FRONT MATTER

Title

- Estimation and Validation of Individualized Dynamic Brain Models with Resting State fMRI

Authors

Matthew F. Singh,[1,2]*, Todd S. Braver,[1] Michael W. Cole,[3] ShiNung Ching[2,4]

Affiliations

[1]Department of Psychology, Washington University in St. Louis, St. Louis, MO, USA.
[2]Department of Electrical and Systems Engineering, Washington University in St. Louis, St. Louis, MO, USA.
[3]Center for Molecular and Behavioral Neuroscience, Rutgers University, Newark, NJ, USA.
[4]Department of Biomedical Engineering, Washington University in St. Louis, St. Louis, MO, USA.
*f.singh@wustl.edu

Abstract

A key challenge for neuroscience is to develop generative, causal models of the human nervous system in an individualized, data-driven manner. Previous initiatives have either constructed biologically-plausible models that are not constrained by individual-level human brain activity or used data-driven statistical characterizations of individuals that are not mechanistic. We aim to bridge this gap through the development of a new modeling approach termed Mesoscale Individualized Neurodynamic (MINDy) modeling, wherein we fit nonlinear dynamical systems models directly to human brain imaging data. The MINDy framework is able to produce these data-driven network models for hundreds to thousands of interacting brain regions in just 1-3 minutes per subject. We demonstrate that the models are valid, reliable, and robust. We show that MINDy models are predictive of individualized patterns of resting-state brain dynamical activity. Furthermore, MINDy is better able to uncover the mechanisms underlying individual differences in resting state activity than functional connectivity methods.

MAIN TEXT

Introduction

To understand human brain function, it is necessary to understand how its components interact across spatial and temporal dimensions. This understanding can take many forms, ranging from statistical descriptions of correlations between brain regions to generative models, which provide a formal mathematical description of how brain activity evolves in time. Indeed, efforts have taken quite different approaches based upon what data is available in human vs. nonhuman subjects. Several international neuroscience initiatives have relied upon nonhuman subjects to collect vast amounts of anatomical and electrophysiological data at the cellular scale (e.g. [1]). Generative models are then formed by integrating these cellular-level observations with known neuronal biophysics at the spatial scale of individual neurons or small populations (e.g. [1]).

In contrast, another set of large initiatives has instead focused on modeling individual human brain function using an approach often referred to as “connectomics” (e.g., Human

Connectome Project, [2]). This approach relies on descriptive statistics, typically correlation between fluctuating activity signals in brain regions assessed during the resting state (``resting state functional connectivity" or rsFC; [3]). As a result, it is sometimes difficult to make mechanistic inferences based upon functional connectivity correlations ([4]). Moreover, neural processes are notoriously nonlinear and inherently dynamic, meaning that stationary descriptions, such as correlation/functional connectivity, may be unable to fully capture brain mechanisms. Nevertheless, rsFC remains the dominant framework for describing connectivity patterns in individual human brains.

Despite the promise of human connectomics, there have been only a few attempts to equip human fMRI studies with the sorts of generative neural population models that have powered insights into non-human nervous systems. Notable advances have occurred in data-driven approaches, with methods being developed to identify directed, causal influences between brain regions (e.g. [5]). Conversely, neural mass modeling approaches have also been extended to study human brain activity in a generative fashion, and these have provided new insights into the computational mechanisms underlying fMRI and MEG/EEG activity dynamics ([6] [7] [8]). Neural mass modeling approaches have also been applied to individuals by incorporating single subject diffusion data and anatomy ([6]), but, unlike (linear) data-driven approaches (e.g. Dynamic Causal Modeling; [5]), these models have not been directly fit to the individual brain activity dynamics they seek to capture. However, both of these approaches have important limitations. In particular, the data-driven approaches are subject to potential misinferences due to assumptions of linearity ([9]) or limitation to a relatively small number of brain regions (e.g. [5]). Likewise, with neural mass modeling approaches, their ability to quantitatively recreate key features of individual-level functional connectivity has also been limited ([10] [7] [8]). This may be because the most common approach is to parameterize connectivity from estimates of white matter integrity from diffusion imaging, which also can lead to potential misinference, since these connectivity estimates are constrained to be symmetric and positive ([11]). Indeed, up until this point, there has been no general solution put forward to directly fit generative models that can successfully capture – and predict – whole-brain activity dynamics at the level of individual humans. This limitation is critical because in order to accurately characterize individual variation in humans – which is the goal of personalized neuroscience and precision medicine initiatives (e.g. [12])– individualized brain models are required.

In the current work, we aim to fill this gap, by advancing high-resolution characterization of the human connectome beyond statistical correlation matrices to nonlinear dynamical systems models of hundreds of interacting neural populations. Most critically, the models are optimized to capture brain activity dynamics at the level of individual human subjects. We present a computationally efficient algorithm to fit these models directly from human resting-state fMRI. The algorithm extends data-driven techniques towards the estimation of biologically plausible models, and conversely enables the parameterization of dynamical neural models in a data-driven, individualized fashion. We describe our efforts to develop and validate these models, demonstrating that they successfully characterize whole-brain activity dynamics at the individual level, and as such can be used as a more powerful alternative to rsFC. Because of this goal, we term our modeling approach MINDy: Mesoscale Individualized Neural Dynamics. In the sections below, we introduce the MINDy modeling framework, highlighting its most innovative and powerful features, and presenting results that validate its utility as an analytic tool for investigating the neural mechanisms and individual differences present in fMRI data

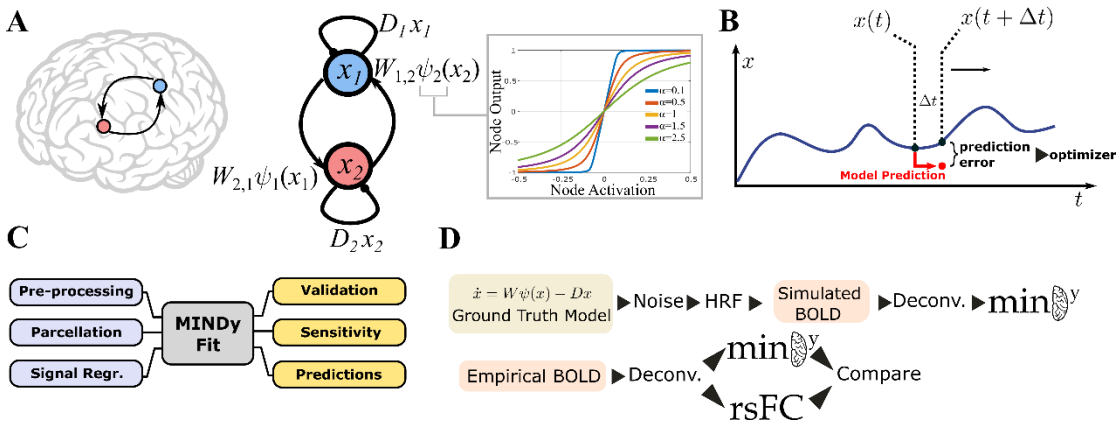


Fig. 1. Overview of Methods Employed. A) The MINDy model consists of coupled 1-dimensional abstracted neural-mass models (Hopfield form). The shape of the transfer-function for each brain region is parameterized by a curvature parameter α. B) Model goodness-of-fit was measured through one-step prediction of the empirical time-series. C) Overview of data processing and analyses: data was processed according to Siegel and colleagues and parcellated. Reported analyses fall into three categories: validation, sensitivity to nuissance parameters, and predictions of brain activity patterns. D) In both simulations and empirical analyses, the BOLD signal was deconvolved with a canonical HRF function (see Methods) before being analyzed with either MINDy or rsFC.

Results

2.1 Robust Estimation of Individualized Neural Model Parameters

The key premise of our approach is an expansion of the architectural description of brain networks from a simple connectivity matrix, to a more biophysically interpretable dynamical model:

$$\dot{x} = W\psi_\alpha(x_t) - Dx_t + \epsilon_t. \quad \textbf{(1)}$$

This model, which resembles a neural mass model (e.g. [13]) describes the evolution of brain activity at each anatomical location (each element of the vector x_t). Unlike true neural-mass models, we model abstracted brain activity commensurate with the fMRI timescale, rather than the evolution of population firing rate over milliseconds (see SI Sec. S.1 for discussion). Our model is similar, however, in that it is described by three components: a weight matrix (W) which identifies pathways of causal influence between neural populations, a parameterized sigmoidal transfer function (ψ) which describes the relation between the local activity of a population and its output to other brain regions (Eq. 2, Fig. 1A) and a diagonal decay matrix (D) which describes how quickly a given neural population will return to its baseline state after being excited (i.e. the time-constant; Fig. 1A). Process noise is denoted ϵ_t and is assumed to be uncorrelated between parcels. The additional parameters (α and D) are revealed to have consistent anatomical distributions (Fig. 2 A,B), which reflect regional variation in intrinsic dynamics (D) and efferent signaling (α). These properties vary with brain network and are consistent even at the finer within-network scale (Fig. 2A,B). Thus, our model, like a neural mass model, parameterize both the interactions between brain regions and the processes that are local to each brain region that make it distinct. While theoretical neural mass models operate in continuous-time, fMRI experiments have limiting temporal sampling rates. Therefore, we approximate **(1)** by fitting a discrete-time analogue with time step Δt (e.g. the sampling TR; Fig. 1B):

$$x_{t+\Delta t} - x_t = (W\psi_\alpha(x_t) - Dx_t + \epsilon_t)\Delta t. \quad \textbf{(2)}$$

Parameter estimation in the MINDy algorithm contains three main ingredients, which ensure that estimates are robust, reliable, and valid. First, the transfer functions of neural mass models are allowed to vary by brain region through the scalar parameter α.

$$\psi_\alpha(x_t) := \sqrt{\alpha^2 + (bx_t + .5)^2} - \sqrt{\alpha^2 + (bx_t - .5)^2}. \quad \textbf{(3)}$$

Each brain region has its own α parameter, fit on a subject-wise basis, while b is a fixed global hyperparameter (b=20/3 for the current case). The use of a parameterized sigmoid allows for additional anatomical heterogeneity in region-wise dynamics. Secondly, we make use of recent advances in optimization to ensure that the fitting procedure is robust. By using Nesterov-Accelerated Adaptive Moment Estimation (NADAM, [14]) we achieve the speed advantage of stochastic gradient descent (SGD) algorithms, while at the same time preventing both over-fitting and under-fitting (see SI Sec. S.2 for discussion). Lastly, we constrain the problem by decomposing the large matrix of connection weights (W) into two simultaneously fit components: a sparse component W_S and a low-dimensional component $W_L := W_1 W_2^T$ in which both W_1 and W_2 are $n \times k$ rectangular matrices with n being the number of neural populations (brain parcels) and $k<n$ being a global constant that determines the maximum rank of W_L. This decomposition is advantageous for concisely representing the interactions of structured networks and is the most important element of the fitting process. Sparseness criteria were achieved through L_1 regularization with the resultant fitting objective being to minimize:

$$\begin{aligned} J = {} & \frac{1}{2} E_T \left[\| (X_{T+\Delta t} - X_T) - [(W_S + W_L)\psi_\alpha(X_T) - DX_T] \|_2^2 \right] \\ & + \lambda_1 \| W_S \|_1 + \lambda_2 Tr(|W_S|) + \lambda_3 (\| W_1 \|_1 + \| W_2 \|_1) + \frac{\lambda_4}{2} \| W_L \|_2^2 \end{aligned} \quad \textbf{(4)}$$

The notation E_T denotes the mean over all temporal samples considered (the "minibatch" of each iteration) so the first term simply corresponds to the mean square error of predictions. Each of the remaining penalty terms have a global regularization constant (λ_i) that is shared across all subjects. This regularization scheme was adopted in order to reduce the dimensionality of the parameter estimation problem, while at the same time, attempting to reflect the consistently observed community-structure of brain connectivity measures. Under this view, brain connectivity patterns can be described in terms of communities (sub-networks) linked together by highly connected hubs. We envision the sparse component of connectivity to mimic the communication between connectivity hubs. By contrast the low-rank component is meant to account for the propagation of signals from hubs to their corresponding subnetworks and vice-versa. From a Bayesian perspective, this penalty function is equivalent to maximum likelihood estimation with fixed Laplace (exponential) priors on the distributions for each of the individual weight matrices and a normal prior on the combined low-rank component. From a linear-algebra perspective, the regularization prioritizes matrices which can be minimally perturbed to produce a skewed eigenvalue spectrum with sparse eigenvectors. It is important to note that the sum of sparse and low-rank matrices (e.g. W) need not be sparse nor low-rank so this decomposition is remarkably flexible. The values of each λ_i and their determination is detailed in SI Table 4 and SI Sec. S.3.

To benchmark the performance of MINDy model estimation and compare with functional connectivity analyses we used resting state fMRI scans from 53 subjects in the Human Connectome Project ([15]). Each subject contributed four scans of 15 minutes, with each scan consisting of BOLD timeseries acquired at 720ms intervals (1200 datapoints / scan). Data were acquired in two separate scanning sessions (two scans × two days), which we leverage for test-retest reliability analyses. Data were preprocessed through the rsFC pipeline proposed by Siegel and colleagues ([16]). The first stage of this pipeline is the HCP minimal pre-processing pipeline (see [15]) with FSL's ICA-FIX correction for rigid body motion ([17]). For most analyses we used the second level pipeline recommended by Siegel ([16]), but we also considered the effects of using three possible variants to test the effects of adding additional preprocessing steps (Sec. 2.6). Lastly, data were deconvolved with a canonical hemodynamic response (HRF) model to

retrieve the corresponding neural time-course (Fig. 1D). As described further below, we also tested the effect of mismatches between "true" and canonical HRF models (Sec. 2.6).

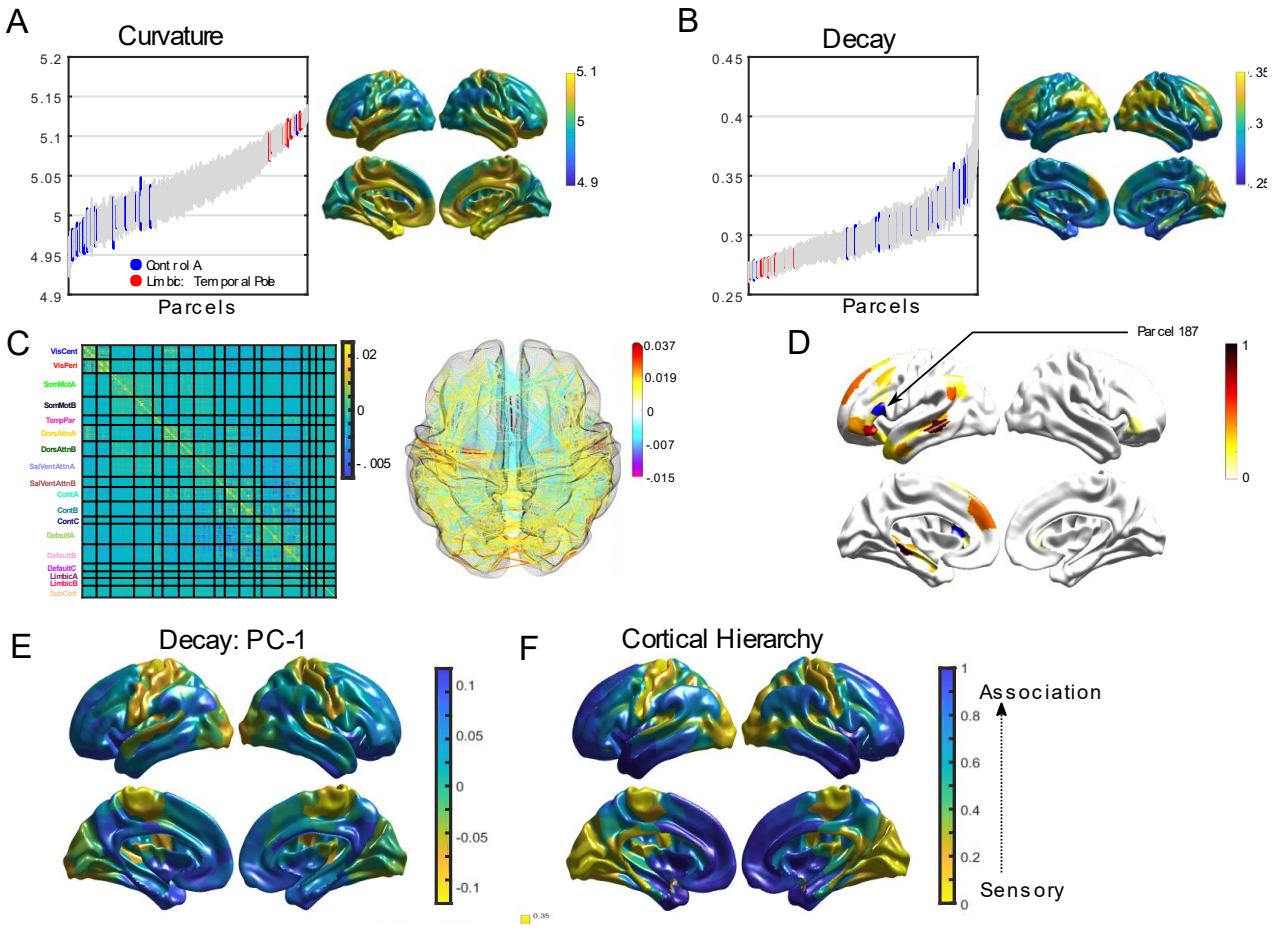


Fig. 2. MINDy Parameters display consistent anatomical distributions. A) The curvature-parameter displays network structure and is consistent across subjects at the finer parcel level. Parcels are ordered from least to greatest value for the curvature parameter (α) averaged across subjects and scanning sessions. Surface plots are for mean value. Two representative brain networks are highlighted (Control-A: red; Limbic-Temporal Pole: blue) to illustrate anatomical gradients in this parameter. B) Same as A but for the decay parameter (D). C) Right-side: Mean connection matrix (W) averaged across subjects and scanning session. Parcels are grouped according to the gwMRF-400 17-network parcellation (hemispheres combined) plus the free-surfer subcorticals. Left-side: thresholded anatomical projection. D) Parcel 187 (gwMRF-400 17-network), near Inferior Frontal Gyrus, had the strongest source-bias for positive connections (more positive out than in). Plotted surface shows the relative magnitude of this bias (only connections with outward-bias) which largely follows left-lateralized regions implicated in language see SI Sec. S.4 for additional, preliminary directed-connectivity results. Blue highlights chosen source-parcel. E) First principle component of individual variation for the decay parameter. F) Hierarchical heterogeneity provided by Demirtas and colleagues based upon erf transform of the T1/T2 ratio (MMP parcelation) and their proposed sensory-association gradient. G) The MINDy weight distribution demonstrates sparser connectivity than rsFC.

