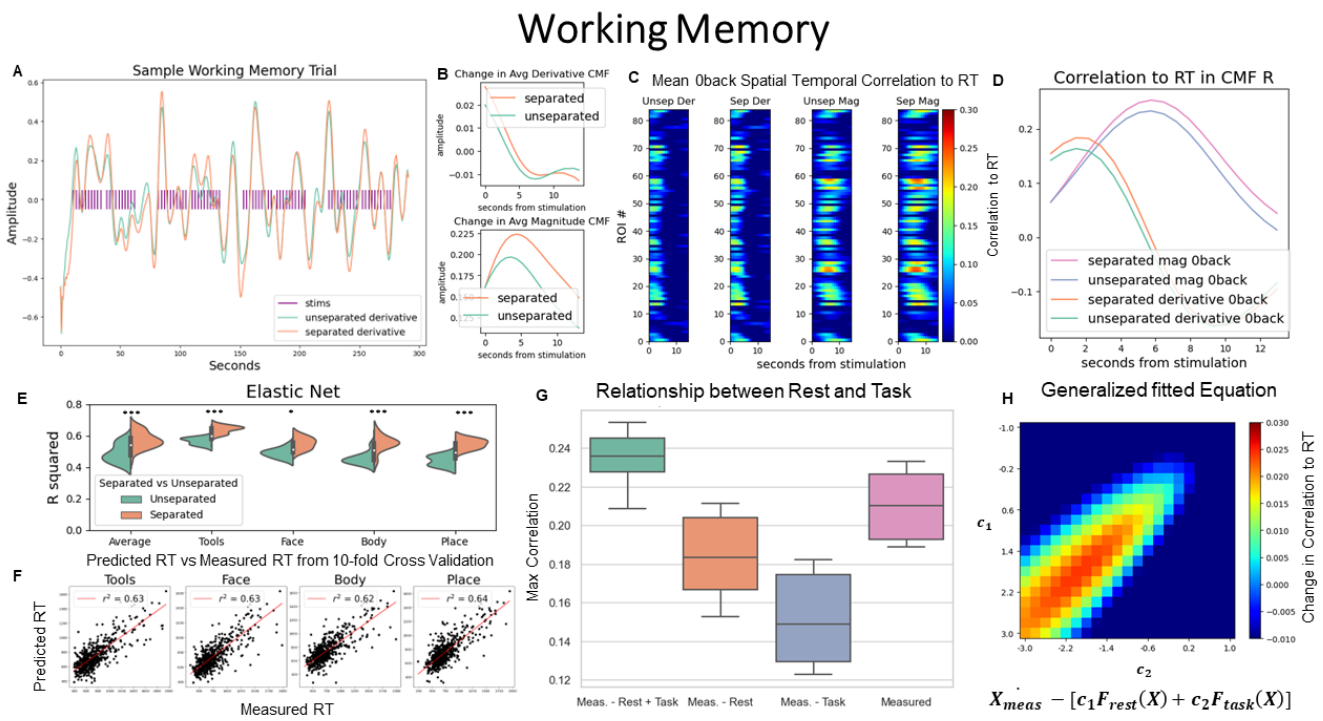

Fig. 6. A) A sample WM ROI timeseries from the CMF, with the measured derivative, the network regressed derivative and the stimulations of each WM task. B) Using the stimulation onset to align across trials, the average derivative and magnitude are calculated along with the regressed derivative and magnitude displayed for the CMF. The regressed signals have a larger derivative and magnitude, which translates into higher correlation to the reaction time collected at each trial shown in C) and for a single ROI at the CMF in D). The derivative correlates also peak earlier than the magnitude correlates seen in D). An elastic net was used to take a weighted sum of the spatial temporal information shown in C) to predict the reaction time in E). The displayed data was computed using 10 fold cross validation across all subtasks individually. In F) we show the scatter plots for the predicted vs actual reaction time from the cross validated Elastic Nets. Plots G) and H) compare using different relationships between rest and task to infer their relationship with the stimulus independent signal. G) is a special case of H) where we assume the stimulus independent signal is the Task model, the Rest model, or the Rest-Task model and remove it from the measured signal. The Rest minus Task model is the only model that significantly improves the correlation to the Behavioral function. The y-axis in this case is maximum correlation of the ROIs, specific to the 10 percent of ROIs highest correlated to the task in the unseperated framework. In H) we compare the change in behavioral correlates by assuming a generalized expression for both rest and task models to model the task independent activity. The diagonal where c1 = -c2 yields the best fit to the data, with a peak around 2.

There are at least two studies that have published similar metrics to these timeseries predictions. One is an 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 correlations above 0.45 (23). 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 has yielded very accurate predictions with a large correlations mean between 0.6 and 0.7(29). 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 in the other paper (29).

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 (32). However, as illustrated in Supplemental Figure 9, the FC metric is the least specific in analyzing dynamical systems and is rather uninformative and we believe that the prediction of the actual timeseries are much more informative of the performance of the model.