Throughout, we use the term "weights" to refer to the matrix W in estimated neural models. This is to differentiate the model connectivity parameter from the term "resting-state

functional connectivity" (rsFC), which instead refers to the correlation matrix of BOLD time-series, rather than the mechanistic concept that it is often assumed to measure (i.e. direct and indirect interactions between brain regions). We reserve the term "effective connectivity" to indicate a causal, monotone relationship in activity between brain regions that evolves over no more than 2s (the typical fMRI sampling rate). Thus, both the fit model weights and the rsFC are ways to approximate the effective connectivity, even though rsFC may not support reverse inferences regarding directedness and causality.

2.2 MINDy Retrieves Individual Differences Missed by Functional Connectivity

A key goal of our investigation was to determine whether MINDy was sufficiently sensitive to reveal individual differences in connectivity weights that have become the focus of recent efforts within the rsFC literature (e.g. [18]). We tested the model by reconstructing individual differences in connectivity weights of simulated subjects and comparing them against both classical rsFC and the partial correlation matrix. Simulated subjects were generated by permuting MINDy parameter sets across individuals (see methods). We then simulated the resultant model with process noise and hemodynamics to generate realistic BOLD fMRI time series (see methods; Fig. 1C). This provided a ground-truth set of simulated fMRI data, from which we could compute the rsFC/partial correlation matrices for each "subject", and also determine the fidelity of recovered parameters (i.e., compared against true parameters used to generate the simulated data). To assess the performance of the model estimation procedure, we considered two metrics: the validity of estimated connectivity weight differences between subjects (Fig. 3B) and the sensitivity of each procedure to different model components (SI Fig. 1A). These sensitivity analyses reveal whether each approach (rsFC matrix, partial correlation matrix, or model estimation) misclassifies variation in some other model component (e.g. decay rates) as being due to a change in weights. To better assess sensitivity, we generated data after varying only one model component at a time across the simulated subjects: the weight matrix (W), transfer functions (α) or decay rates (D).

Results indicated that MINDy was able to accurately recover the weight matrix for each individual (single-subject: R=.800±.025, inter-subject: R=.544±.049; Fig. 3A, B). Thus, the simulated weight changes that differentiated one individual from another were recovered well by the MINDy parameter estimation approach. Moreover, MINDy weight estimates were found to significantly outperform rsFC (computed on the simulated timeseries data) in their ability to accurately recover both the ground-truth connectivity matrix of simulated individual subjects (FC: R=.436±.065, *paired-t*(33)=45.629, $p\approx0$), as well as the differences between individuals (FC: R=.293±.046; *paired-t*(33)=58.618, $p\approx0$; Fig. 3B) This finding suggests that the modest relation between rsFC and ground-truth connectivity weights is primarily driven by the group-average connectivity as opposed to individual differences. However, the rsFC may be a weak comparison as it does not typically permit sparseness commensurate with empirical MINDy weights (Fig. 2G). Therefore, we used partial correlations as an additional benchmark. While partial correlations quantitatively improved upon rsFC estimates (single-subject: R=.537± .032, inter-subject: R=.392 ± .027), performance remained significantly lower than MINDy (single-subject: *paired-t*(33)=40.51, $p\approx0$, inter-subject: *paired-t*(33)=23.62, $p\approx0$).

After we established that MINDy outperforms rsFC and partial correlations in retrieving true individual differences in weights, we benchmarked the sensitivity of each approach to other sources of individual variation. Rather than measuring how well each procedure correctly retrieves connectivity, these tests quantify how well each approach selectively measures connectivity as opposed to other sources of variation (see Methods). We quantified sensitivity in terms of how often MINDy and rsFC reported that a connection differed between simulated

models, when in reality only the curvature or decay terms were altered (SI Fig. 2A). Results indicate that MINDy correctly detects the sources of individual variation when due to local changes such as decay rate (7.6±0.6% of connections changed vs. 5% by chance) and transfer function shape (6.3±.2%), as these have no appreciable impact on MINDy's connectivity estimates (the false positive rate is near that expected by chance). By contrast, rsFC measurements were highly sensitive to the decay rate (decay: 27.5±12%, *paired-t*(339)=31.753, $p \approx 0$ vs. MINDy) and slightly more sensitive to curvature (7.9±2.7%, *paired-t*(339)=10.639), indicating that some individual differences in FC may be reflective of purely local brain differences as opposed to connectivity between brain regions. These results indicate that MINDy promises to improve both the mechanistic sensitivity and the anatomical accuracy of inferences based upon individual differences in resting-state fMRI.

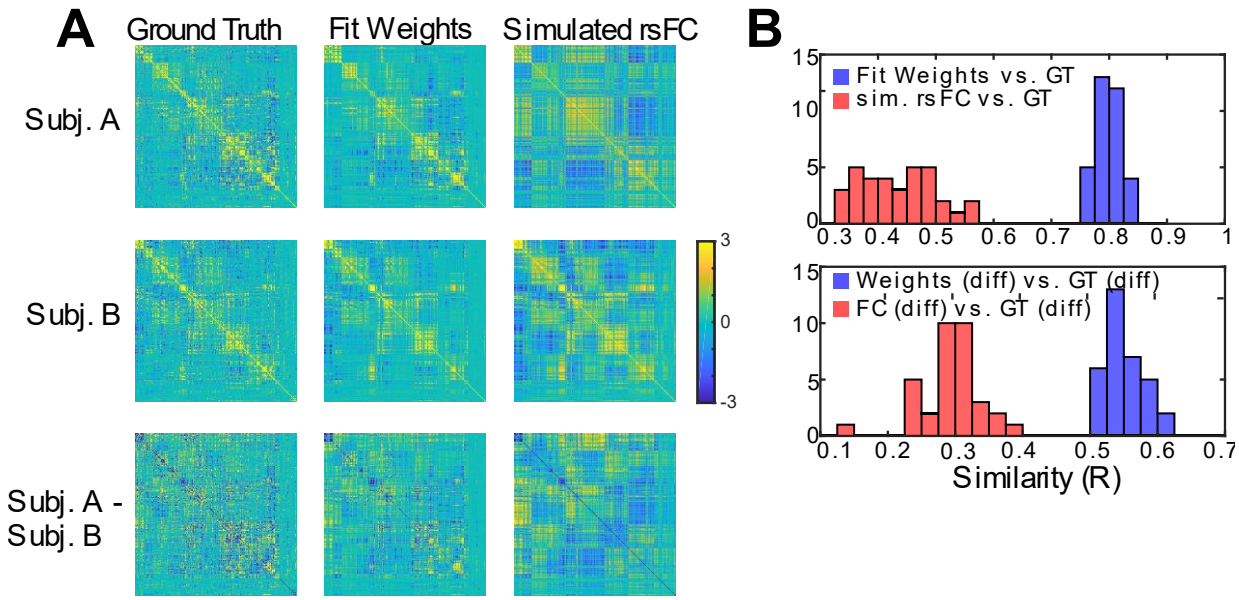


Fig 3. Ground-truth validation of MINDy and rsFC. Validation is in terms of single-subject parameters and inter-subject variation. A) First column: Example ground-truth weight matrices for two simulated subjects (top two rows) and the difference between ground-truth weights (bottom-row). Second column: Recovered weight matrices using MINDy for both subjects and their difference. Third column: same as second but using the rsFC. Fit weight matrices and simulated FC matrices are shown in standard-deviation (SD) units with SD computed across the off-diagonal elements of each individual matrix. The ground-truth matrices are displayed in units 2/3×SD to aid visual comparison. B) Top row: histogram of performance at the single-subject level (correlation with ground-truth [GT]) for MINDy (blue) and rsFC (red). Bottom row: same as top but for predicting the differences between subjects.

2.3 MINDy Parameters are Reliable and Personalized

In addition to determining the validity of MINDy parameters, it is also critical to establish their reliability. We examined this question by analyzing measures of test-retest reliability of the parameter estimates obtained for human subjects contributing resting-state scans on two separate days (30 minutes each). Results indicated that MINDy had high test-retest reliability for all parameter estimates (Fig. 4A). The reliability of weight estimates was significantly higher than rsFC reliability, although the mean difference was modest ($\Delta R \approx .045$, SI Table 1,2). By contrast, the variability in reliability was noticeably smaller for MINDy, meaning that while the mean advantage of MINDy in terms of reliability was modest, its performance was much more consistent across subjects (less variable reliability; SI Table 2).

For sake of comparison with FC we have thus far emphasized the ability of MINDy to extract brain connectivity. However, MINDy fits brain models, with the connectivity weights comprising just one component. For the approach to faithfully reflect the stable differences among individual brains, it is important that it not just accurately estimates the neural parameters that describe human brains, but that these parameters accurately capture individual differences and predict brain activity. Using the "connectome fingerprinting" approach ([19]), we compared whether MINDy parameter estimates and the combined model uniquely identify individuals within a sample. This analysis was conducted in two ways. First, we computed separate parameter estimates for each individual in each testing day session. Then we examined whether the parameters estimated from one day showed the highest similarity to the same individual on the other day (relative to all other individuals in the dataset; Fig. 4A). Secondly, we used the estimated model from one day to test whether the estimated parameters provided the best fit to the fMRI data timeseries recorded on the second day, again relative to the estimated parameters from other subjects. Specifically, this second analysis provides a strong form of cross-validation testing and we performed it for both predictions of the empirical timeseries (Fig. 4B) and for predictions of each subject's empirical rsFC, both cross-validated across sessions (Fig. 4C). In all analyses, we found that the best predicting model for every subject was almost always their previously fit model (SI Table 1). In particular, we achieved 100% accuracy when conducting connectome fingerprinting based on MINDy weight parameters (SI Fig. 2B), and when computing cross-validated goodness of fit/cross-validated predicted rsFC (Fig. 4B,C). For pairwise analyses of subjects, see SI Fig. 2F).

Similar patterns emerged but also some important differences, when conducting parallel analyses using rsFC. Replicating prior findings ([19]), 100% accuracy was also achieved in connectome fingerprinting (SI Fig. 2E). However, between-subject similarity was significantly lower in the rsFC analysis. Conversely, in rsFC the distinction between across-sessions within-individual similarity scores (i.e. test-retest similarity) and the average similarity obtained between subjects was greater than that observed in the MINDy model weights (SI Table 1). These results suggest that rsFC may result in an exaggerated picture of the idiosyncratic nature of connectivity, in that within MINDy individual differences are partitioned not only into weights, but also into other mechanistic parameters that are attributed locally, to the node/parcel (i.e., the decay [D] and curvature [α] parameters). In other words, MINDy may provide a richer and more variegated perspective on the nature of individuality, than what can be obtained with rsFC which lumps together what may be multiple dimensions of individual difference, into a simple, undifferentiated measure. For applications such as biomarker discovery, these properties may not be relevant in that the apparent magnification of individual differences in rsFC over MINDy weights could prove beneficial despite the mechanistic ambiguity of rsFC. However, we also note that MINDy provides additional parameters (curvature and decay) which may also prove useful for biomarker discovery. Lastly, the relevant dimensions for biomarker discovery are in terms of separating phenotypes, rather than separating all individuals. Since MINDy can robustly separate individuals, it has the potential to influence biomarker discovery, but whether it possesses quantitative advantages over rsFC will need to be investigated in the context of explicit biomarker questions (and may be phenotype-specific).

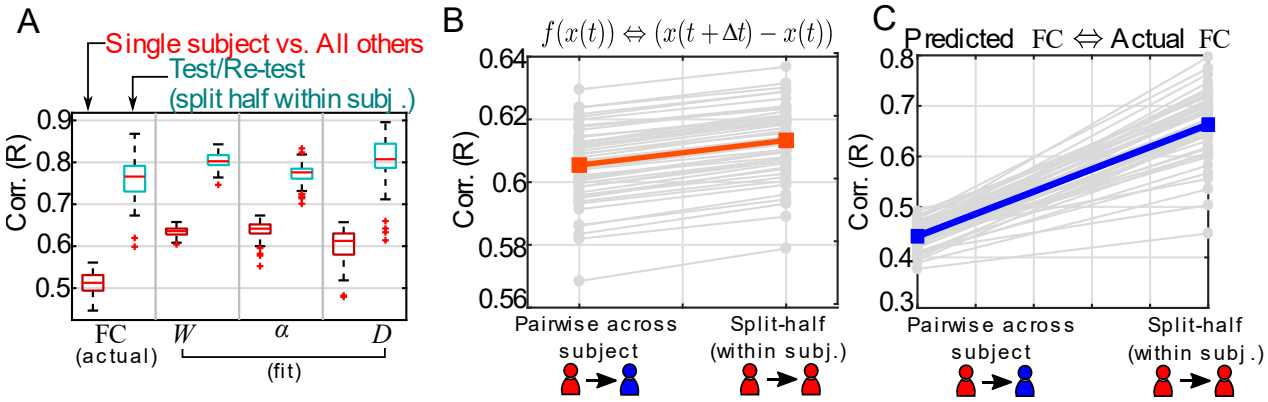


Fig. 4. MINDy parameters and predictions are personalized and reliable. A) Comparison of the test-retest similarity between subjects (red) and the test-retest reliability (blue) for rsFC and the MINDy parameters. B) Goodness-of-fit for a single time-step prediction is uniformly (but minutely) greater for comparing test-retest predictions within a subject vs. between subjects. Performance is in terms of predicting the difference time series. Red line indicates group-mean C) This relationship magnifies across time steps as evidenced by far greater similarity in test-retest predicted FC from model simulations of the same subject vs. different subject. Performance is in terms of predicting the empirical rsFC on a different scanning session. For similarity to the same or both scanning sessions see SI Fig. 6. Blue line indicates mean.

2.4 Novel MINDy Parameters show reliable individual and anatomic variation

Interestingly, we observed important additional functional utility from examining the novel MINDy parameters that are unavailable in standard rsFC. With regard to individual variation and fingerprinting analyses, we found that even ignoring the weights completely, the transfer function curvature parameter (α) associated with each node showed high consistency across sessions within an individual, and also unique patterns across individuals, such that 100% accuracy could also be achieved in fingerprinting analyses (SI Table 1). A slightly lower accuracy (94.3%) was observed when using the MINDy decay (D) parameters, though even here performance was still significantly above chance (1.89%) in identifying individuals (SI Table 1). Pair-wise, between-subject, comparisons of similarity in these parameters are reported in SI Fig. 2B-E).

Although the above analyses focused on individual differences in the unique MINDy parameters, these parameters also exhibited common patterns across individuals (SI Fig. 1B,F) that revealed interesting anatomical structure and gradients (Fig. 2A,B). These may reflect regional variation in intrinsic dynamics (D) and efferent signaling (α) that vary across brain networks (SI Fig. 1 A,E), but also exhibit consistency even at the finer within-network scale (Fig. 2A,B; SI Fig. 1B,F). Group-mean values show the same anatomical gradient across the gwMRF ([20]; SI Fig. 1D,H) and MMP ([21]SI Fig. 1C,G) atlases. It is important to note that the decay parameter only describes temporal integration at time scales commensurate with fMRI sampling. Thus, the decay parameter should not be conflated with the time-constant of traditional neural mass models just as the latter is distinct from the membrane time constant of neuronal models. Interestingly, however, the decay parameter in MINDy appears to reflect components of both temporally-extended signal integration and the time-constant of local sub-second integration. Whereas the mean value of the decay parameter correlates with absolute global brain connectivity (i.e. the sum of absolute values along a row of the rsFC matrix; $r(377)=.911$, $p\approx0$) the principal dimensions of individual variation (Fig. 2E, SI Fig. 1I,J) recreate the hierarchical organization of primate cortex (Fig. 2F): as derived from the T1/T2 ratio map ($r(358)=.518$, $p\approx0$; using the MMP Hierarchy map by Demirtas and colleagues [7]). This hierarchy (Fig. 2F) has been the subject of

recent studies into its relationship with local excitation/inhibition ([7] [8]) which is one physiological mechanism we suspect underlies the decay construct (see SI Sec. S.1.3). This hierarchy also predicts the time-scales of local microcircuits, patterns of gene-expression, myelin density, and function (sensory-processing hierarchy; see [7] [8])

In addition to the curvature and decay parameters, MINDy improves upon rsFC by allowing asymmetries within the connectivity/weight matrix *W*. As a simple proof-of-concept, MINDy identified a region of left Inferior Frontal Gyrus (IFG) as the parcel with the greatest asymmetry in positive connections. Specifically, this region showed a positive outward-bias in connectivity with the bias primarily exhibited in its feed-forward positive connections to ipsilateral medial temporal lobe, inferior parietal lobule (IPL), and dorsal/ventrolateral PFC (Fig. 2D). Excitatory connections of the left IFG with temporal cortex are essential features of the "language network" (e.g. [22]). Additional results revealing other brain regions showing directionality biases in connectivity are reported in SI (Sec. S.3, SI Fig. 3). In a later section, we explicitly test the robustness of asymmetry estimates and how they are affected by assumptions regarding hemodynamics and model mismatch.

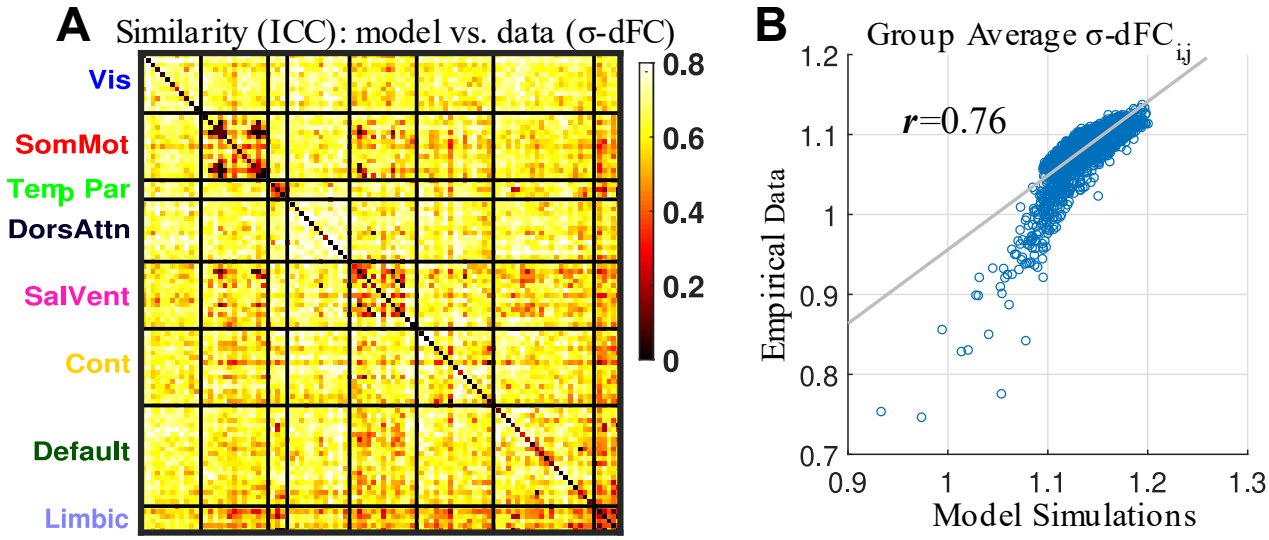


Fig. 5. MINDy models predict dynamic functional connectivity (dFC) measures. A) Similarity between model and data for predicting each subject's σ-dFC for each pair of brain regions (using the 100-parcel variant of the gwMRF atlas) and collapsing the 17-network grouping down to 8). B) Scatterplot and Pearson correlation of group-average σ-dFC for data vs. model.

2.5 MINDy Predicts Individual Brain Dynamics

We next focused our analyses on the dynamic patterns observed in brain activity, since this has been an area of rapidly expanding research interest within the rsFC literature, termed dynamic functional connectivity or dFC (e.g. [23]). Critically, the question of whether MINDy models can predict more slowly fluctuating temporal patterns in the recorded brain data for individual subjects is qualitatively distinct from the ability to predict activity over very short timescales (i.e., 1-step). This is because small biases in fitting individual time points can lead to very different long-term dynamics (e.g. compare panels B,C Fig. 4). To test model accuracy in capturing longer-term dynamic patterns, we used fitted model parameters for each subject to then generate simulated fMRI timeseries, injecting noise at each timestep to create greater variability (see Methods). We then used this simulated timeseries to identify the temporal evolution of short-term correlations between brain regions and compared results with those obtained from the recorded data. Correlation timeseries were estimated using Dynamic Conditional Correlation

(DCC; [24]), a method which has been recently shown to improve reliability in the same data-set as compared to sliding-window estimates ([25]). We then attempted to recreate dFC measures of individual subjects which have shown the greatest reliability in the actual data. Recent reliability analyses have indicated that simple statistics of temporal variation in individual correlation pairs such as standard deviation of the conditional correlation time-series ([25]) and excursion ([26]) are more reliable than state-based descriptions for the HCP resting-state data ([25]). Therefore, we used these measures to validate dynamics within the model. To avoid confusion, we use the term σ-dFC to refer to the temporal standard-deviation of time-varying correlations which is used as a measure in dFC. Alternatively, the σ-dFC may be conceptualized as the signal power of the moving-correlation time series and has proven to be one of the most reliable measures of dFC ([25]). MINDy performed slightly better on recreating group-average excursion, so we chose to be conservative and display the results from σ-dFC for main-text figures rather than using excursion dFC. Results using excursion dFC and the corresponding figures are provided in SI.

Results indicate that individual differences in the simulated dynamics of models fit to separate test-retest sessions are at least as reliable as summary dFC measures of individual differences in the original data (SI Fig. 4A,B). The image intraclass correlation (I2C2, [27]) for the model was .555 for σ-dFC and .481 for excursion. The reliability of the model was actually greater than the original experimental data itself, as the I2C2 reliabilities of the data were .527 for σ-dFC and .380 for excursion. Moreover, individual differences in the dFC of simulated models were highly correlated with those of the original data for most region-pairs (Fig. 5A). Lastly, we analyzed whether the simulated data recreates the central tendency of observed data. In general, the group-mean σ-dFC (SI Fig. 4D) and excursion (SI Fig. 4E) estimates were highly similar between the simulated and observed data for both the σ-dFC (Fig. 5B; $r(4948)=.761$, $p\approx 0$) and excursion metrics (SI Fig. 4C; $r(4948)=.836$, $p\approx 0$). Thus, MINDy models recreate measures of dFC at the level of both individual differences and the group-level. Moreover, in some cases (i.e., the excursion metric), MINDy models generate more reliable estimates than those of the original data. This potential advantage of the model for dFC investigations is likely due to the ability to simulate an arbitrarily large amount of data in MINDy models that are also free from nuisance signals/motion.

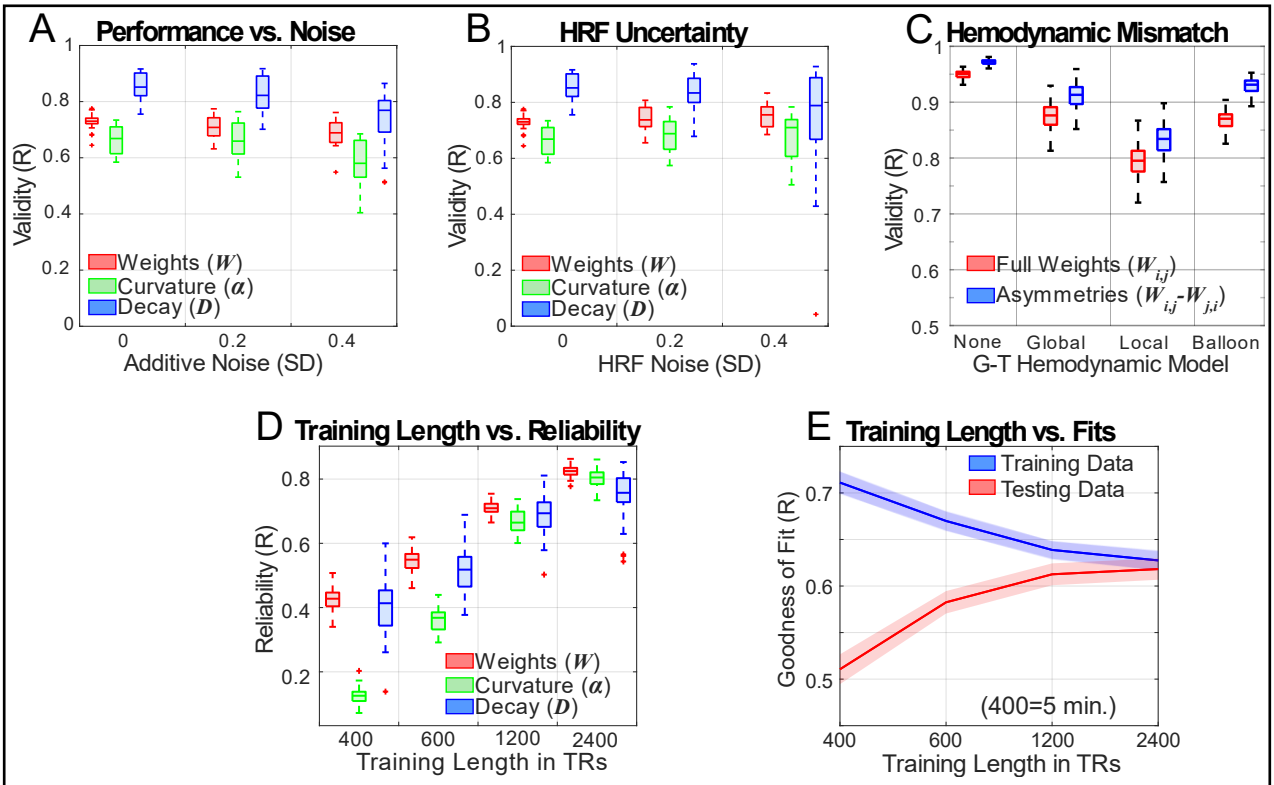


Fig. 6 MINDy is robust for 15+ minutes of data. A) Increasing the amount of additive measurement noise slightly decreases MINDy performance in recovering ground-truth parameters. B) Mean performance is unaffected by the uncertainty in the ground-truth HRF, although performance does become more variable. C) MINDy performance in retrieving the ground-truth weight matrix (original matrix: red; asymmetric part: blue) under mismatch between the ground-truth hemodynamics and the canonical HRF assumed by MINDy (left to right: No hemodynamic modeling, random spatially homogeneous HRF, spatially heterogeneous HRF, nonlinear Balloon-Windkessel model). D) Test-retest reliability of MINDy parameters as the amount of (contiguous) training data is varied. E) MINDy goodness-of-fit for 1-step prediction in training data (blue) and cross-validated with another scanning session (red). The difference between these lines indicates the degree of overfitting. Shading indicates standard deviation.

2.6 MINDy is Robust

We addressed the degree to which MINDy fitted parameters are influenced by potential sources of contamination or artifact in the observed fMRI data. Resting-state fMRI data is thought to be vulnerable to three main contaminants: noise in the BOLD signal, biases induced from post-processing pipelines that attempt to remove this noise, and idiosyncratic variation in the hemodynamic response function that relates the BOLD signal to underlying brain activity. For the first case, we considered two sources of noise in the BOLD signal: additive measurement error and motion artifact. The former case can result from random fluctuations in magnetic susceptibility, blood flow, and responsiveness of radiofrequency coils among other factors. We examined this issue using the ground-truth simulations described above, but additionally with systematic variation in the amount of measurement noise added at each time-step. Results indicated that although the performance of MINDy decreased with the amount of noise added (Fig. 6A), similarity to the ground-truth values generally remained high. At the highest level of noise considered, Weight and Decay parameters correlated $R \approx 0.7$ with ground-truth, while the curvature parameter correlated $R \approx 0.6$. We note that data exhibiting such a high level of noise would (hopefully) fail quality control.

We next examined the impact of motion on MINDy estimates. In this case, we used three standard measures of motion that were derived from the observed fMRI timeseries data: 1) the number of total frames censored due to crossing critical values of frame-wise displacement or DVARS (see Methods), 2) the median framewise-displacement of the subjects head across scanning sessions, and 3) the spatial standard deviation of temporal difference images (DVARS; [28]). We then examined whether variability in these parameters across individuals contributed to the quality of MINDy parameter estimation and individuation, using test-retest reliability (of estimated parameters from each session) as the index of quality. If MINDy estimated parameters reflected vulnerability to the degree of motion present in an individual, then we would expect higher test-retest reliability in the parameters for the individuals with the lowest estimated motion (e.g., highest data quality). Instead, we found that test-retest reliability was relatively un-impacted by any measure of motion (SI Fig. 5B-D, I). A parallel analysis used cross-validated fit, in which MINDy parameters were estimated from one session, and then used to simulate data in the held-out session, computing goodness-of-fit of the model to the observed data in this session (in terms of variance explained). In this case, we examined a subset of participants that had relatively low motion in one session, but relatively higher in the other compared against a second subset that had similar levels of motion in both sessions. If the increased motion in this latter session was problematic, it should reduce the goodness-of-fit (either when used for parameter estimation or when used for cross-validation, as the held-out session). In fact, the cross-validated fits were relatively similar in each group (SI Fig. 5E,J). Together, these results suggest that participant motion may not be a strong factor in determining how well MINDy model parameters can be estimated from observed fMRI data timeseries.

We next examined whether secondary data pre-processing pipelines, which are typically applied to rsFC data prior to analysis produce biases on MINDy parameter estimates, again examining this issue in terms of test-retest reliability. We considered three variants of a standard published preprocessing pipeline ([16]), one with motion-correction only, one which adds to this CompCor (a standard method that removes noise components associated with white matter and CSF; [29]), and a final, full variant that additionally includes global signal regression (GSR). We compared test-retest reliability for data-processed with each pipeline (SI Table 2) and the similarity of parameter estimates obtained when the same data were processed using different pipelines (SI Table 2). Results indicated that MINDy parameters had high test-retest reliability regardless of preprocessing choices (all median R>.7, SI Table 2) and that similar parameter estimates are obtained regardless of preprocessing choices (all median r>.85, SI Table 2). By comparison, when a parallel analysis was conducted on rsFC values, the rsFC parameters showed lower test-retest reliability, particularly when more pre-processing was performed on the data, and showed a larger impact of a change in pre-processing on test-test reliability. A direct comparison of the test-retest of MINDy weight parameters relative to rsFC revealed that these were significantly higher (all p's<.05), were more consistent (lower variance of reliability) across subjects (all p's<.001; SI Table 2), and were less impacted by changing preprocessing pipelines (all p's<.001; SI Table 2). Together, this set of analyses indicate that the choice of preprocessing pipeline will not have a large effect on estimated MINDy parameters.

Lastly, we considered the effect of poor estimation of the hemodynamic response function (HRF). Currently, for simplification, the MINDy estimation procedure assumes a canonical HRF model that is constant across individuals and parcels. However, existing literature suggests that this assumption is likely to be incorrect ([30]). To examine the impact of mis-fitting the HRF, we first used ground-truth simulations, in which the activity timeseries were convolved with spatially homogeneous, but randomly parameterized HRFs with incrementally greater variability. We then attempted to recover MINDy parameters while again assuming the fixed canonical HRF model (

[31]). Results of this analysis suggest that, on average, the MINDy parameters recovered from this analysis remain consistently similar to the ground truth parameters (mean similarity of all parameters, $R \approx 0.75$, Fig. 6B). However, the variability of the fits increased across simulations, as the HRF became more variable across individuals ($r(49)=.735$, $p<4.1\times10^{-10}$).

We also considered the effect of spatially heterogeneous HRFs and nonlinear hemodynamics. These effects are expected to be most pronounced in estimating asymmetric connections as unaccounted lags can potentially reverse the direction of inferred causality in many other techniques such as Granger Causality. For these simulations, we generated complex networks from a hyperdistribution whose characteristic parameters were randomly sampled at each run which allowed us to sample over a wide range of qualitatively different network structures (see methods, Sec. 4.8). Thus, these simulations did not depend upon previous empirically-fit MINDy models. We tested the ability of MINDy to recover the weight parameter (Fig. 6C) from a simulated model with four levels of hemodynamic variability: 1) no hemodynamics, 2) random, spatially-uniform HRF, 3) random, spatially-heterogeneous HRF, and 4) nonlinear hemodynamics simulated by the Balloon-Windkessel model ([32] [7], see methods Sec. 4.9). In the last case, the nonlinear hemodynamic transformation varies implicitly and systematically in space due to spatial variation in the firing-rate distribution. Results indicate that MINDy can recover asymmetric connections of ground-truth networks ($W_{i,j} - W_{j,i}$) for all cases considered, but performance depends upon the degree of HRF complexity (Fig. 6C). When no hemodynamics were included in the model (MINDy received the downsampled neural time-series) performance was near-perfect ($r=.949\pm.009$ overall; $r=.971\pm.007$ for asymmetries, $n=1700$). Performance also was high for random, spatially homogeneous HRF's both overall ($r=.874\pm.024$) and at estimating asymmetries ($r=.910\pm.023$, $n=1600$). Spatial heterogeneity of the HRF decreased MINDy performance in recovering overall ground-truth connectivity ($r=.793\pm.029$; $t(3071.8)=-86.72$, $p\approx0$; unequal-variance), but did not differentially impair the estimation of asymmetries ($r=.832\pm.028$; $t(3057.7)=11.74$, $p\approx1$, *1-tailed*, unequal variance).