Analysis of Effective Connectivity. The resting state ODE identified by the SINDy algorithm exhibits both expected properties and surprising differences from previous studies describing the underlying ODE for fMRI dynamics. 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 rather different than the connectivity inferred through white matter tractography. The interhemispheric connections are dominated by strong contralateral connections compared to those that exist the true connectome, and very few connections between other regions across the two hemispheres. We believe that this is probably an artifact due to the omission of the cross terms in the SINDy dictionary due to memory limitations, where coordination could be driven using the contralateral region and subsequently their intra hemispheric connections. Another way that would lead to synchronization between the two hemispheres without the need for additional wiring is by simultaneously driving both hemispheres with identical sensory input. This would cause similar activation in both hemispheres, and cause the algorithm to overestimate the connections between them. Thus further research is needed to understand the differences between the SINDy estimated EC and the white matter measured SC. The two features captured in the SINDy model that were most similar to our earlier understanding of macroscale dynamics were the intrahemispheric connections and the sigmoidal activation function. The intrahemispheric connections contribute to the majority of the connections of the connectome. 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 this processes can thereby be represented by a sigmoidal function with an argument being the product of the SC and the activity of brain regions(16) (8). The formulation is also quite similar to connectome harmonics, where the graph Laplacian is a linear version of the identified EC (33). Interestingly, the first order coefficients also model captured were enough to explain most of the variance in the dynamics as shown in Figure 5D, which is in line with conclusions of previous publications as well (17) (34).

Analysis of HRF impulse response. Our predictions of the impulse response and its comparison to the HRF function are also rather novel. Although a lot of work has been done using the Balloon Windkessel model (BW), there has been less work done on predictions of the impulse response due to task input from the BNM or the Neural Field Model community (35). This is because in most of these models, a BW model is used to transform neuronal activity to BOLD response and hence by definition the impulse response of these models are designed to match the HRF. Here the impulse response naturally arises while setting a proper value for λ. The model also predicts variability of the HRF based on its location of in the network. In practice 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 (16). It is also interesting to note that λ needs to be set as a constraint in the algorithm and 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 local input from task processing the latter which SINDy assumes to be noise. Thus due to the two conflicting sources the SINDy algorithm might be unable to differentiate them. By setting the decay value to a constant and assuming that the local input will tend to seem like Gaussian noise over long periods of time, we can estimate the rest of the EC as shown in 8. While we can predict variation in the simulated HRF, it is difficult to compare with experimental predictions as it is difficult to design task paradigms to elicit HRF responses unique to each ROI (36). 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.

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. Early DCM papers representing task input as an additional cross coupling matrix, and later BNM models using noise addition in place of task input (11) (34) (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 (7), 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 activation a certain subset of processes are coordinated to perform the given task. This switching is probably achieved

DRAFT

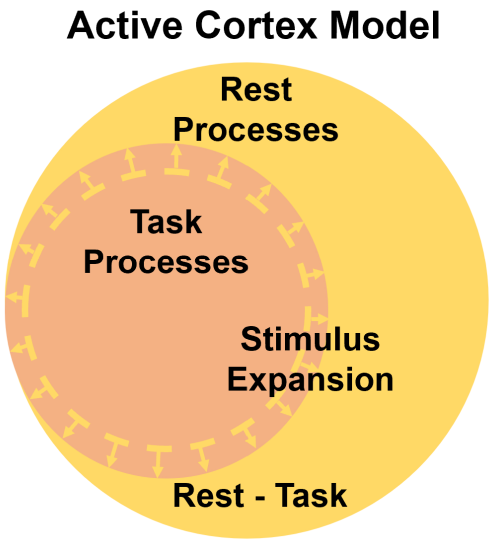

Fig. 7. The Active Cortex Model describes the relationship between Rest and Task, where Task is a subset of Rest. This can be understood that all of the subnetworks are active during task, as spontaneous input drives the cortical loops, hence the Active Cortex. During task, the response of certain subnetworks are elevated due to the input, a stimulus expansion, while subprocesses not associated with the current task are still active although smaller amount. Using the SINDy algorithm we estimate the rest processes training on rest data, and when the task is active, the strongest features of the system namely the task processes are captured by the second SINDy model. Then estimating the relative complement, rest - task models, we can estimate the dynamics independent of task. These can then be removed from the measured signal to isolate the task processes and enhance comparison with behavioral estimates.

using a processes of stimulus expansion of task networks or stimulus quenching of task independent networks (2). In order to separate this specific task from the unseparated measured signal, we can isolate the processes that are independent of the task by taking the relative complement of the rest and task sets using subtraction, and finally remove them from the measured signal to isolate task processes. The resulting separated signal are shown to be more correlated with our behavioral measures across a variety of subtasks and complexities. Moreover, the large overlap between rest and task would explain how rest and task networks have very similar FC, as many processes common to exist in both paradigms (30). 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 (12). 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 increase 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 mach the reaction time per trail (31). 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) (31). This significant 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 (37). Another recent paper that uses network modeling to make inferences about WM using HCP data, but instead looks at average RT metrics across WM trials rather than predicting them on a per trail basis (32). Our results compared to this paper are contradictory, as it claims that there are fast and slow learners and proceeds to separate the HCP population into 5 distinct groups. Here we have much better fits to the actual behavioral correlates using a single equation to model the entire population. This could be due to the fact that the paper makes a lot of assumptions on the structure of what network model should be used to model task fMRI and that these results actually don't translate to the actual data. The short comings of these findings illustrate that it is much more important to focus on the transient signals especially when using these network models that are differential equations rather than take average metrics over time (32). 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.

Connections to other whole brain models for fMRI and its future. Our approach to solving for the fMRI equation is most similar to one of the first approaches used to characterize fMRI, namely DCM in 2003. 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 (11). 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 equations proposed in Friston, 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 12 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 pull a similar structure from fMRI data directly. A recent paper also showed that the simulations of the NFM are similar to those of BNM (38), 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

DRAFT

connected with each other (39) (29). Therefore, the BNM and NFM are roughly equivalent. Both are able to show that the simulated signal had dynamic properties, i.e FC, multiple states, which are similar to those extracted from resting state fMRI. However, importantly both methods did not allow for a direct comparison between the simulated and the measured signal. This limitation was due to the unknown local neural activity modeled as noise in the simulations and thus they relied upon metrics that summarized the signal over time and space in order to compare the simulations with the predictions of the BNM or NFM (40). 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(41) (42). 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 resting state 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.