It is significant that MINDy still performed well in recovering asymmetric connectivity when a nonlinear (Balloon-Windkessel) ground-truth hemodynamic model ($r=.865\pm.022$ overall; $r=.927\pm.019$ for asymmetries, $n=2020$) was used to generate simulated fMRI time-series data as compared to when a spatially homogeneous, linear HRF model ($t(3073.3)=23.03$, $p\approx0$; unequal variance) was used. Thus, violations of spatial homogeneity in the hemodynamic response appear more relevant to MINDy than violations of linearity. However, performance was still strong in all cases considered (median $r\geq.80$). We also conducted preliminary tests of MINDy's ability to recover synaptic conductance's (weights) from the simulated BOLD signal (Balloon-Windkessel) of a biophysically parameterized neural mass model ([13]) which evolves at a much faster timescale than the fMRI TR. MINDy was generally able to recover connection weights (synaptic conductance in the neural-mass model) for this case as well ($r=.684\pm.039$ overall). However, unlike in the other simulations, performance in recovering asymmetries ($r=.624\pm.052$) was lower than that of the overall weight matrix (*paired-t*$(1399)=-109.172$, $p\approx0$). This result indicates that the difference in time-scales between neuronal and BOLD activity is a more relevant constraint on directional inferences than hemodynamic variability. Although these simulations represent but a small subset of possible ground-truth models, they indicate that the directionality of MINDy connectivity estimates remains largely accurate under violations of the assumed spatially homogeneous hemodynamic response.

2.7 MINDy Requires 15-20 minutes of Data

In most fMRI experiments, scanner-time is a precious resource and particularly so with sensitive populations. While the Human Connectome Project affords a full 60 minutes of resting-

state scan time, this quantity of data may not be a reasonable expectation for other datasets, so we varied the training data length to determine how much data was necessary for MINDy to reliably estimate models. We first evaluated reliability in terms of test-retest on MINDy parameters estimated from separate scanning days using only a subset of the total data for model fitting. As expected, when the length of data used to estimate parameters increased, the test-retest reliability of the estimated parameters also increased, up to the maximum interval considered (30 minutes). Nevertheless, acceptable levels of reliability ($R \approx .7$) were obtained with 15 minutes of data (Fig. 6D). We next examined bias or overfitting of MINDy parameters by comparing the fit to the trained data, relative to a cross-validation approach, examining the fit to held-out (testing) data. As would be expected with over-fitting bias, as the length of the training data increased, the fit to the trained data decreased. In contrast, the fit to the held-out (test) data increased, and the two values converged at around 15 minutes of data in training and test sets (Fig. 6E). Thus, we do believe the current method is too prone to over-fitting biases and unacceptably low reliability with fewer than 15 minutes of total scan time using the HCP scanning parameters. However, since the data does not need to be acquired in a single continuous run, we believe that this requirement is reasonable and in concert with current recommendations for rsFC analyses ([18]). Future study with other fMRI acquisition techniques may illuminate how data-requirements change with sampling rate (e.g. shorter TRs may potentially compensate for less scan time).

Discussion

3.1 Relationship with Functional Connectivity

There are three primary advantages to using MINDy over rsFC. First, rsFC is limited as a statistical descriptive model. This means that even though rsFC may be found to be reliable and powerful as a biomarker that can characterize individuals and effects of experimental variables, it is unable to predict how the nature of an experimental manipulation relates to the observed changes in brain activity dynamics and interactions between brain regions. By contrast, MINDy is a true mechanistic causal neural model, which attempts to capture the physical processes underlying resting-state brain activity in terms of neurobiologically realistic interactions and nonlinear dynamics. This feature is powerful, as it enables investigators to perform exploratory analysis in how altering a physical component of the brain (e.g. the connection between two brain regions) will affect brain activity ([6]).

Second, MINDy provides a richer description of brain mechanisms than rsFC. While rsFC and MINDy both attempt to parameterize the connection strength between brain regions, MINDy also describes the local mechanisms that govern how each brain region behaves. Neural processes are thought to involve the combination of anatomically local computations and spatially-extended signal propagation, so it is important that descriptions of brain activity be able to define the degree to which this activity is generated by local vs. distributed mechanisms. Although the elements of the rsFC matrix are often interpreted as reflecting interregional components of neural processing, we have demonstrated that the rsFC is also sensitive to purely local characteristics of brain regions, such as their decay rate (SI Fig. 2D). Conversely, we have demonstrated that both the transfer function and decay rate of brain regions can also serve as reliable markers of individual differences and anatomical structure. By using MINDy, researchers can identify which neural mechanisms (i.e. specific MINDy parameters) give rise to functionally-relevant individual differences.

Lastly, MINDy greatly improves upon rsFC's characterizations of effective connectivity between brain regions. Unlike the elements of a correlation matrix, MINDy's weight parameters can describe asymmetric connectivity strengths and thus make inferences regarding the directional flow of activity between brain regions. We provide tantalizing illustration of the

potential utility of these types of findings (e.g., Fig. 2D, SI Fig. 3). Further, we demonstrated that MINDy may prove a more valid measure of brain connectivity and individual differences in connectivity than rsFC (Fig. 3), potentially because it is able to capture these connectivity asymmetries.

3.2 Relationship with other Models

There are currently two classes of generative models used to study fMRI: proxy-parameterized neural-mass models (e.g. models using diffusion-imaging data as a proxy for synaptic efficacy [6] [13]) and directly-parameterized linear models (e.g. Dynamic Causal Modeling; [5]). These two approaches represent opposite ends of a trade-off between realism and tractability for mesoscale human brain modeling. The first case (proxy-parameterized neural-mass models) excels in terms of interpretability of the model framework as the state-variables can always be traced back to population firing rates. These models operate at relatively fast time-scales and can produce predictions with the spatial resolution of MRI and the temporal resolution of EEG, which make them a parsimonious and general-purpose investigative tool that can be utilized across temporal scales. Current approaches in this respect are limited, however, due to the manner by which these models are parameterized. Even in state-of-the-art techniques (e.g. [7] [8]), most parameters are fixed a priori (local neural-mass parameters), or determined from diffusion imaging data, with only a limited subset taken from fMRI functional connectivity estimates, and then only at the group-average level. Thus, the vast majority of parameters are not sufficiently constrained by the relevant individual-level data, and instead are adapted from measurement of proxy variables, which is likely to limit the accuracy of model predictions. Diffusion imaging data, for instance is inherently unsigned and undirected, so the resultant models are unable to consider hierarchical connection schemes or long-distance connections that depress activity in the post-synaptic targets. Moreover, it remains unknown how to convert units from white matter volume to synaptic conductance even when these assumptions are met. In practice the conversion is performed by choosing a single scaling coefficient, which assumes that this relationship is linear with a universal slope. Due to these sources of uncertainty, proxy-parameterized models do not necessarily fit/predict raw functional time series. To be fair, however, this limitation may not be relevant for all scientific questions (e.g. studying long-term phenomena such as the structure of static FC, [10] [7] [8]).

By contrast, the ability of a model to fit the observed time-series implies that its predictions are valid within the vicinity of observed data. This property holds even when the underlying model is likely to be inaccurate in its long-term predictions. Evidence of this can be seen in the success of Dynamic Causal Modeling approaches, which can recreate task-driven activity ([33]) despite using a biologically implausible linear model. However, the downside of using a linear modelling framework is that the long-term predictions of these models are most likely inaccurate, given that brain activity in a linear model will always converge to a noise-driven stationary distribution. Thus, even though these models may be more accurate than forward-parameterized neural-mass models in short-term predictions (by virtue of fitting parameters), the linear form guarantees that they will be unable to capture the extended pattern of brain spatiotemporal dynamics. Analytically, it is known that linear dynamical systems cannot exhibit non-trivial deterministic dynamics and are characterized by a steady-state covariance when driven by noise (which can be calculated by solving a Lyapunov equation). For this reason, they are often employed as surrogate models for testing whether proposed measures of DFC can distinguish between noise-driven trivial (linear) dynamics and those observed in the data (e.g. [34]). Thus, DFC measures which have been shown to be non-spurious through surrogate methods cannot, by definition, be reproduced by a linear dynamical system with or without noise. Likewise, these models will also be limited in their ability to identify the neural mechanisms

underlying predictions. Since the model does not take a biologically plausible form, it remains unknown whether the coefficients fit to the linear models are the same as would be retrieved by fitting a more realistic model (e.g. neural mass model). That is not to say that the coefficients are uninterpretable; indeed, meaningful predictions have been made by inferring effective connectivity from the model coefficients (e.g. [5]). On the other hand, the models' simplicity may lead to nonlinear components of brain activity being mixed into the linear model estimates, just as we have shown that intrinsic dynamics influence FC estimates (SI Fig. 2A).

MINDy attempts to leverage the advantages from both approaches. Like current neural-mass models, MINDy employs a nonlinear dynamical systems framework, which is capable of generating long-term patterns of brain dynamics. However, MINDy is also a data-driven approach, in that models are fit from the ground-up using fMRI time-series rather than using surrogate measures such as diffusion imaging (although in principle, such information could be used to initialize or constrain MINDy parameter estimates). From the perspective of biologically-plausible models, MINDy extends parameter fitting from the relatively small number of local parameters that constitute the current state-of-the-art ([7] [8]) to fitting every parameter in biologically-plausible whole-brain mesoscale models (i.e., increasing the number of fitted parameters by orders of magnitudes). Likewise, MINDy extends data-driven modeling of resting-state data from linear models containing tens of nodes ([5]) to nonlinear models containing hundreds. It is also worth noting that the computational innovations made in the optimization process make MINDy parameterization many orders of magnitude faster than comparable techniques ([8] [5]) despite fitting many more parameters (e.g. over 176,000 free parameters can be estimated in a minute vs. several hours to fit hundreds of parameters). This efficiency has enabled us to interrogate the method's validity and sensitivity in ways that would probably not be computationally feasible for less efficient methods (e.g., building error distributions for sensitivity analyses in Sec. 2.2).

3.3 Comparison with Diffusion Imaging Seeded Neural Mass Models

Although we emphasize the ability to generate individualized brain models, previous studies using neural-mass models with weights seeded by diffusion imaging have emphasized predicting group-level data ([10] [7] [8]). Two recent studies fit free parameters with the explicit optimization objective of predicting the group-average rsFC matrix ([7] [8]). By contrast, MINDy seeks to predict the short-term evolution of the neural activity time series for single subjects, which often results in the simulated individual rsFC correlating highly with the empirical rsFC (Fig. 4C; SI Fig. 6). Averaging across simulated rsFC's produces a group-level simulated rsFC that correlates extremely well with the empirical group-average rsFC ($r(87{,}398)=.94$; see SI Fig. 6). As such, the group-average MINDy fit compares very favorably with the analogous measures for diffusion-parameterized models which typically don’t surpass $r=.6$, ([7] [8]).

3.4 Limitations

There are two primary limitations of MINDy. The first relates to the properties of fMRI data: MINDy is limited by the spatial and temporal resolution at which data is gathered. This means that MINDy is more sensitive to slow interactions that occur over larger spatial scales and is limited to predicting infraslow dynamics (as opposed to higher-frequency bands). Interactions that are more heterogeneous in time or space may also be missed by MINDy as the model assumes that the transfer function is monotone. While the strength of signaling between regions is allowed to vary according to the transfer function, the sign of signaling (inhibitory vs. excitatory)

is not. This feature is inherent in region-based modeling and so this limitation is not unique to MINDy.

A second limitation relates to the model used to specify MINDy. Unlike the conventional neural mass models (e.g. [7]), MINDy employs a single population rather than two or more local subpopulations of excitatory and inhibitory neurons. However, the use of a single population for modeling fMRI is not unique to MINDy (e.g. [13] [8]). The model does contain local competitive nonlinearities in that the decay term (D) competes with the recurrent connectivity of W but the precise mechanisms underlying these dynamics are not explicated. By comparison, multipopulation neural mass models contain separate terms for the interactions between subpopulations and the time-constants of firing rate within each subpopulation, both of which likely influence the local parameters of MINDy. Similarly, while MINDy can specify that the directed interaction between a pair of regions is positive, it cannot distinguish between excitatory projections onto an excitatory subpopulation and inhibitory projections onto an inhibitory subpopulation (both of which could be net excitatory; see SI Sec S.1.1).

3.5 Future Applications and Directions

We view MINDy models as providing a rich and fertile platform that can be used both for computationally-focused explorations, and as a tool to aid interrogation and analyses of experimental data.

The most immediate potential of MINDy is in providing new parameter estimates for studies of individual differences or biomarkers. There is also immediate potential for MINDy in model-driven discovery of resting-state dynamics (e.g. [13]), in which case MINDy simply replaces diffusion imaging as a method to parameterize neural mass models. The potential benefit of using MINDy over diffusion imaging is that MINDy identifies signed, directed connections in a data-driven manner which may improve realism. Going forward, other applications of MINDy may be in bridging the gap between resting-state characterizations of brain networks and evoked-response models of brain activity during task contexts. We envision two lines of future work in this domain: one in improving estimates of task-evoked effects, and the other concerning the effect of task contexts or cognitive states on brain activity dynamics.

3.6 Isolating Task-Evoked Signals

One future use of MINDy may be in improving estimates of task-related brain activity. Current methods of extracting task-related brain signals are based upon comparing BOLD time courses during windows of interest using generalized linear models. However, the effects of task conditions are related to both ongoing brain activity ([35]) and intrinsic network structure ([36]). Viewing the brain as a dynamical system, any input to the brain will have downstream consequences, so brain activity observed during task contexts likely contains some mixture of task-evoked activity and its interaction with spontaneous activity. Using MINDy, it may be possible to isolate task-evoked responses by subtracting out what would have been predicted to occur via the resting-state MINDy model. The resultant estimate for task-related activity would be the time-series of MINDy prediction errors (i.e. residuals), ideally adjusted for the model's error at rest. Thus, MINDy has the potential to improve estimates of task-evoked activity from fMRI data, although future validation is needed.

3.7 Illuminating Dynamics

Present results indicate that MINDy is able to replicate some features of infraslow brain-dynamics observed in the data (see Sec. 2.5). Although these slower frequency bands have been less studied in task-contexts, growing evidence implicates them in slowly evolving cognitive states such as states of consciousness ([37]). MINDy may benefit future studies in these domains by providing a formal model by which to identify the mechanisms underlying dynamical regimes. Moreover, MINDy may illuminate the behavioral significance of infraslow dynamics. Previous studies have found that timing of pre-cue brain activity and infraslow dynamics interact to predict behavioral performance ([38]), so future characterizations of task-activation may benefit from considering how exogeneous stimuli interact with endogenous neural processes. Generative models of resting-state brain activity may prove critical in these efforts by predicting how endogenous brain states modulate the effects of exogeneous perturbations.

3.8 Conclusion

We have developed a novel and powerful method for constructing whole-brain mesoscale models of individualized brain activity dynamics, termed MINDy. We demonstrate that MINDy models are valid, reliable, and robust, and thus represent an important advance towards the goal of personalized neuroscience. We provide initial illustrations of the potential power and promise of using MINDy models for experimental analysis and computational exploration. It is our hope that other investigators will make use of MINDy models to further test and explore the utility and validity of this approach. Towards that end, we have made MATLAB code and documentation for developing and testing MINDy models available via the primary author's GitHub.

Materials and Methods

4.1 Experimental Design

The objective of the current study was to rigorously validate a new approach for data-driven whole-brain modeling (MINDy). The study design consisted of numerical simulations to validate the accuracy of models with respect to a known ground-truth and empirical analyses of HCP resting-state data. The latter analyses were designed to test whether MINDy adds additional value in-practice and to quantify its performance in the presence of known experimental confounds (e.g. motion).

4.2 Dataset

Data consisted of resting state scans from 53 subjects in the Human Connectome Project (HCP) young adult cohort, 900 subject release (for acquisition and minimal preprocessing details, see [15]; WU-Minn Consortium). Each subject underwent two scanning sessions on separate days. Each scan session included two 15-minute resting-state runs for a total resting state scan time of 60 minutes (4800 TRs). The two runs for each session corresponded to acquisitions that had left-right and right-left phase-encoding directions (i.e., balanced to account for potential asymmetries in signal loss and distortion). The TR was 720ms and scanning was performed at 3T. The subjects were selected by starting with an initial pool of the first 150 subjects and then excluding subjects who had at least one run in which more than 1/3 of frames were censored (i.e. 400 bad frames out of 1200).

Although this criterion greatly decreased the number of usable subjects from the initial pool of 150 to 53 (attrition=65%), it should be noted that it is likely to be overly conservative. We employed such a strongly conservative criterion for this first-stage validation effort to provide the cleanest data from which to test the model. Likewise, we had the luxury of drawing upon a very large-sample dataset. In contrast, we believe that the exclusion criteria will not need to be as

conservative in a research setting for which model cross-validation is not performed on every subject (i.e., it is probably overly stringent to require that all four sessions be clean, since we only used two sessions at a time). In particular, the use of cross-validation required that two models be fit for every subject using disjoint data so that the validation required twice as much data as would normally be required. Moreover, we required that the data be uniformly clean so that we could parametrically vary the amount of data used (i.e. criteria were in terms of absolute cleanness for each scanning session rather than number of clean frames). However, there is no reason why the models could not be fit to only the clean segments of scanning sessions.

4.3 Preprocessing

Data were preprocessed through the rsFC pipeline proposed by Siegel and colleagues ([16]). The first stage of this pipeline is the HCP minimal pre-processing pipeline (see [15]) with FSL's ICA-FIX correction for rigid body motion ([17]). We then applied one of three second-stage pipelines developed by Siegel and colleagues ([16]), to test the effects of including various additional preprocessing steps. In all three pipelines, drift was mitigated by detrending data. The pipelines also all included motion scrubbing, using both Framewise Displacement (FD) and the temporal derivative of variation (DVARS). Frames that exceeded the cutoffs for FD (.2mm) or DVARS (5% above median) were replaced via linear interpolation ([39]). Respiratory artifact was mitigated with a 40th-order .06-.14 Hz band-stop filter applied to FD and DVARS for all pipelines ([16]).

The three second-stage pipeline variants differed however, in the number of regressors included to remove nuisance signals. The first variant mainly corrected frame-to-frame motion artifact, which has been found to induce systematic errors in functional connectivity studies, i.e. generating spurious short-distance correlations while diminishing long distance ones ([28]). In addition to data scrubbing, motion correction was performed using the 12 HCP motion regressors and their temporal derivatives. The second, more extensive pipeline variant, known as CompCor, also removed cardiac and respiratory signals, by additionally regressing out principal components of the white matter and cerebrospinal fluid signals ([29]). Lastly, the third pipeline variant also added global signal regression (GSR), in which the mean signals from white matter, cerebrospinal fluid, and grey matter are also included as regressors. As the variables included are cumulative, these three pipelines form a representative hierarchy of preprocessing approaches, that optionally includes CompCor or CompCor+GSR in addition to motion scrubbing.

After the second-stage preprocessing pipelines, we deconvolved the parcellated data (see below) with the generic SPM hemodynamic kernel ([31]) using the Weiner deconvolution. For the Weiner deconvolution, we used noise-power to signal-power parameter .02. The value of this parameter dictates the degree of temporal filtering during the deconvolution with smaller values being more parsimonious (less additional filtering). We then smoothed by convolving with the [.5 .5] kernel (2 point moving average) and z-scored the result. To test the robustness of the fitting procedure, we compared the effect of the second-stage preprocessing pipelines for some analyses. Based upon these results, we chose the third variant pipeline (GSR+CompCor+motion) for all other analyses. For all empirical rsFC analyses we use the deconvolved data to prevent bias from the deconvolution procedure in comparing MINDy and rsFC.

We defined derivatives in terms of finite differences. Since HCP employed unusually fast scanner TRs, we temporally downsampled the estimated derivatives for calculating goodness-of-fit in non-simulation analyses to represent the anticipated benefits to typical fMRI protocols and improve SNR: $dx_t \approx (x_{t+2}-x_t)/2$.

4.4 Parcellation Atlases

In the present framework we define whole-brain models in terms of connected neural populations. Thus, the approach demands that the neural populations be defined a-priori. For the present case of fMRI data, we define these populations to be anatomical brain regions corresponding to subcortical structures and cortical parcels. For subcortical regions, we follow the HCP protocol in considering 19 subcortical regions as defined by FreeSurfer. For cortical parcels, we generally employed the gradient-weighted Markov Random Field (gwMRF) parcellation with 200 parcels per hemisphere ([20]) with the 17-network organization. The gwMRF parcellation is optimized to align with both resting-state and task fMRI and has been found to demonstrate improved homogeneity within parcels relative to alternative parcellation techniques. However, for anatomical analyses we compared with an additional atlases (SI Fig. 1C,G) to ensure generality : the MMP atlas ([21]) which was also derived from a combination of rest and task-based data. The MMP (Multi-Modal Parcellation) atlas is symmetric with 180 parcels per hemisphere.

4.5 Compensating for Regularization Bias

In order to retrieve parsimonious weight matrices and reduce overfitting, we employed regularization to each weight matrix (both the sparse and the low-rank matrices) during the fitting process. One consequence of regularization, however, is that the fitted weights may be unnecessarily small as weight magnitudes are penalized. After fitting, we therefore performed a global rescaling of weight and decay contributions for each model using robust regression as implemented by MATLAB2018a. Specifically, we fit two scalar parameters: p_W, p_D in regressing $dx_t = p_W \psi(x_t) - p_D D x_t$ collapsed across all parcels. Here p_W and p_D represent global rescaling coefficients for the weights and decay, respectively. As this compensating step only used global rescaling for W and D, it had no effect upon the relative values for each parcel, only the total magnitude of the W and D components. Since only two values are estimated, this step does not reintroduce overfitting.

4.6 Selecting SGD Hyperparameters and Initialization

The proposed fitting procedure requires two sets of hyperparameters: the four regularization terms specific to our procedure and the four NADAM parameters ([14]). By "hyperparameters" we refer to free constants within an algorithm which distinguishes them from the "parameters" of an individualized model.

Hyperparameters were hand-selected for model goodness-of-fit and reliability, based upon prior numerical exploration with a subset of 10 subjects who did not belong to the "data source" subjects. Thus, these subjects were not included in any further analyses so the hyperparameter selection procedure did not artificially inflate model performance. The chosen set of hyperparameters was then constant for all test subjects. Hyperparameter values and discussion are included in SI (Table 4). The initialization distributions for the algorithm were similarly selected using the same subjects and are included in the SI (Table 4).

4.7 Realistic Whole-Brain Simulations

For the analyses of sensitivity and individual differences we generated new, synthetic individuals by randomly sampling model parameters from the parameter distributions estimated from the full dataset (i.e. N=53 participants). The decay and curvature parameters (α, D) were independently sampled for each parcel from that parcel's population distribution. The weight matrices, however, were sampled as a whole rather than sampling each individual connection as we found that the latter approach led to pathological behavior in simulations. For the robustness analyses, ground truth models were drawn from those fit to experimental sessions. The ground-truth models were simulated as stochastic differential equations ($dx = f(x)dt + \sigma_W dWt$) with $f(x)$ the

deterministic MINDy model and units time measured in terms of the fMRI TR. Models were Euler-Maruyama integrated with dt=1/4 and σ_W=.45 in units TR (720 ms) to generate simulated neural activity time-series. Neural-activity was then downsampled to 1 TR resolution (as opposed to the simulation's time-step of dt×TR) and convolved with the HRF kernel ([31]):

$$h(t;\alpha_{\{1,2\}},\beta_{\{1,2\}},c) := \frac{t^{\alpha_1-1}e^{-\beta_1 t}\beta_1^{\alpha_1}}{\Gamma(\alpha_1)} - \frac{t^{\alpha_2}e^{-\beta_2 t}\beta_2^{\alpha_2}}{c\Gamma(\alpha_2)} \quad \textbf{(5)}$$

Here Γ is the gamma function (equal to factorial for integer values). The parameters describe two gamma-distributions (one α, β pair per distribution) and a mixing coefficient (c) to generate a digamma function. Parameters were set to their default values (α_1 =6, α_2=16, β_1 =1, β_2=1, c=1/6) except for the simulation featuring HRF variability in empirically-parameterized models. In this case, random perturbations were added to each parameter and were drawn from the normal distribution with mean zero and SD as indicated. The final simulated BOLD signal was then generated by adding white, gaussian noise with the indicated SD (Fig. 1D).

4.8 Randomized Network Simulations

To prevent circularity, we also tested MINDy's ability to invert simulated networks drawn from random distributions independent of previous analyses. These distributions were designed to possess complex network structures by superimposing three simpler network structures: community-structure (M_1), sparse structure (M_2), and low rank structure (M_3). These distributions are characterized by standard-deviation parameters σ_1 and σ_2. An asymmetry parameter σ_a characterizes the degree to which the resultant network is asymmetric. Each of these parameters were randomly sampled on each instance from normal distributions: σ_1, $\sigma_a \sim N(4, .05^2)$ and $\sigma_2 \sim N(3, .05^2)$. Connectivity matrices were then randomly parameterized as follows:

$$\begin{aligned} M_1 &\sim [N(0,1/\sigma_1^2) + N(0,1/\sigma_1^2)^3]_{n/p\times n/p} \\ M_2 &\sim [N(0,1/\sigma_2^2)^3]_{n\times n} \\ M_3 &\sim [N(0,1/\sigma_1^2) + N(0,1/\sigma_1^2)^3]_{n\times k} \times [N(0,1/\sigma_1^2) + N(0,1/\sigma_1^2)^3]_{k\times n} \end{aligned} \quad \textbf{(6)}$$

Here, the bracket's subscript outside each matrix denotes its size with n=40 denoting the total number of nodes, q denoting the number of nodes per community (randomly set to either 1 or 2 with equal probability), and k=5 denoting the rank of the low-rank component. We denote the Kronecker product $\otimes$ and use it to copy the community level matrix (M_1) among each node belonging to the community: $M_1 := 1_{q\times q} \otimes M_1$. The three component matrices are then combined as follows:

$$Q = \hat{M}_1 + M_2 + M_3, \quad \hat{Q} = (Q + (Q - Q^T)/\sigma_a) \quad \textbf{(7)}$$

The final matrix C is formed by censoring elements of $\hat{Q}$ whose absolute value is below 1/4 the standard deviation of $\hat{Q}$. This same technique was used to randomly generate asymmetric continuous Hopfield models ([40]) with homogeneous, heterogeneous, or nonlinear hemodynamic effects and realistically-parameterized neural mass models with nonlinear hemodynamics. The Hopfield models are similar in form to the MINDy model, but use a tanh transfer function:

$$dx = \left(Wtanh(b_0 \circ x) - D \circ x\right)dt + \sigma_W dW_t \quad \textbf{(8)}$$

Here, the slope vector $b_0 \in \mathbb{R}^n \sim N(6, (.5)^2)$ and diagonal elements of the decay matrix D were drawn from $N(.4, (.1)^2)$ (non-diagonal elements are zero). As elsewhere, the symbol '∘' denotes the Hadamard product (element-wise multiplication). Models were simulated via Euler-Maruyamma integration with dt=.1s, σ_W =.2, TR=.7s, and total simulation length t=10,000. We considered the case in which no hemodynamics are present, in which case MINDy is fed x_t downsampled according to TR, and the case in which x_t is convolved with spatially heterogeneous hemodynamics and deconvolved with the canonical HRF before being fit by MINDy. In the latter case, the HRF function was parameterized as before, but with the ground-truth α_1 parameter for each brain region drawn from $N(6, (.5)^2)$ and the β_1 parameter drawn from $N(1, (.5/6)^2)$. The simulated BOLD was produced by convolving the simulated time-series with the ground-truth HRF before temporal downsampling. In both cases, initial conditions for each node were independently drawn from $N(0,1)$ and the first 100 samples were dropped. Since the total number of nodes was approximately one-tenth of those used in the HCP data, we rescaled the dimension of the low-rank component by one-tenth (from 150 to 15). Similarly, we rescaled the regularization terms inversely proportionate to the effect of rescaling W by a factor of 10: (λ_1,λ_3 by 1/10, λ_2 by $1/\sqrt{(10)}$ and λ_4 by $1/10^2$). For simulations using the Balloon-Windkessel model of hemodynamics, x_t was rescaled to the range of average synaptic gating via $5S_t=1+\tanh(x_t/10)$. This transformation of x_t into S_t was then substituted into the nonlinear hemodynamic model (below) to generate simulated BOLD signal. In all cases, time series were z-scored, smoothed via nearest-neighbor ([.5 .5] kernel) and run through MINDy for 150,000 iterations (approximately 70 seconds) with the original batch size of 250.

4.9 Neural Mass and Windkessel-Balloon Model Simulation

For our biophysical neural-mass simulation we largely followed the approach of Wang and colleagues ([8]) in using single-population neural mass models (20 masses/simulation) with Windkessel-Balloon model hemodynamics ([32]). Similar to the MINDy model, the neural mass model ([13]) contains a monotone nonlinearity ($\hat{\psi}$) and linear decay via τ_S^{-1}.

$$\begin{aligned} dS_i &= [-S_i / \tau_S + r(1-S_i)H(x_i)]dt + \sigma_W dW_t \\ H(x_i) &= \hat{\psi}(-d(ax_i - b)) \\ x_i &= \sum_j C_{i,j} S_j + I_{Sub} \end{aligned} \qquad \textbf{(9)}$$

The variable S describes the average synaptic gating, while H describes the population firing-rate. We used the default parameter settings: τ_S =.1s, a=270n/C, b=108 Hz, d=.154s, r=.641. Unlike Wang and colleagues ([8]), we used a logistic sigmoid transfer function for $\hat{\psi}(x) = 1/(1+exp(-x))$ instead of the rectified linear transfer function: x/(1-exp(-x)), as the former is less prone to pathological behavior in random networks. Subcortical input was I_{Sub}=5. Connection weight matrices were randomly generated as described in the previous section, but with 1.5 added to all recurrent connections and the resultant matrix scaled by a factor of 100. Simulated neural activity is converted into BOLD signal through the Windkessel-Balloon model ([32]):

$$
\begin{aligned}
\frac{z_i}{dt} &= S_i - \kappa z_i - \gamma(f_i - 1) \\
\frac{f_i}{dt} &= z_i \\
\tau \frac{v_i}{dt} &= f_i - v_i^{\alpha_G^{-1}} \\
\tau \frac{q_i}{dt} &= \frac{f_i}{\rho}\left[1-(1-\rho)^{1/f_i}\right] - q_i v_i^{\alpha_G^{-1}-1}
\end{aligned}
\qquad \textbf{(10)}
$$

The variables z, f, v, and q model vasodilation, inflow, blood volume, and deoxyhemoglobin content, respectively. Parameters were: ρ=.34, κ=.65s^{-1}, γ=.41s^{-1}, τ=.98s, α_G=.32. The simulated BOLD signal at each TR is then modeled as:

$$
BOLD(v_i, q_i) = V_0\left[k_1(1-q_i) + k_2\left(1-\frac{q_i}{v_i}\right) + k_3(1-v_i)\right] \qquad \textbf{(11)}
$$

Resting blood volume fraction is denoted V_0=.02. Scanning parameters k_1, k_2, k_3 were set to 3T values according to Demirtas and colleagues ([7]): k_1=3.72, k_2=.53 k_3=.53. Simulations were run with dt=25ms and σ_W=.005 for total length t=40,000. Sampling was performed every 29 time-steps (TR=725ms) and the first 10% of samples were dropped. The resulting time-series were deconvolved with the canonical HRF assumed by MINDy and z-scored. MINDy hyperparameters were identical to the rate-model case and MINDy was run for 10,000 iterations (approximately 6 seconds) with batch size 250. Initial conditions for hemodynamic variables were randomly sampled from the absolute normal distribution $|N(0,1)|$. Initial conditions for the neural variable (S) were generated by first sampling $S_0 \sim |N(0,1)|$ and then performing the transformation $S=S_0/(1+S_0)$.

4.10 Simulations for dFC Analyses

For analyses of dynamic functional connectivity, we used the same time-scale as the validation models (.5 TR). However, whereas the validation simulations employed process noise containing constant variance across parcels, we used a naïve estimate of process noise for each parcel, that was based upon the residual error of model fits over subsequent time-steps. We avoided doing this in the validation stage so that ground-truth parameters could not be recovered simply by observing noise. The residual error covaried with the decay parameter across parcels at the group-level, but not at the individual level, despite individual differences in both noise and decay being reliable within parcel. We reintroduced parcel-based variation into the dFC simulations to obtain maximum realism. We considered both the case in which process noise was allowed to vary by parcel but not by individual within a test-retest group (e.g. using the mean noise across subjects for each session separately), as well as the case in which process noise was determined on a subject-wise basis. Results obtained with either method were near-identical for the dFC reliability analyses, so we present results using the session-wise group-mean process noise (e.g. the mean process noise for each parcel averaged across all day 1 scans or all day 2 scans). Initial conditions were drawn from each subject's observed data for that scanning session. Simulations were run for 2600 time steps (1300 TRs) using 15 different initial conditions per session and temporally downsampled back to the scanning TR. After simulation, we downsampled from the 400 parcel to the 100 cortical parcel variants of gwMRF ([20]) and

removed subcortical ROIs in order to reduce computational complexity of subsequent dFC analyses.

4.11 dFC Analyses

DFC analyses consisted of the standard deviation and excursion of the time-varying correlation between brain regions. To calculate time-varying correlations we used Dynamic Conditional Correlation (DCC; [24]). To ensure numerical stability, we repeated the DCC algorithm 10 times per case (simulation or true data) and used the median estimated time-series for time-varying correlations. The excursion measure was calculated according to ([26]). Reliability was computed for each pair of regions' dFC statistics using the Pearson's ICC of group-demeaned dFC metrics between scanning session. Overall reliabilities collapsed across all regions were calculated using Image Intraclass Correlation ([27]).

4.12 Sensitivity Analyses

For the sensitivity analyses we first simulated a resampled individual multiple times to generate a distribution of trial-to-trial variability ("within-subject") in elements of MINDy's weight matrix and the rsFC matrix. We then held the weights of the ground-truth neural mass model constant while resampling either the curvature (α) or decay (D) parameters and calculating MINDy weights and rsFC from simulations of the new model. Changes in the estimated connectivity (weights or rsFC) were deemed significant if they occurred with $p<.05$ for the corresponding "within-subject" distribution.

4.13 Statistical Analyses

Statistical testing was primarily within-subject between method/condition (e.g. paired t-tests). We used the conservative Bonferroni method for all multiple-comparison corrections. All reported p-values are calculated for two-tailed tests unless indicated otherwise. We use $p\approx 0$ to denote p-values calculated as less than 10^{-20} for which precise numerical estimates may deteriorate.