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. Moreover, the full expansion including cross terms were not included as well in this analysis due to memory constraints of the algorithm. 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 and the inclusion of higher order cross terms (43). 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 (44). 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 in under 10 minutes using the whole HCP resting state data on 447 subjects.

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 9, 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 timepoints 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.

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.

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.

Funding

PR 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.

References

1. VB Mountcastle, The columnar organization of the neocortex. *Brain: A J. Neurol.* **120 (Pt 4)**, 701–722 (year?).
2. TO Laumann, AZ Snyder, Brain activity is not only for thinking. *Curr. Opin. Behav. Sci.* **40**, 130–136 (year?).
3. P Wobst, R Wenzel, M Kohl, H Obrig, A Villringer, Linear aspects of changes in deoxygenated hemoglobin concentration and cytochrome oxidase oxidation during brain activation. *NeuroImage* **13**, 520–530 (year?).
4. A Berkovich-Ohana, M Harel, A Hahamy, A Arieli, R Malach, Alterations in task-induced activity and resting-state fluctuations in visual and DMN areas revealed in long-term meditators. *NeuroImage* **135**, 125–134 (year?).
5. P Ritter, M Schirner, AR McIntosh, VK Jirsa, The virtual brain integrates computational modeling and multimodal neuroimaging. *Brain Connect.* **3**, 121–145 (2013).
6. VK Jirsa, Neural field dynamics with local and global connectivity and time delay. *Philos. Transactions. Ser. A, Math. Phys. Eng. Sci.* **367**, 1131–1143 (2009).
7. G Deco, VK Jirsa, PA Robinson, M Breakspear, K Friston, The dynamic brain: From spiking neurons to neural masses and cortical fields. *PLOS Comput. Biol.* **4**, e1000092 (2008) Publisher: Public Library of Science.
8. M Breakspear, Dynamic models of large-scale brain activity. *Nat. Neurosci.* **20**, 340–352 (2017).
9. PA Robinson, et al., Eigenmodes of brain activity: Neural field theory predictions and comparison with experiment. *NeuroImage* **142**, 79–98 (year?).
10. J Cabral, E Hugues, O Sporns, G Deco, Role of local network oscillations in resting-state functional connectivity. *NeuroImage* **57**, 130–139 (2011).
11. KJ Friston, L Harrison, W Penny, Dynamic causal modelling. *NeuroImage* **19**, 1273–1302 (2003).
12. RW Cox, AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Comput. Biomed. Res.* **29**, 162–173 (1996).
13. SL Brunton, JL Proctor, JN Kutz, Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proc. Natl. Acad. Sci.* **113**, 3932–3937 (2016) Publisher: Proceedings of the National Academy of Sciences.