Bibliography

[1] H. Markram, K. Meier, T. Lippert, S. Grillner, R. Frackowiak, S. Dehaene, A. Knoll, H. Sompolinsky, K. Verstreken, J. DeFelipe, S. Grant, J.-P. Changeux and A. Saria, "Introducing the Human Brain Project," *Procedia Computer Science,* vol. 7, pp. 39-42, 2011.

[2] D. C. Van Essen, S. M. Smith, D. M. Barch, T. E. J. Behrens, E. Yacoub and K. Ugurbil, "The WU-Minn Human Connectome Project: An overview," *NeuroImage,* vol. 80, pp. 62-79, 2013.

[3] B. Biswal, F. Z. Yetkin, V. M. Haughton and J. S. Hyde, "Functional connectivity in the motor cortex of resting human brain using echoplanar mri," *Magnetic Resonance in Medicine,* vol. 34, pp. 537-541, 1995.

[4] R. L. Buckner, F. M. Krienen and B. T. T. Yeo, "Opportunities and limitations of intrinsic functional connectivity MRI," *Nature Neuroscience,* vol. 16, pp. 832-837, 2013.

[5] A. Razi, M. L. Seghier, Y. Zhou and P. McColgan, "Large-scale DCMs for resting-state fMRI," *Network Neuroscience,* vol. 1, 2017.

[6] V. K. Jirsa, T. Proix, D. Perdikis, M. M. Woodman, H. Wang, J. Gonzalez-Martinez, C. Bernard, C. Bénar, M. Guye, P. Chauvel and F. Bartolomei, "The Virtual Epileptic Patient: Individualized whole-brain models of epilepsy spread," *NeuroImage,* vol. 145, pp. 377-388, 2017.

[7] M. Demirtaş, J. B. Burt, M. Helmer, J. L. Ji, B. D. Adkinson, M. F. Glasser, D. C. Van Essen, S. N. Sotiropoulos, A. Anticevic and J. D. Murray, "Hierarchical Heterogeneity across Human Cortex Shapes Large-Scale Neural Dynamics," *Neuron,* vol. 101, pp. 1181-1194, 2019.

[8] P. Wang, R. Kong, X. Kong, R. Liégeois, C. Orban, G. Deco, M. P. Heuvel and B. T. Thomas Yeo, "Inversion of a large-scale circuit model reveals a cortical hierarchy in the dynamic resting human brain," *Science Advances,* vol. 5, 2019.

[9] C. Tu, R. P. Rocha, M. Corbetta, S. Zampieri, M. Zorzi and S. Suweis, "Warnings and caveats in brain controllability," *NeuroImage,* vol. 176, pp. 83-91, 2018.

[10] C. J. Honey, O. Sporns, L. Cammoun, X. Gigandet, J. P. Thiran, R. Meuli and P. Hagmann, "Predicting human resting-state functional connectivity from structural connectivity," *Proceedings of the National Academy of Sciences,* vol. 106, pp. 2035-2040, 2009.

[11] S. Knock, A. McIntosh, O. Sporns, R. Kötter, P. Hagmann and V. Jirsa, "The effects of physiologically plausible connectivity structure on local and global dynamics in large scale brain models," *Journal of Neuroscience Methods,* vol. 183, no. 1, pp. 86-94, 2009.

[12] T. D. Satterthwaite, C. H. Xia and D. S. Bassett, "Personalized Neuroscience: Common and Individual-Specific Features in Functional Brain Networks," *Neuron,* vol. 98, pp. 243-245, 2018.

[13] E. C. A. Hansen, D. Battaglia, A. Spiegler, G. Deco and V. K. Jirsa, "Functional connectivity dynamics: Modeling the switching behavior of the resting state," *NeuroImage,* vol. 105, pp. 525-535, 2015.

[14] T. Dozat, "Incorporating Nesterov momentum into Adam," *Proceedings of 4th International Conference on Learning Representations, Workshop Track, 2016,* 2016.

[15] M. F. Glasser, S. N. Sotiropoulos, J. A. Wilson, T. S. Coalson, B. Fischl, J. L. Andersson, J. Xu, S. Jbabdi, M. Webster, J. R. Polimeni, D. C. V. Essen and M. Jenkinson, "The minimal preprocessing pipelines for the Human Connectome Project," *NeuroImage,* vol. 80, pp. 105-124, 2013.

[16] J. S. Siegel, A. Mitra, T. O. Laumann, B. A. Seitzman, M. Raichle, M. Corbetta and A. Z. Snyder, "Data Quality Influences Observed Links Between Functional Connectivity and Behavior," *Cerebral Cortex,* vol. 27, pp. 4492-4502, 2017.

[17] G. Salimi-Khorshidi, G. Douaud, C. F. Beckmann, M. F. Glasser, L. Griffanti and S. M. Smith, "Automatic denoising of functional MRI data: Combining independent component analysis and hierarchical fusion of classifiers," *NeuroImage,* vol. 90, pp. 449-468, 2014.

[18] E. M. Gordon, T. O. Laumann, A. W. Gilmore, D. J. Newbold, D. J. Greene, J. J. Berg, M. Ortega, C. Hoyt-Drazen, C. Gratton, H. Sun, J. M. Hampton, R. S. Coalson, A. L. Nguyen, K. B. McDermott, J. S. Shimony, A. Z. Snyder, B. L. Schlaggar, S. E. Petersen, S. M. Nelson and N. U. F. Dosenbach, "Precision Functional Mapping of Individual Human Brains," *Neuron,* vol. 95, pp. 791 - 807.e7, 2017.

[19] E. S. Finn, X. Shen, D. Scheinost, M. D. Rosenberg, J. Huang, M. M. Chun, X. Papademetris and R. T. Constable, "Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity," *Nature Neureoscience,* vol. 18, pp. 1664-1671, 2015.

[20] A. Schaefer, R. Kong, E. M. Gordon, T. O. Laumann, X.-N. Zuo, A. J. Holmes, S. B. Eickhoff and B. T. T. Yeo, "Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI," *Cerebral Cortex,* pp. 1-20, 2017.

[21] M. F. Glasser, T. S. Coalson, E. C. Robinson, C. D. Hacker, E. Harwell, K. Ugurbil, J. Andersson, C. F. Beckmann, M. Jenkinson, S. M. Smith and D. C. Van Essen, "A multi-modal parcellation of human cerebral cortex," *Nature,* vol. 536, pp. 171-178, 2016.

[22] A. D. Friederici and S. M. E. Gierhan, "The language network," *Current Opinion in Neurobiology,* vol. 23, pp. 250-254, 2013.

[23] E. A. Allen, E. Damaraju, S. M. Plis, E. B. Erhardt, T. Eichele and V. D. Calhoun, "Tracking Whole-Brain Connectivity Dynamics in the Resting State," *Cerebral Cortex,* vol. 24, pp. 663-676, 2014.

[24] R. Engle, "Dynamic Conditional Correlation," *Journal of Business & Economic Statistics,* vol. 20, pp. 339-350, 2002.

[25] A. S. Choe, M. B. Nebel, A. D. Barber, J. R. Cohen, Y. Xu, J. J. Pekar, B. Caffo and M. A. Lindquist, "Comparing test-retest reliability of dynamic functional connectivity methods," *NeuroImage,* vol. 158, pp. 155-175, 2017.

[26] C. Zhang, S. A. Baum, V. R. Adduru, B. B. Biswal and A. M. Michael, "Test-retest reliability of dynamic functional connectivity in resting state fMRI," *NeuroImage,* vol. 183, pp. 907-918, 2018.

[27] H. Shou, A. Eloyan, S. Lee, V. Zipunnikov, A. N. Crainiceanu, M. B. Nebel, B. Caffo, M. A. Lindquist and C. M. Crainiceanu, "Quantifying the reliability of image replication studies: The image intraclass correlation coefficient (I2C2)," *Cognitive, Affective, & Behavioral Neuroscience,* vol. 13, pp. 714-724, 2013.

[28] J. D. Power, K. A. Barnes, A. Z. Snyder, B. L. Schlaggar and S. E. Petersen, "Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion," *NeuroImage,* vol. 59, pp. 2142-2154, 2012.

[29] Y. Behzadi, K. Restom, J. Liau and T. T. Liu, "A component based noise correction method (CompCor) for BOLD and perfusion based fMRI," *NeuroImage,* vol. 37, pp. 90-101, 2007.

[30] G. K. Aguirre, E. Zarahn and M. D'Esposito, "The Variability of Human, BOLD Hemodynamic Responses," *NeuroImage,* vol. 8, pp. 360-369, 1998.

[31] K. J. Friston, P. Fletcher, O. Josephs, A. Holmes, M. D. Rugg and R. Turner, "Event-Related fMRI: Characterizing Differential Responses," *NeuroImage,* vol. 7, pp. 30-40, 1998.

[32] K. J. Friston, A. Mechelli, R. Turner and C. J. Price, "Nonlinear Responses in fMRI: The Balloon Model, Volterra Kernels, and Other Hemodynamics," *NeuroImage,* vol. 12, pp. 466-477, 2000.

[33] K. J. Friston, K. H. Preller, C. Mathys, H. Cagnan, J. Heinzle, A. Razi and P. Zeidman, "Dynamic Causal Modeling Revisited," *NeuroImage,* vol. 199, pp. 730-744, 2019.

[34] T. O. Laumann, A. Z. Snyder, A. Mitra, E. M. Gordon, C. Gratton, B. Adeyemo, A. W. Gilmore, S. M. Nelson, J. J. Berg, D. J. Greene, J. E. McCarthy, E. Tagliazucchi, H. Laufs, B. L. Schlaggar, N. U. F. Dosenbach and S. E. Petersen, "On the Stability of BOLD fMRI Correlations," *Cerebral Cortex,* vol. 27, pp. 4719-4732, 2017.

[35] B. J. He, "Spontaneous and Task-Evoked Brain Activity Negatively Interact," *Journal of Neuroscience,* vol. 33, pp. 4672-4682, 2013.

[36] M. W. Cole, T. Ito, D. S. Bassett and D. H. Schultz, "Activity flow over resting-state networks shapes cognitive task activations," *Nature Neuroscience,* vol. 19, pp. 1718-1726, 2016.

[37] A. Mitra, A. Z. Snyder, E. Tagliazucchi, H. Laufs and M. E. Raichle, "Propagated infra-slow intrinsic brain activity reorganizes across wake and slow wave sleep," *eLife,* vol. 4, p. e10781, 2015.

[38] M. D. Fox, A. Z. Snyder, J. L. Vincent and M. E. Raichle, "Intrinsic Fluctuations within Cortical Systems Account for Intertrial Variability in Human Behavior," *Neuron,* vol. 56, pp. 171-184, 2007.

[39] J. D. Power, B. L. Schlaggar and S. E. Petersen, "Recent progress and outstanding issues in motion correction in resting state fMRI," *NeuroImage,* vol. 105, pp. 536-551, 2015.

[40] J. J. Hopfield, "Neurons with graded response have collective computational properties like those of two-state neurons," *Proceedings of the National Academy of Sciences,* vol. 81, pp. 3088-3092, 1984.

[41] T. Schreiber, "Measuring Information Transfer," *Phys. Rev. Lett.,* vol. 85, no. 2, pp. 461-464, 2000.

[42] A. Schroeter, F. Schlegel, A. Seuwen, J. Grandjean and M. Rudin, "Specificity of stimulus-evoked fMRI responses in the mouse: The influence of systemic physiological changes associated with innocuous stimulation under four different anesthetics," *NeuroImage,* vol. 94, pp. 372-384, 2014.

[43] F. Schlegel, A. Schroeter and M. Rudin, "The hemodynamic response to somatosensory stimulation in mice depends on the anesthetic used: Implications on analysis of mouse fMRI data," *NeuroImage,* vol. 116, pp. 40-49, 2015.

[44] N. Todd, S. Moeller, E. J. Auerbach, E. Yacoub, G. Flandin and N. Weiskopf, "Evaluation of 2D multiband EPI imaging for high-resolution, whole-brain, task-based fMRI studies at 3T: Sensitivity and slice leakage artifacts," *NeuroImage,* vol. 124, pp. 32-42, 2016.

[45] H. R. Wilson and J. D. Cowan, "Excitatory and inhibitory interactions in localized populations of model neurons," *Biophysical Journal,* vol. 12, pp. 1-24, 1972.

[46] G. Deco, V. K. Jirsa and A. R. McIntosh, "Emerging concepts for the dynamical organization of resting-state activity in the brain," *Nature Reviews Neuroscience,* vol. 12, pp. 43-56, 2011.

[47] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *CoRR,* vol. abs/1412.6980, 2014.

[48] A. R. Aron, T. W. Robbins and R. A. Poldrack, "Inhibition and the right inferior frontal cortex," *Trends in Cognitive Sciences,* vol. 8, pp. 170-177, 2004.

[49] D. Swick, V. Ashley and A. U. Turken, "Left inferior frontal gyrus is critical for response inhibition," *BMC Neuroscience,* vol. 9, p. 102, 2008.

Acknowledgements

MS was funded by NSF-DGE-1143954 from the US National Science Foundation. TB acknowledges R37 MH066078 from the US National Institute of Health. SC holds a Career Award at the Scientific Interface from the Burroughs-Wellcome Fund. Portions of this work were supported by AFOSR 15RT0189, NSF ECCS 1509342 and NSF CMMI 1537015, NSF NCS-FO 1835209 and NIMH Administrative Supplement MH066078-15S1 from the US Air Force Office of Scientific Research, US National Science Foundation, and US National Institute of Mental Health, respectively.

H2: Supplementary Materials

S.1 Interpreting Model Parameters

S.1.1 Interpreting Model Weights

In our model, the connectivity matrix defines the causal ability of mean regional activity in the sender region to monotonically change mean regional activity levels in the receiving region within a specific time window. This causal influence has standardly been termed effective connectivity within the fMRI (and EEG) literature. More precisely, however, in the model the effects must begin within the duration of one TR (720ms in our case), and last long enough to invoke a metabolic response. As such, our definition is slightly more specific than the notion of effective connectivity, as we specify that these relations must be weakly monotone: all else being equal, increasing (decreasing) the activity of region A will never decrease (increase) the activity of region B. We use the term weakly monotone as regions may exhibit saturated activity within our model and thus have little room to increase/decrease. In contrast with our definition, effective connectivity does not specify the nature of the relationship between regions. In this case, model-free methods such as transfer entropy ([41]) can be employed to study non-monotone relationships within a very small number of dimensions. We also define our temporal range of interactions to be between 500ms and 2s. We do not use a tighter temporal range such as 500ms-1s as temporal variations inherent in BOLD imaging, such as physiological changes in the hemodynamic response (e.g. under anesthesia; [42] [43]) lead to some uncertainty in timing. In addition, there are methodological limitations inherent in rapid acquisition methods such as multiband imaging ([44]), which have led some investigators to prefer TR's closer to 2s. In either case, our definition limits the duration of interest to the order of a typical fMRI trial.

The monotone and temporal constraints can also differentiate our W matrix from structural connectivity, the latter of which does not necessarily reflect how regions interact. If two regions communicate in a very heterogeneous manner and/or these interactions only result in very transient changes, these regions would not be connected in our W matrix, even if a direct white matter tract linked them. Of course, this scenario also suggests that those portions of brain would also not meet the definition of a cortical parcel due to their heterogeneity. Finer parcellation schemes lead to correspondingly more homogeneous "regions", so, with a sufficiently high resolution parcellation, we expect that most forms of structural connectivity would meet the monotone requirement, with a single cell as the theoretical limiting case. In summary, our form of connectivity in the W matrix describes not just the ability of regions to causally influence each other but to do so with easily predictable (monotone) consequences in a specific time scale. For ease of presentation, however, we use the term "effective connectivity" to refer to this matrix and also make connection with existing terminology.

S.1.2 Interpreting Model Curvature

In the original theory of neural mass models ([45] [46]), the decay-term and transfer function were meant to capture phenomenological components of the individual population without corresponding to a singular biological feature. For instance, the transfer function of neural mass models is usually derived from the probability of neuronal spiking as a function of excitation. If cells within each population are assumed homogeneous, the population level activity is proportional to the individual spiking probability when refractory periods are negligible. Under this homogeneity assumption, inter-parcel variation in the transfer function slope would directly reflect variation in the cellular spiking probability between parcels. For cortical neurons with low-firing rates at rest, the spiking threshold is essentially constant (unlike bursting cells for instance), so a high slope might be interpreted as low noise. Since the ground-truth relation with excitation is binary for each cell at a given time ("all or none" spiking), all deviations from that relation must be due to variations in how much of the population-level excitation each neuron receives.

However, if we instead allow parcels to be internally heterogeneous, the transfer function slope parameter may indicate heterogeneities in either the spiking threshold or how excitation is distributed within the parcel. For a simple leaky integrate-and-fire model of neurons the individual transfer functions are binary (infinite slope). However, as the variation in firing thresholds between cells increases, the cumulative probability of population spike count becomes more graded corresponding to the sum of binary functions with different thresholds. Other sources of variation such as noise or inhomogeneities in projections to cells within the population would have a similar effect. Thus, although the exact source of variation (i.e., between regions or individuals) in the transfer function slope is unknown, a likely contributor is the degree of within-parcel variation, which may be due to inhomogeneities in internal/external inputs or neuronal dynamics.

However, there are at least three other potential physiological influences in transfer function slope. The first is the relationship between neuronal activity and the BOLD response. The neural components of the BOLD signal are more closely related to synaptic activity than neuronal spiking, so the likelihood of synaptic activity achieving a spike may also be a factor. For instance, for a given number of excitatory synaptic events, the likelihood of the post-synaptic cell firing generally increases with the synchrony of these events. Thus, the degree of synchronization could be another factor in the transfer function slope with parcels having greater synchronization of excitatory inputs having a higher slope. Alternatively, variation in neurovascular coupling between regions may affect the relationship. Regions with less predictable or less uniform hemodynamics would likely receive a lower transfer function slope similar to the case of neuronal variation. In this case, however, the lower slope results from uncertainty in observations rather than variation ("noise") within the generative system.