DRAFT

14. HR Wilson, JD Cowan, A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik* **13**, 55–80 (1973).
15. BH Jansen, VG Rit, Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. *Biol. Cybern.* **73**, 357–366 (1995).
16. P Sanz-Leon, SA Knock, A Spiegler, VK Jirsa, Mathematical framework for large-scale brain network modeling in the virtual brain. *NeuroImage* **111**, 385–430 (2015).
17. A Ashourvan, et al., A dynamical systems framework to uncover the drivers of large-scale cortical activity (2019).
18. J Ashburner, et al., SPM12 manual. *SPM* (2021).
19. DM Barch, et al., Function in the human connectome: Task-fMRI and individual differences in behavior. *NeuroImage* **80**, 169–189 (2013).
20. DC Van Essen, et al., The WU-minn human connectome project: an overview. *NeuroImage* **80**, 62–79 (2013).
21. RS Desikan, et al., An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage* **31**, 968–980 (2006).
22. CJ Holmes, et al., Enhancement of MR images using registration for signal averaging. *J. Comput. Assist. Tomogr.* **22**, 324 (1998).
23. M Schirner, AR McIntosh, V Jirsa, G Deco, P Ritter, Inferring multi-scale neural mechanisms with brain network modelling. *eLife* **7**, e28927 (2018) Publisher: eLife Sciences Publications, Ltd.
24. AM Dale, B Fischl, MI Sereno, Cortical surface-based analysis. i. segmentation and surface reconstruction. *NeuroImage* **9**, 179–194 (1999).
25. P Hagmann, et al., Mapping the structural core of human cerebral cortex. *PLOS Biol.* **6**, e159 (2008) Publisher: Public Library of Science.
26. JD Tournier, F Calamante, A Connelly, MRtrix: Diffusion tractography in crossing fiber regions. *Int. J. Imaging Syst. Technol.* **22**, 53–66 (2012) _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/ima.22005.
27. O Civier, RE Smith, CH Yeh, A Connelly, F Calamante, Is removal of weak connections necessary for graph-theoretical analysis of dense weighted structural connectomes from diffusion MRI? *NeuroImage* **194**, 68–81 (2019).
28. BT Thomas Yeo, et al., The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106**, 1125–1165 (2011).
29. A Ponce-Alvarez, ML Kringelbach, G Deco, Critical scaling of whole-brain resting-state dynamics. *Commun. Biol.* **6**, 1–11 (year?) Number: 1 Publisher: Nature Publishing Group.
30. SM Smith, et al., Correspondence of the brain's functional architecture during activation and rest. *Proc. Natl. Acad. Sci.* **106**, 13040–13045 (2009).
31. W Cai, S Ryali, R Pasumarthy, V Talasila, V Menon, Dynamic causal brain circuits during working memory and their functional controllability. *Nat. Commun.* **12**, 3314 (2021) Number: 1 Publisher: Nature Publishing Group.
32. M Schirner, G Deco, P Ritter, Learning how network structure shapes decision-making for bio-inspired computing. *Nat. Commun.* **14**, 2963 (year?) Number: 1 Publisher: Nature Publishing Group.
33. S Atasoy, I Donnelly, J Pearson, Human brain networks function in connectome-specific harmonic waves. *Nat. Commun.* **7**, 10340 (2016) Number: 1 Publisher: Nature Publishing Group.
34. JZ Kim, DS Bassett, Linear dynamics & control of brain networks (2019).
35. KJ Friston, A Mechelli, R Turner, CJ Price, Nonlinear responses in fMRI: The balloon model, volterra kernels, and other hemodynamics. *NeuroImage* **12**, 466–477 (2000).
36. GK Aguirre, E Zarahn, M D'Esposito, The variability of human, BOLD hemodynamic responses. *NeuroImage* **8**, 360–369 (year?).
37. M Emch, CC von Bastian, K Koch, Neural correlates of verbal working memory: An fMRI meta-analysis. *Front. Hum. Neurosci.* **13**, 180 (2019).
38. JC Pang, et al., Geometric constraints on human brain function. *Nature* **618**, 566–574 (2023).
39. DS Margulies, et al., Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proc. Natl. Acad. Sci. United States Am.* **113**, 12574–12579 (2016).
40. A Kashyap, S Keilholz, Dynamic properties of simulated brain network models and empirical resting-state data. *Netw. Neurosci. (Cambridge, Mass.)* **3**, 405–426 (2019).
41. RM Hutchison, et al., Dynamic functional connectivity: Promise, issues, and interpretations. *NeuroImage* **80**, 360–378 (year?).
42. EA Allen, et al., Tracking whole-brain connectivity dynamics in the resting state. *Cereb. Cortex (New York, N.Y.: 1991)* **24**, 663–676 (2014).
43. Sparse matrices (scipy.sparse) — SciPy v1.11.3 manual (year?).
44. K Champion, B Lusch, JN Kutz, SL Brunton, Data-driven discovery of coordinates and governing equations. *Proc. Natl. Acad. Sci.* **116**, 22445–22451 (year?) Publisher: Proceedings of the National Academy of Sciences.

Supplemental Figures

DRAFT

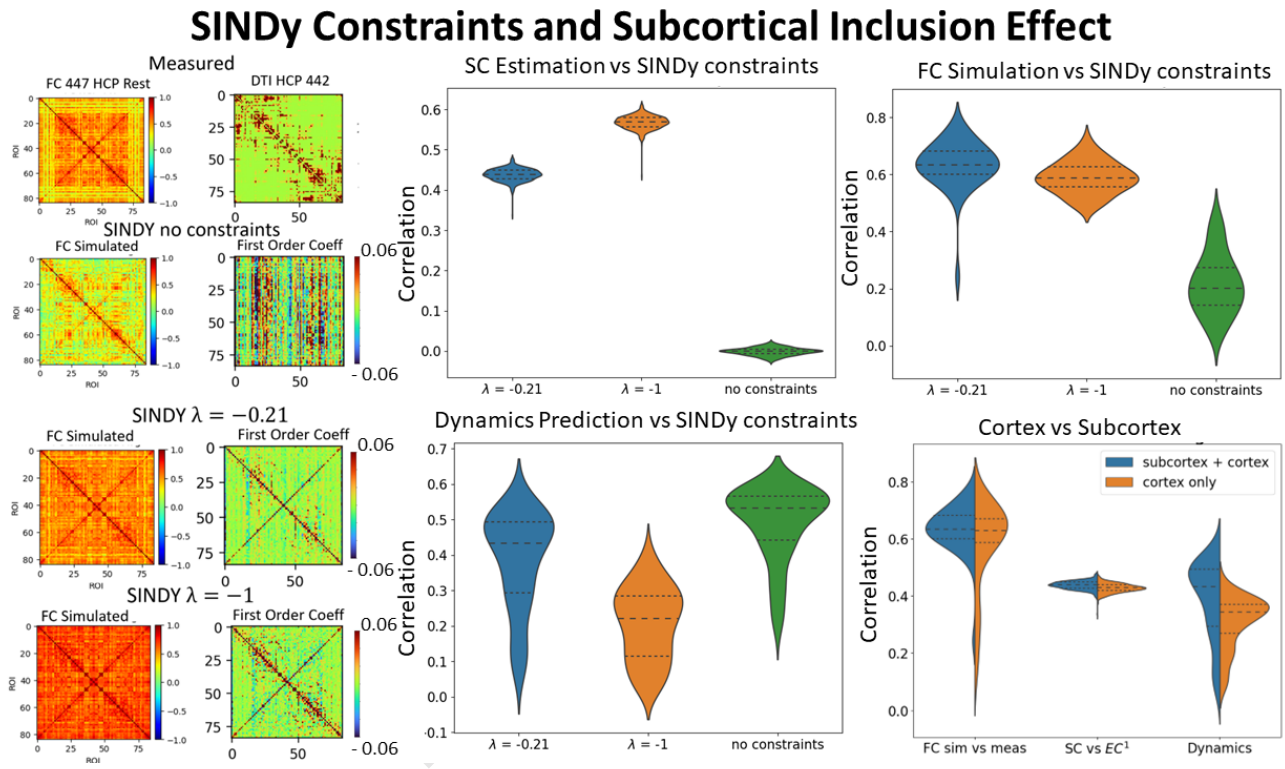

Fig. 8. Comparison between different constraint options using the SINDy algorithm. Left panel, the simulated FC and 1st Order Coefficient are provided for no constraints, constraint close to the decay rate for the measured HRF, amd a large constraint of -1. No constraint has the best fit for the dynamics 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.

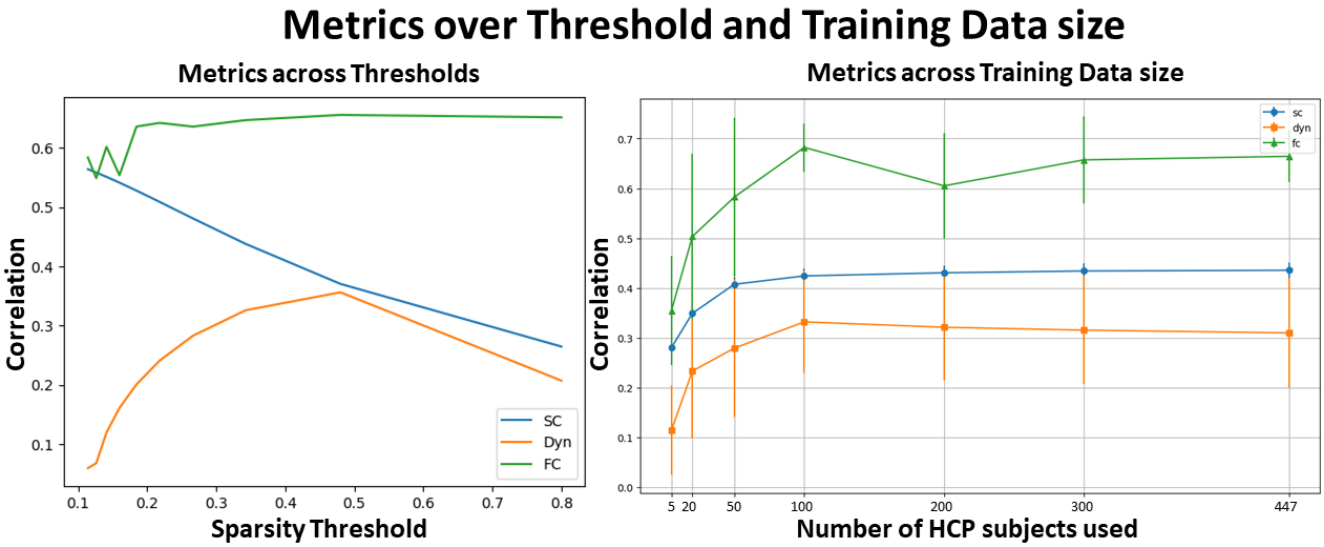

Fig. 9. Left: Looking at the performance of multiple metrics over the value of sparsity. While the correlation of the first order coefficents to the SC, and the prediction of the derivative vastly change over the range of sparsity threshold, the simulated functional connectivity barely changes across the whole range. Therefore we justified using SC and dynamics as well as the HRF as the main metrics to distinguish between the models. Right: 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.