A final factor may be the intrinsic dynamics of each population. As the BOLD-based observations are temporally coarse (i.e., low resolution), the activity level of each population is more reflective of the average level of synaptic activity over hundreds of milliseconds. Thus, the transfer function seeks to relate the sum of parcel output over hundreds of milliseconds to the sum of parcel input (internal and external) over hundreds of milliseconds. Populations with more temporal integration (better "memory") are less sensitive to variation in input timing so transfer function slope might also increase with parcel memory. However, results actually indicated the opposite: parcels with greater slopes consistently had parameters reflecting less temporal integration (larger decay; see Results). Temporal integration within our model is reflected by the decay parameter, with high decay indicating less temporal integration.

S.1.3 Interpreting Model Decay

For neural mass models, the decay term describes how quickly a homogeneous population returns to its baseline level of activity. It is assumed that, in the absence of external inputs, the time course will be exponential, leading to the linear term -*D*x. Many cellular models also contain a linear decay term corresponding to the leak current, with *D* equal to the membrane time constant. At the population level, however, the decay term cannot be easily related to any biophysically comparable parameter, e.g., leak potassium conductance. Instead, the decay parameter should be considered as a phenomenological fit to the general pattern of homogeneous populations returning to some baseline rest level. In the current model, however, we relax the assumption of linear decay by also allowing “self-connections” in the connectivity matrix. That is not to say that the individual population members (neurons) contain autoconnections, but that by allowing both a nonlinear term and a linear term we allow a greater range of possible intrinsic dynamics including self-excitation at the population level. When the model is fit to the HCP data, all individuals were found to display nonnegative values for the nonlinear self-interaction. The resultant intrinsic dynamics for each isolated parcel consist of a nonlinear self-excitation and a linear self-inhibition which can lead to either a single stable equilibrium (near the mean BOLD signal) or bistability wherein initial conditions sufficiently above the mean will all converge to one equilibrium and those sufficiently below the mean converge to another. The bistable case generally results when the maximal slope of the self-excitatory component is larger than the decay term. In general, we expect that the decay parameter is related to the relative proportions of local excitation/inhibition within each parcel. The anatomical distribution of decay terms across parcels was highly consistent across subjects (Fig. 2D).

S.2 Accelerated Stochastic Gradients through NADAM

To fit the models, we use a variant of the stochastic-gradient descent (SGD) method: NADAM (Nesterov-accelerated adaptive moment estimation [14]) which builds upon the earlier ADAM algorithm ([47]). Gradient descent methods are algorithms that attempt to minimize a cost function, by updating parameters based upon the cost function's current slope (gradient). For an error function *E* and a parameter θ, the original gradient descent algorithm updates the estimate of the parameter (denoted θ_k) at each iteration (k) of the algorithm according to:

$$\theta_{k+1} = \theta_k - \eta \frac{\partial E}{\partial \theta_k} \qquad \textbf{(S.1)}$$

In which η is the user-chosen learning rate parameter. Although highly efficient, gradient descent algorithms are not guaranteed to reach a global minimum for non-convex problems; further, the original gradient-descent method is prone to getting "trapped" in local minima. Additionally, global-minima of highly non-convex problems may not be desirable as they sometimes poorly generalize.

Since the development of first-generation gradient-descent algorithms, substantial progress has been made in generalizing the method to handle non-convex surfaces, often by adding a “momentum” term. Momentum in SGD makes the system’s evolution a function of not only the current gradient, but also past gradients. Like physical momentum, this memory allows the algorithm to ``roll past" small dips in the error surface. The NADAM algorithm is one of the most recent advances in momentum-based SGD ([14]). Rather than just updating the parameter estimate (θ_k) at each time step, NADAM also updates a moving average of the gradient (m_k) and the squared gradient (n_k). The moving average of the gradient adds momentum, while the moving average of the squared gradient is used to adaptively scale updates according to the mean square error. The memory of the moving average gradients and squared gradients are controlled by the

hyperparameters μ and ν, respectively. A "regularization" hyperparameter (ε) stabilizes the learning rate and prevents division by zero. The NADAM algorithm thus updates as follows:

$$
\begin{aligned}
m_{k+1} &= \mu\, m_k + (1-\mu)\frac{\partial E}{\partial \theta_k} \\
n_{k+1} &= \nu\, n_k + (1-\nu)\frac{\partial E}{\partial \theta_k}^{2} \\
\theta_{k+1} &= \theta_k - \eta \frac{\frac{1-\mu}{1-\mu^{k+1}}\frac{\partial E}{\partial \theta_k} + \frac{\mu}{1-\mu^{k+2}} m_{k+1}}{\sqrt{\frac{n_{k+1}}{1-\nu^{k+1}}} + \varepsilon}
\end{aligned}
\qquad \textbf{(S.2)}
$$

Like its predecessor, the ADAM algorithm ([47]), NADAM makes use of momentum to avoid converging to shallow minima and also incorporates estimates of the error surface curvature ([14]). However, like all SGD methods, the NADAM algorithm is still only guaranteed to converge to a local minimum. The advantage, however, is that the NADAM algorithm improves the depth and breadth of that local minimum. Due to the limited amount of data per subject we prioritize robustness over goodness-of-fit so the global-minimum is not necessarily desirable and might actually correspond to over-fitting. There are thus two main advantages to using modified SGD over a global-optimizer: 1) computational efficiency, which enables us to fit very large networks, and 2) emphasis on robust solutions, which improves cross-validation and prevents over-fitting.

S.3 Hyperparameters in Model Fitting

In deconstructing the connectivity matrix, we produce three terms: one $n \times n$ sparse component (W_S) and one $n \times m$ rectangular matrix for each of the two diffuse components (W_1, W_2^T) in which n denotes the number of parcels and $m<n$ denotes the chosen dimensionality of the diffuse matrix. Hence, $W_{Full} := W_S + W_1 W_2^T$. The sparsity of W_S is achieved with L_1 regularization with penalty λ_1 and both of the diffuse components are also L_1 penalized with the same coefficient (λ_2) for both halves. The full diffuse matrix $W_1 W_2^T$ also receives L_2 regularization. The full, integrated cost function which includes the regularization penalty is thus:

$$
\begin{aligned}
J &= \frac{1}{2} E_T[\| (X_{T+1} - X_T) - [(W_S + W_L)\psi_A(X_T) - DX_T] \|_2^2] \\
&\ldots + \lambda_1 \| W_S \|_1 + \lambda_2 Tr(|W_S|) + \lambda_3(\| W_1 \|_1 + \| W_2 \|_1) + \frac{\lambda_4}{2} \| W_L \|_2^2
\end{aligned}
\qquad \textbf{(S.3)}
$$

with the notation E_T denoting the expected value over all time points within the minibatch. The NADAM algorithm itself involves four parameters: an update rate parameter, two NADAM decay parameters for computing moving averages, and one "regularization" parameter ([14]). Unlike the regularization parameters for the weight matrices, which factor into the error and steer the model towards sparse solutions, the NADAM regularization parameter simply serves to stabilize the speed of updates and prevent division by zero. We chose parameters for each variable: W_S, W_L, α, D. As with the regularization terms, we used the same parameters for the two halves of the diffuse component: W_1 and W_2. We found that the least impactful hyperparameters are the NADAM decay rate hyperparameters, which only need to be slightly less than one. The most impactful hyperparameters are the L_1 regularization penalties for the weight matrices which control the balance between over-fitting and under-fitting.

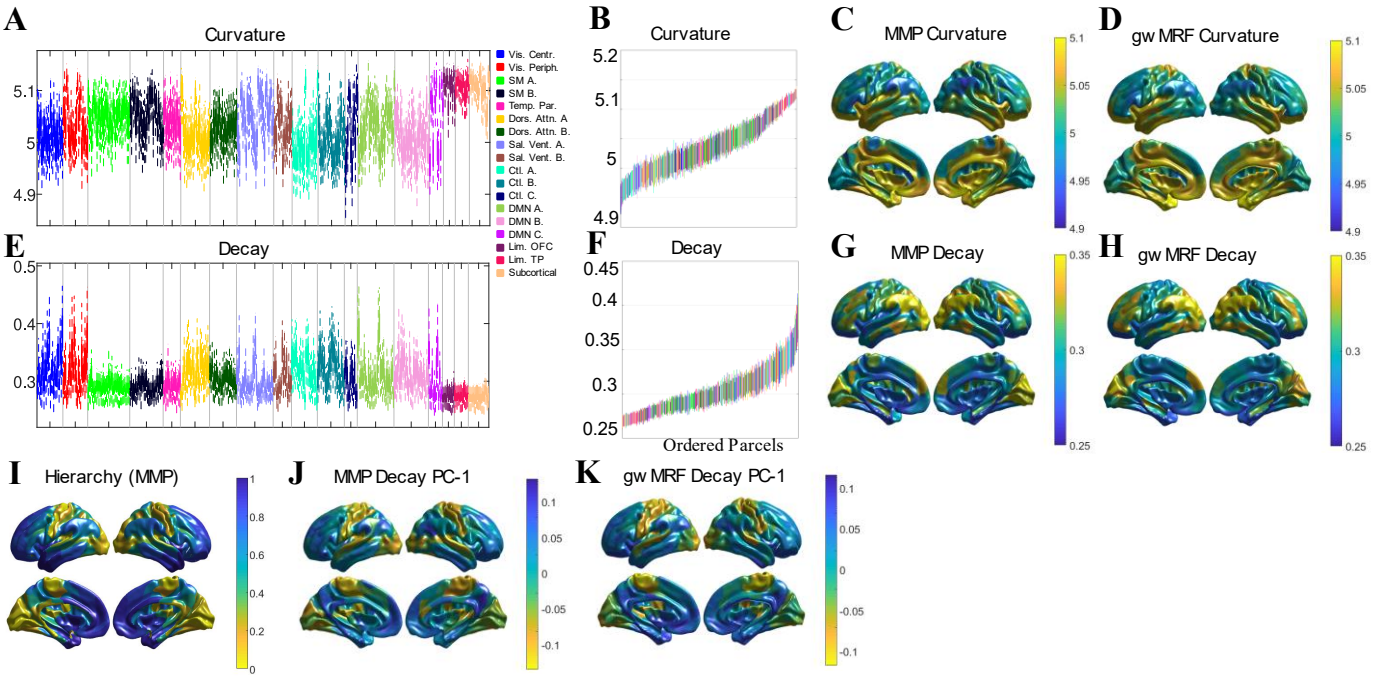
S.4 Directed Connectivity Identified by MINDy

The simplest way to characterize connection asymmetries is in terms of regions being sinks (input weights greater than output weights in absolute value) vs. sources (output weights greater than input weights in absolute value). For now, we focus upon sources and do so separately for positive (SI Fig. 3A) and negative connections (SI Fig. 3B). For positive sources MINDy most strongly identifies inferior frontal gyrus (IFG), bilateral parieto-occipetal sulcus, and dorsal prefrontal cortex (SI Fig. 3A). MINDy identifies the strongest excitatory source as a region of left IFG (see main text; SI Fig. 3C) and identified bilateral IFG as the strongest negative sources with negative outward-biased connections primarily to components of the Default Mode Network (IPL and medial PFC) with a general contralateral bias (SI Fig. 3D,E). The role of right IFG in inhibition is well-documented within neuroimaging (e.g. [48]) and lesion studies suggest an inhibitory role for left IFG as well ([49]). These results indicate that the asymmetries within MINDy weights are functionally interpretable. However, these initial findings only scratched the surface of possible analyses.

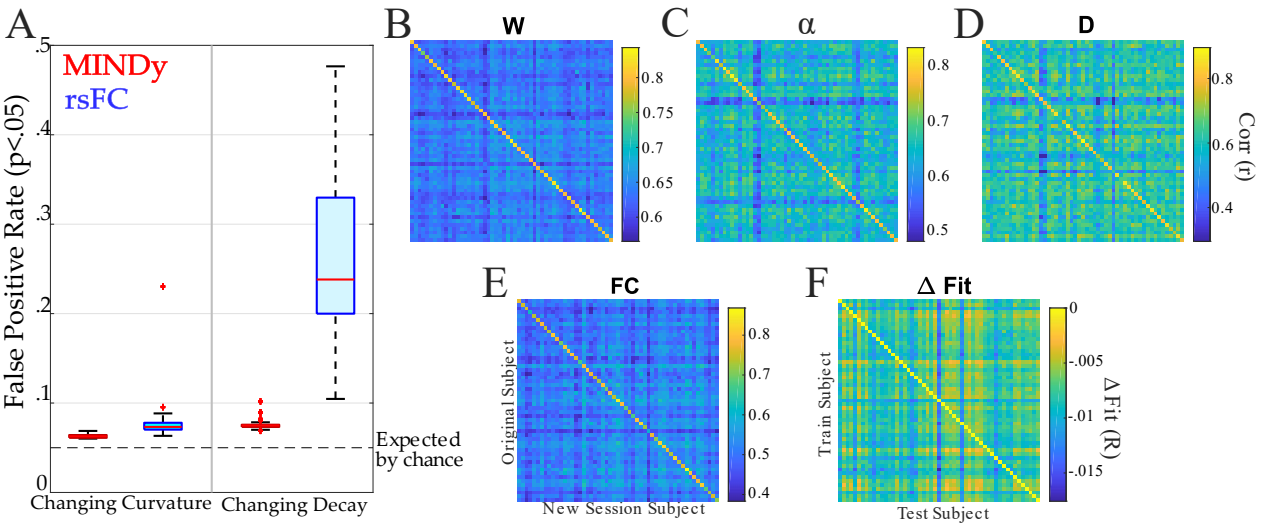
S.5 MINDy Optimization: Under the Hood

To stabilize MINDy's fitting procedure we use two changes of variable during the fitting process. Instead of directly fitting the term D we fit D_2 satisfying the relation $D{:=}D_{min}+D_2^2$. This change keeps the estimated parameters away from the pathological conditions in which D is either negative or very small which can cause simulated models to explode in the long term. The term D_{min} is a constant, positive hyperparameter. This step does not significantly alter computational complexity and we found that it did not alter results for our current initialization setting of D as our estimates never approached the pathological regions. However, we included this change of variable in the code as a safeguard should it become relevant for future users.

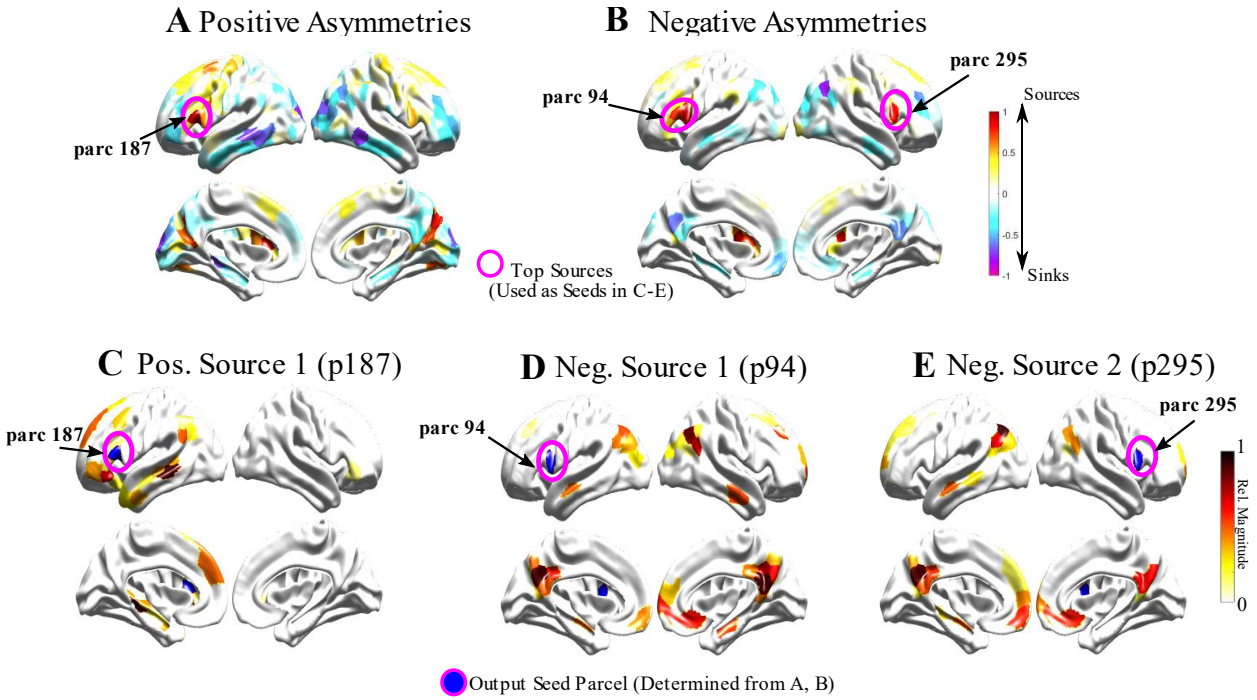
The second change of variable served to linearize the effects of the nonlinear curvature parameter α. Rather than explicitly fitting α we fit the variable $\xi{:=}b/\sqrt{(\alpha^2+.25)}$ which satisfies $\xi=\max_x\{d\psi_\alpha/dx\}$. This transformation smooths the relation between the nonlinear parameter (ξ instead of α) and its effects on the model's vector field. The new parameter ξ is constrained to be smaller than or equal to $2b$ so that $\alpha^2 \geq 0$. Efficient gradient calculations were performed by first calculating the variables in SI Table 3 (left) and then calculating gradients as in SI Table 3 (right). In all cases MINDy was run for 5000 iterations with batch size 300 (300 time-points used in each iteration).