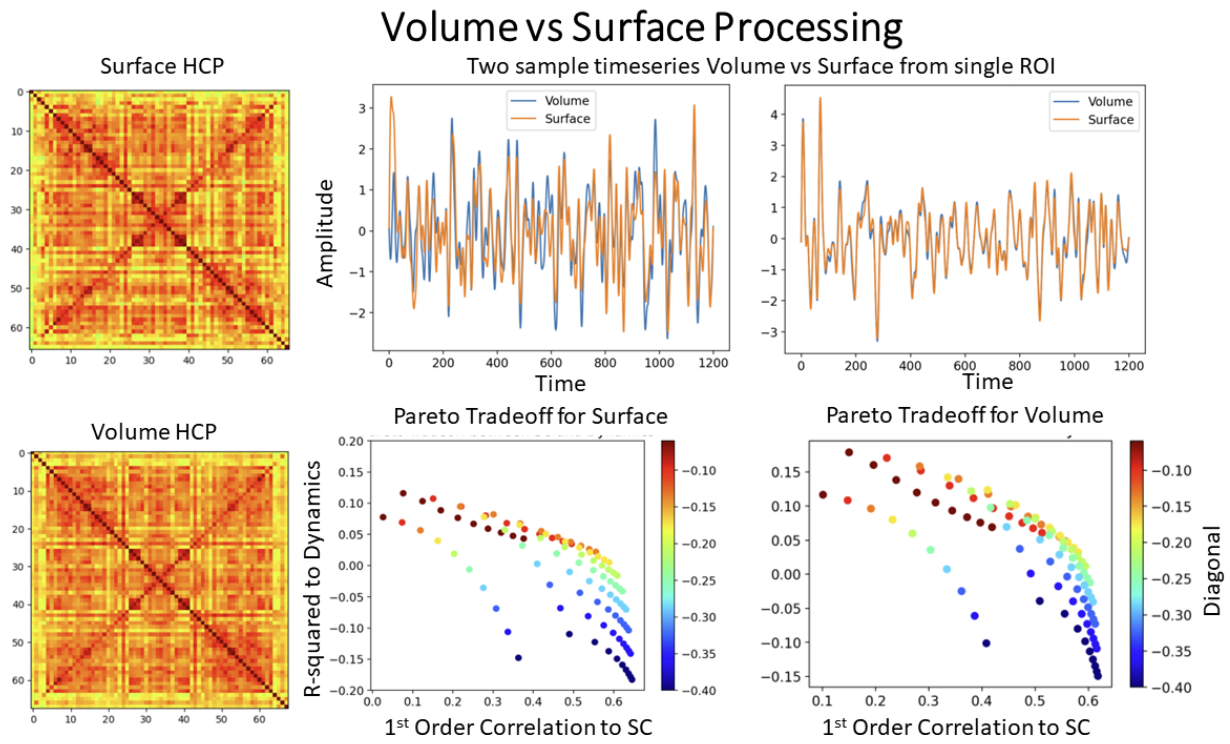

Fig. 10. Comparison of different types of preprocessing in the HCP pipelinee. The Surface HCP refers to files that have gone through the surface HCP pipeline and then using the provided parcellations in surface space, averaged across each ROI. The volume HCP refers to files that were segmented in their native space (not group space) and the ROI timeseries extracted from volume ROI. 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 surface HCP files had higher correlations of their first order coefficients to the SC while the volume had much better fit to the dynamics. Since our primary intention is to separate the signal dynamics the HCP Volume preprocessing was used.

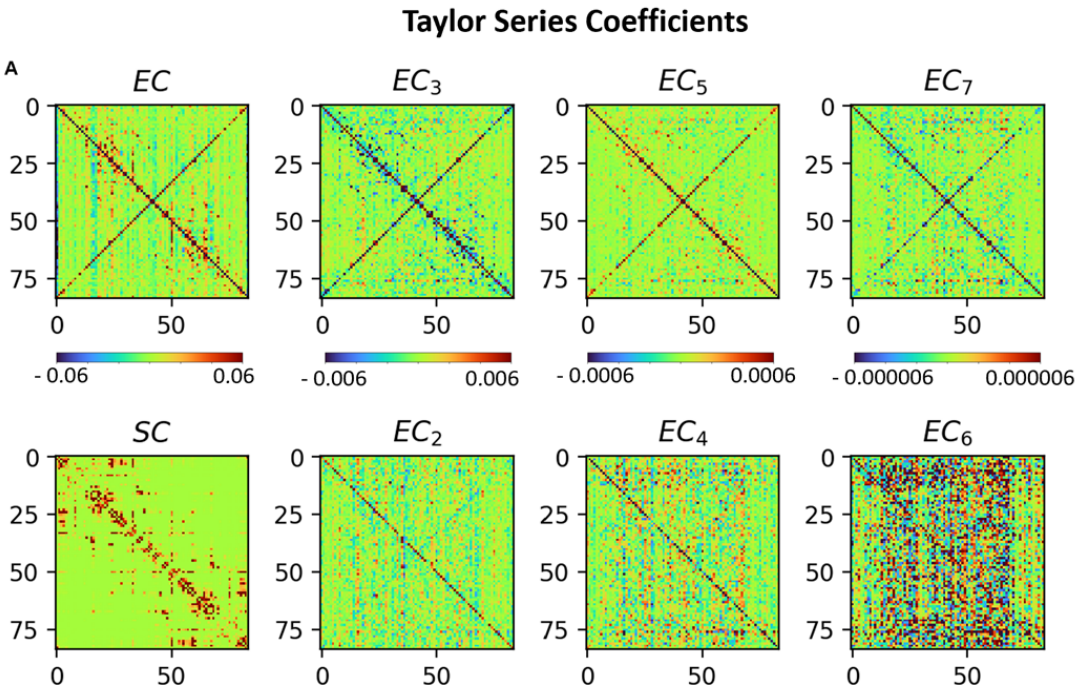

Fig. 11. A) The first seven Taylor series coefficients for $\lambda = -0.21$. Top and bottom have the same colorbar. This are the full coefficients displayed in Figure 4 and at the working biased point of the SINDy algorithm for fMRI.

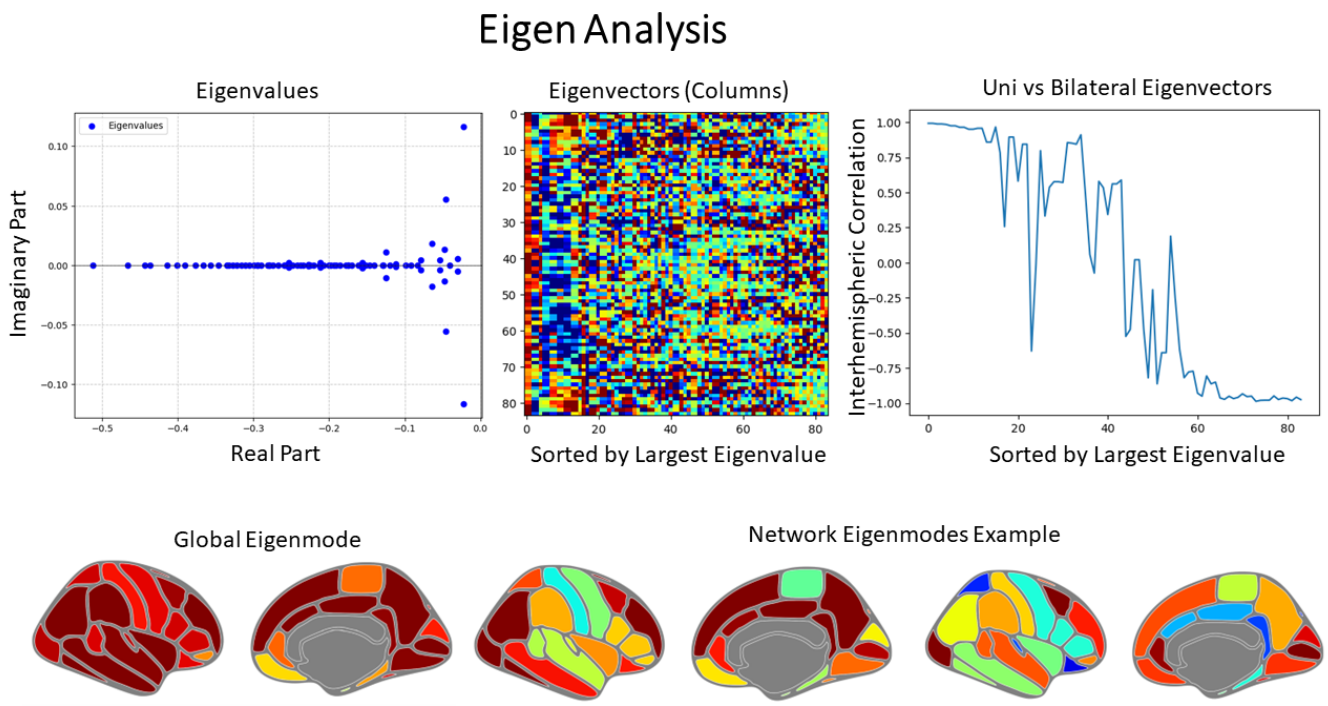

Fig. 12. Eigen Analysis of the Jacobian (first order taylor coefficients). The Eigenvalues shown in top left are plotted on the complex plane. All eigenvalues have a real value less than one, suggesting the system is stable, although some modes have real components fairly close to 0 suggesting the system 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.

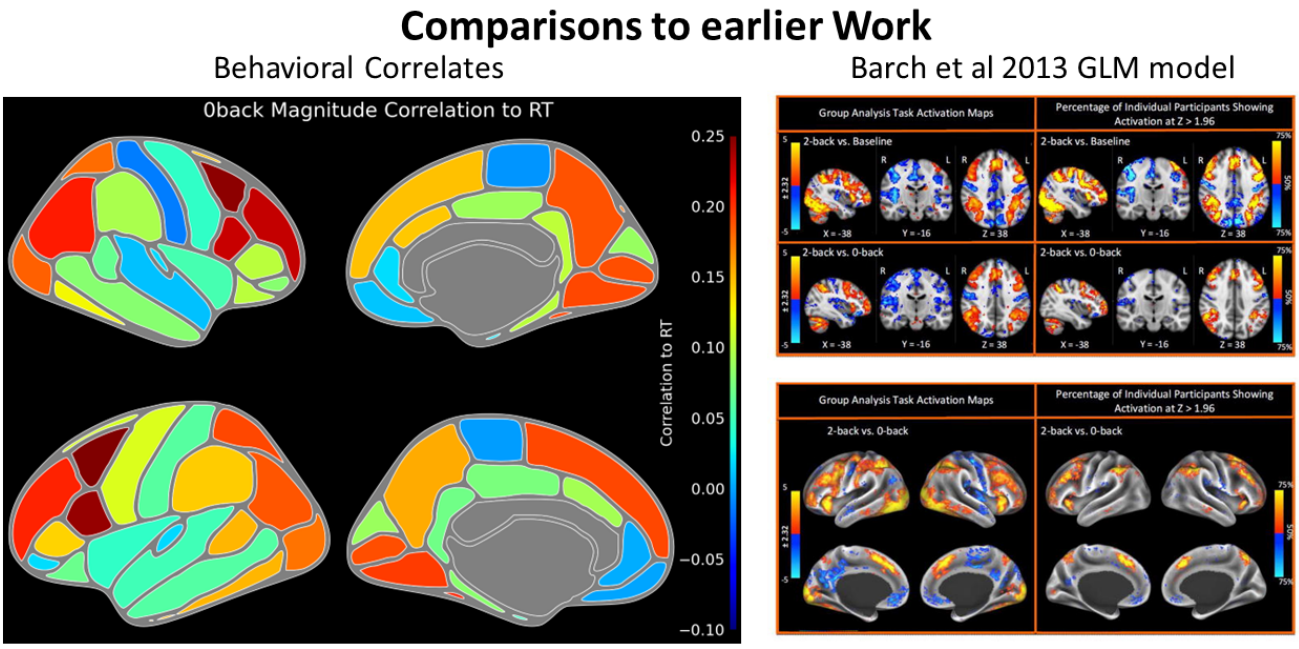

Fig. 13. Comparisons between our results spatially vs earlier GLM models on HCP WM data. Left the correlates for each brain region calculated based on reaction time for 0back averaged across all subtasks. Top Right: GLM model for accessing the HRF function and the significant voxels highlighted. Bottom Right: The voxels that activate for one particular task over another.