SI Fig. 1. Local MINDy parameters exhibit consist anatomical structure. This structure occurs within and between networks. A) Distribution of curvature parameter values for each brain parcel grouped according to each of the 17 networks. B) Curvature parameters reordered according to mean demonstrate that within-network variability is also consistent. C) Anatomical profile of group-mean curvature for the MMP atlas. D) Profile for the gwMRF parcellation. E-H) same as A-D but for the decay parameter. I) Hierarchical Heterogeneity map by Demirtas and colleagues using group T1/T2 ratio. J,K) same as G,H but for the first principle component of decay across subjects.



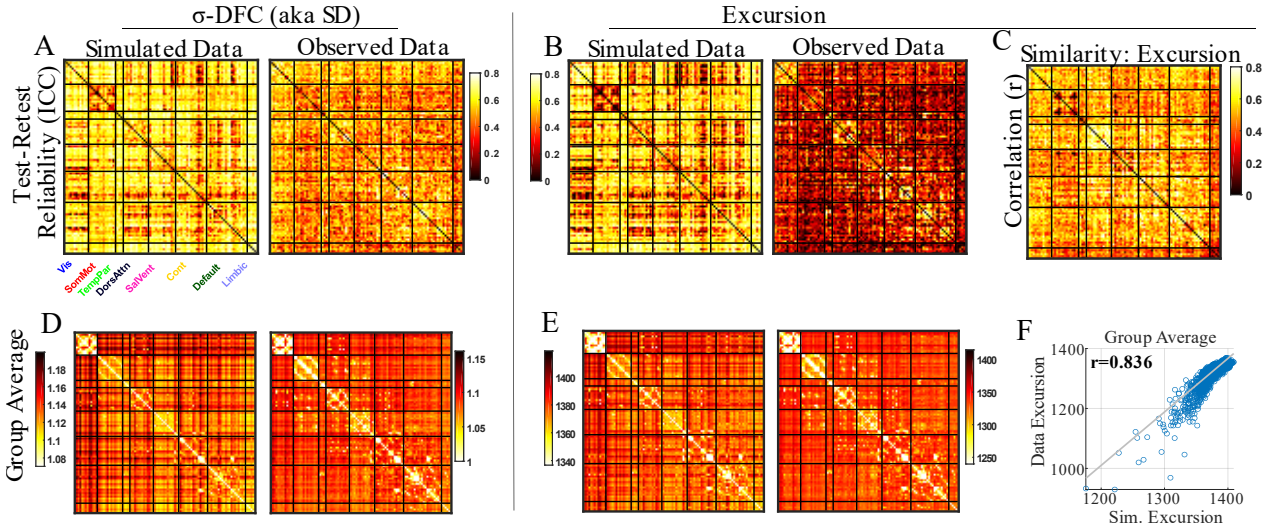
SI Fig. 2. MINDy identifies the source of individual differences. A) Changing only the curvature (left) or decay (right) parameters of a simulated subject has more impact on the simulated rsFC (blue) then on MINDy weight estimates refit to the new simulated data (red). B) Weight matrices are individualized: weight matrices derived from different scans of the same subject are universally more similar than weights fit to another subject. C,D) same as B) but for the curvature and decay parameters, respectively. E) The rsFC matrix is more similar for different scans of the same subject than between subjects. F) Individualized models better generalize to new data from the same subject than to a new subject.



SI Fig. 3. MINDy Connections are Directed. A) Difference of total input weights minus output weights for positive connections only (normalized units). B) Same as (A) but for negative connections only (using difference in magnitude of input/output weights). C) Difference of output and input weights for positive output-biased connections in the parcel with the greatest positive output-bias. D) Same as C), but for the parcel with greatest negative output bias. E) The parcel with the second-greatest negative output bias is the contralateral analogue to the parcel in (D). Parcel numbers are labeled for the 17-network gwMRF parcellation with 400 total parcels.

SI Table 1. Test-retest correlation for MINDy parameters. Sessions were either drawn from the same subject (n=53) or from different subjects between scanning sessions. Group-level statistics are present in mean correlation (SD) form. Group level permutation testing (100,000 each) produced *p*'s≈0 for all parameters vs. chance. Accuracy is in correct assignment for subjects based upon maximal similarity between sessions (e.g. how often is the subject most similar to themselves?). Statistical tests are for the contrast weight vs. FC. W/in=within subject, Btwn=between subject, Diff=w/in subject minus between, Acc.=accuracy. We denote significance with *=*p*<.05 and **=*p*<.001, 2-tailed and Bonferroni corrected

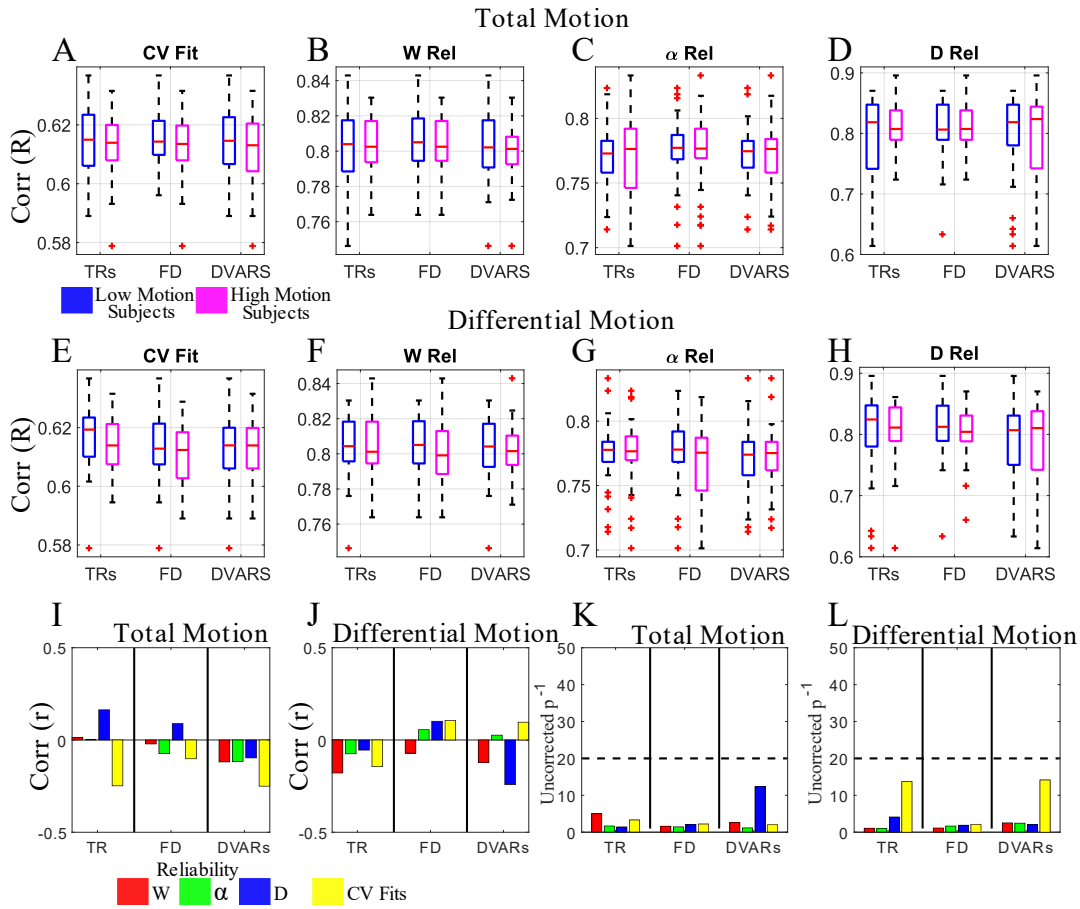
	FC	Weight (*W*)	*t* (mean)	Curv. (*α*)	Decay (*D*)
W/in.	.757 (.052)	.802 (.018)	7.883**	.772 (.029)	.798 (.062)
Btwn.	.511 (.044)	.635 (.021)	38.143**	.637 (.041)	.600 (.069)
Diff.	.246 (.038)	.167 (.019)	-23.076**	.135 (.023)	.198 (.052)
Acc.	100	100	0	100	94.3



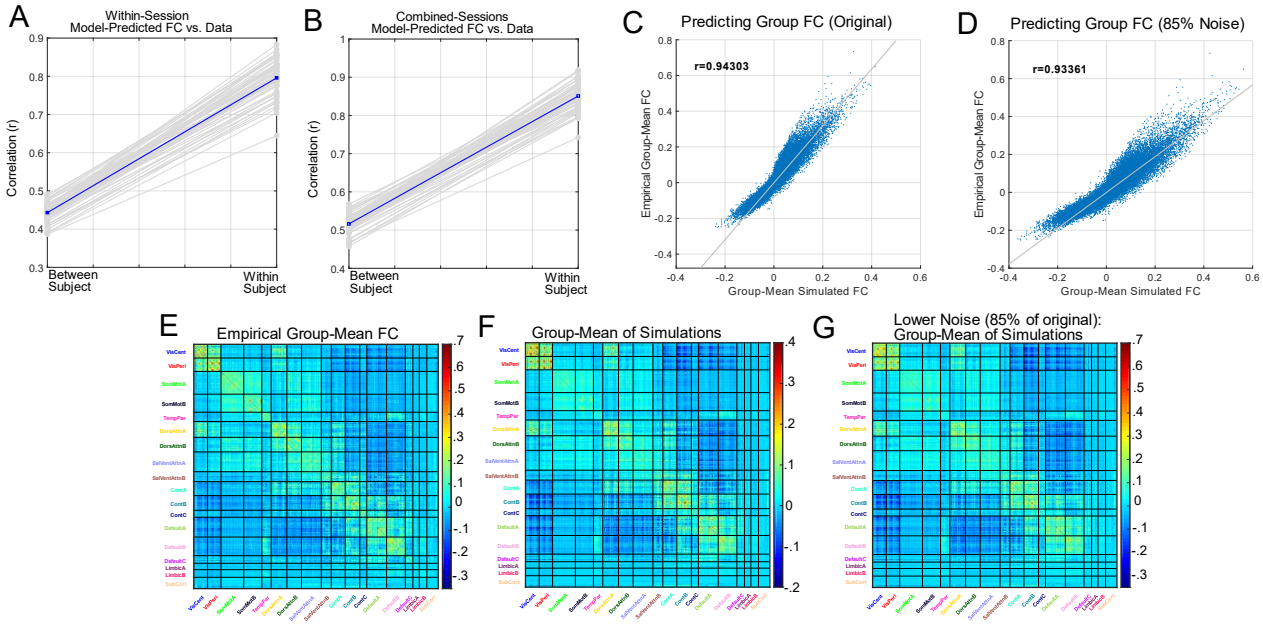
SI Fig. 4. MINDy reproduces dynamic functional connectivity (dFC). We use σ-dFC to disambiguate the standard-deviation measure for dFC from other uses of standard-deviation. A) Data simulated from test-retest models (models fit to separate sessions) has at least as high reliability on average as the original data for σ-dFC. B) Same as A) but for the excursion metric of dFC. C) Correlation between observed and simulated excursion across subjects by region-pair (combining across scanning sessions). D) Predicted group average σ-dFC for the model simulations (left) and recorded data (right) combining across scanning sessions. E) Same as D) but for excursion. F) Correlation between observed and predicted group-average excursion across region-pairs. The σ-dFC analogues of C and F are reported in the main text (Fig. 5).

SI Table 2. Preprocessing sensitivity of MINDy and FC. Comparing the test-retest reliability (top) and pre-processing sensitivity (bottom) of the MINDy parameters and the resting state functional connectivity. The pipelines correspond to using motion without CompCor or GSR correction, using motion + CompCor or using motion + CompCor + GSR (default). Results are presented in mean(SD) form for the group distribution of individual test-retest correlations or correlations between different levels of preprocessing applied to the same session. Statistical tests consisted of paired t-tests for the mean correlation, and F-tests for testing heterogeneity of variance (n=53). Results generally favored the MINDy connectivity matrix over the FC matrix (greater mean reliability and less variation) but the absolute differences, although highly statistically significant, are not profound. We denote significance with *=p<.05 and **=p<.001, 2-tailed and Bonferroni corrected.

	Weights	FC	t (mean)	F (var)	Curv (α)	Decay (D)
Test-Retest Correlations by Pipeline						
Motion	.810 (.027)	.791 (.062)	3.151*	5.322**	.689 (.081)	.827 (.076)
M+CompCor	.804 (.019)	.748 (.050)	10.510**	7.378**	.757 (.031)	.801 (.065)
M+C+ GSR	.802 (.018)	.757 (.052)	7.883**	8.420**	.772 (.029)	.798 (.062)
Similarity (R) Across Pipelines						
M vs. C	.932 (.018)	.851 (.058)	13.633**	10.725**	.856 (.084)	.894 (.049)
M vs. G	.912 (.019)	.791 (.078)	14.150**	16.482**	.835 (.091)	.854 (.065)
C vs. G	.974 (.004)	.967 (.014)	4.216**	14.627**	.976 (.007)	.976 (.013)



SI Fig. 5. MINDy fits are robust to motion. Fitting performance was measured by the cross-validated goodness of fit (A,E) and the reliability for each parameter (B-D,F-H). Individual differences in motion were quantified by either membership in median-split high vs. low motion groups (first two rows) or as a continuous variable (bottom row). Groups were assigned for each combination of motion measurement (number of TRs censored, median Framewise Displacement, or Median Absolute Deviation (MAD) of DVARS) and motion type: either the total motion of a subject averaged across scanning sessions (A-D, I,K) or the absolute difference in motion artifact between sessions (E-H,J,L). There was no significant relationship with motion as a discrete characteristic (e.g. high vs. low: A-H) or as a continuous characteristic: group level correlations between motion measures and fitting performance in (I,J) and the associated (uncorrected) inverse *p*-values in (K,L).



SI Fig. 6. MINDy predicts individual and group FC. A) Correlation between empirical FC and simulated FC from the same scanning day for either the same subject or a different subject (mean across other subjects). Blue line indicates group mean. Correlations are averaged across scanning sessions. B) Same as A) but for with all data combined across sessions (simulations from each session's model were combined) C) Group-mean of empirical FC vs. group-mean of simulated FC (both combined across sessions). Notice that while the correlation is high, the magnitude of simulated FC is smaller than empirical. D) Slightly decreasing noise produces a predict group-mean FC that is nearly indistinguishable from that observed empirically. E) Group-average empirical FC combined across sessions. F) Group-average FC of simulated data combined across sessions. G) Same as F) but with a slightly decreased noise term (85% of original).

SI Table 3. MINDy gradient equations. Left: Variables used in MINDy gradient calculations. Right: Negative error gradient for MINDy parameters. Note that gradient updates are in terms of the transformed variables $\xi := b/(\alpha^2+.25)^{1/2}$ and $D_2 := (D-D_{min})^{1/2}$ rather than directly in terms of α and D, respectively. X denotes the data matrix (minibatch) of each iteration. n_{Batch} denotes the number of time-points in each minibatch.

Name	Formula	Meaning
W	W_S+W_L	Full Weights
$P_1(X)$	$\sqrt{\xi^{-2}+X\circ(X+b^{-1})}$	Part of ψ
$P_2(X)$	$\sqrt{\xi^{-2}+X\circ(X-b^{-1})}$	Part of ψ
$\psi(X,\varepsilon)$	$b(P_1-P_2)$	Transfer Fct.
D	$D_2^2+D_{min}$	Full Decay
R	$dX-W\psi(X)-DX$	Residual Error
Q	$W_L^T R$	$2[\partial R/\partial\psi]$
Z	$R\psi(x)^T/n_{Batch}$	$E_T[\partial R/\partial W]$
Y	$Z-\lambda_4 W_L$	$\partial J/\partial W_L$

Variable	Neg. Error Grad: $-[\partial J/\partial\theta]$
W_S	$Z-\lambda_1\mathrm{sgn}(W_S)-\lambda_4\mathrm{diag}(\lvert W_S\rvert)$
W_1	$YW_2^T-\lambda_3\mathrm{sgn}(W_1)$
W_2	$W_1Y-\lambda_3\mathrm{sgn}(W_2)$
ξ	$-\xi^{-3}bE_T[Q\circ(1/P_1-1/P_2)]$
D_2	$-2D_2\circ E_T[R\circ X]$

SI Table 4. MINDy hyperparameters. Top: NADAM hyperparameters and initial distributions associated with each model parameter. Bottom: Other hyperparameters.

Parameter	Rate ($\eta\times10^5$)				

W_S	2.5	.9	.95	.15	$N(0, .01^2)$	
$W_{1,2}$	6.25	.9	.95	.15	$N(0, .01^2)$	
A	12.5	.9	.95	.2	$.1+\sqrt{.25+\lvert N(0,.25^2)\rvert}+b^2/4$	
D_2	1750	.9	.95	200	$1.75+\sqrt{\lvert N(0,.25^2)\rvert}$	
Hyperparameter	λ_1	λ_2	λ_3	λ_4	D_{min}	b
Value	.075	.2	.05	.05	.1	20/3