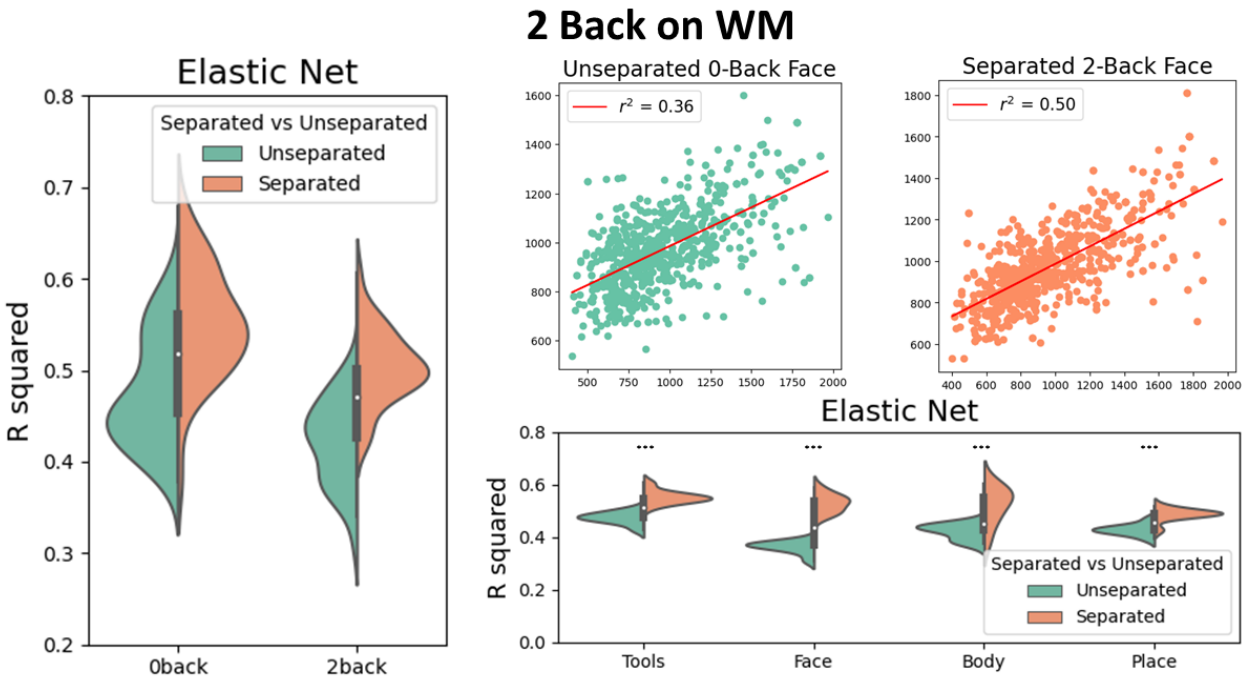

Fig. 14. The decoding of the 2back WM subtasks. The fit to the two back data is slightly worse than for 0back, but yields very similar results in the terms of unseparated/seperated data as shown in the left panel. Across the subtasks bottom right, we can see significant changes for each subtask for 2back as well. Note, a single model for WM was trained using all of the WM data, for both 0 and 2back for all subtasks. However, for the final elastic net step, in order to integrate the spatial temporal signal, a separate linear model was trained for each of the 8 different subtasks.

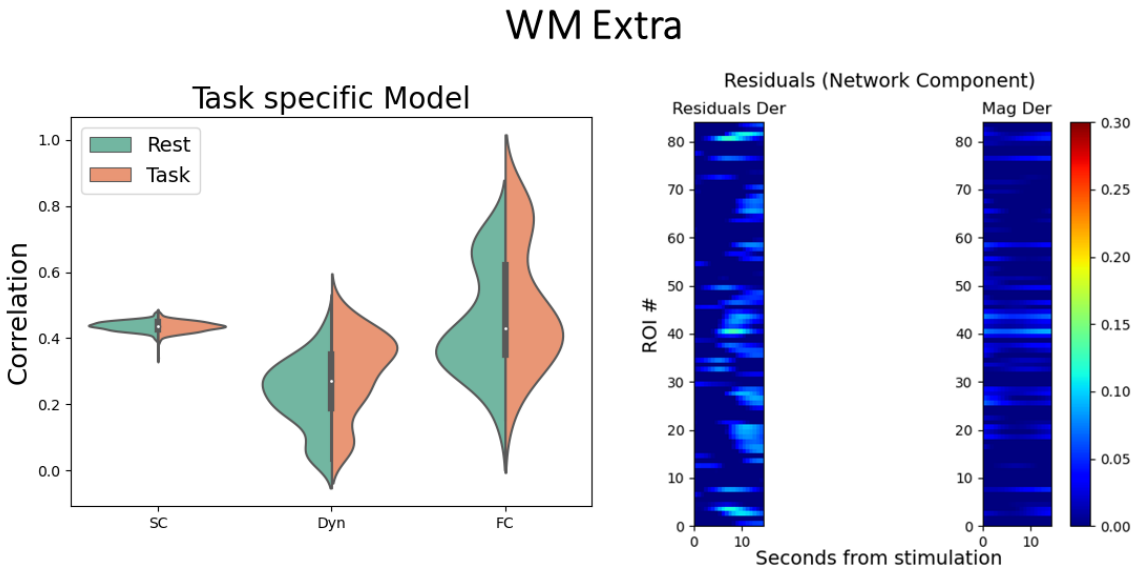

Fig. 15. Left) The specificity of the Task ODE identified by SINDy. The Rest and Task ODE models were compared how well they predict the derivative during a task block (Dyn) as well as the FC computed from the task dataset. The rightmost shows they have roughly same coeffcients (about 0.95 correlated with each other) and the first order coefficients are equidistant from the measured SC. Right) The Stimulus Independent component estimated as Rest-Task is correlated with the reaction time to see how much of this component is still correlated with the task. There is a small amount correlation around 0.1 left in this signal and probably could be extracted using a more complicated nonlinear separation of rest and task